



Pontificia Universidad
Católica del Ecuador

UNIDAD ACADÉMICA:

OFICINA DE POSTGRADO

TEMA:

ANÁLISIS DE METODOLOGÍAS PARA LA IMPLEMENTACIÓN DE UN DATA
WAREHOUSE APLICADO A LA TOMA DE DECISIONES DEL INSTITUTO
NACIONAL DE PATRIMONIO CULTURAL REGIONAL 3

Proyecto de Investigación y Desarrollo previo a la obtención del título de

Magister en Gerencia Informática

Línea de Investigación, Innovación y Desarrollo principal:

Sistemas de Información y/o Nuevas Tecnologías de la Información y Comunicación y sus
aplicaciones

Caracterización técnica del trabajo:

Aplicación

Autor:

Geovanny Euclides Silva Peñafiel

Director:

José Marcelo Balseca Manzano, Ing. Mg.

Ambato – Ecuador

Marzo 2018

Análisis de metodologías para la implementación de un data warehouse aplicado a la toma de decisiones del Instituto Nacional de Patrimonio Cultural Regional 3

Informe de Trabajo de Titulación
presentado ante la
Pontificia Universidad Católica del Ecuador Sede Ambato

por

Geovanny Euclides Silva Peñafiel

En cumplimiento parcial de
los requisitos para el Grado de
Magister en Gerencia Informática



Pontificia Universidad
Católica del Ecuador

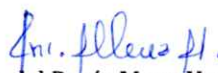
Oficina de Postgrado
Marzo 2018

Análisis de metodologías para la implementación de un data warehouse aplicado a la toma de decisiones del Instituto Nacional de Patrimonio Cultural Regional 3

Aprobado por:



Concepción del Carmen Bedón Vaca, Mg.
Presidenta del Comité Calificador
Coordinadora de Postgrados



Liliana del Rocío Mena Hernández, Mg.
Miembro Calificador



José Marcelo Balseca Manzano, Ing. Mg.
Miembro Calificador
Director de Proyecto



Dr. Hugo Rogelio Altamirano
Secretario General



Pontificia Universidad
Católica del Ecuador

SECRETARIA GENERAL
VICERRECTORIA



Dennis Vinicio Chicaiza Castillo, Mg.
Miembro Calificador

Fecha de aprobación:
Marzo 2018



Pontificia Universidad
Católica del Ecuador

BIBLIOTECA

Ficha Técnica

Programa: Magister en Gerencia Informática

Tema: Análisis de metodologías para la implementación de un *Data Warehouse* aplicado a la toma de decisiones del Instituto Nacional de Patrimonio Cultural Regional 3

Tipo de trabajo: Proyecto de Investigación y Desarrollo

Clasificación técnica del trabajo: Aplicación

Autor: Geovanny Euclides Silva Peñafiel

Director: José Marcelo Balseca Manzano, Ing. Mg.

Líneas de Investigación, Innovación y Desarrollo

Principal: Sistemas de Información y/o Nuevas Tecnologías de la Información y Comunicación y sus aplicaciones

Resumen Ejecutivo

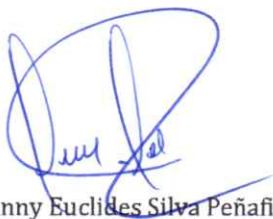
El constante crecimiento de la información está generando grandes cantidades de datos que no prestan las condiciones necesarias para generar e interpretar la información que día a día se va gestionando, lo cual retarda la toma de decisiones. El análisis de datos se ha venido perfeccionando mediante la aparición de los sistemas Business Intelligence, los cuales presentan un mayor potencial a nivel tecnológico y gerencial (toma de decisiones) para una adecuada utilización de los recursos de la empresa.

La investigación descriptiva-explicativa pretende evaluar parámetros de las Metodologías Hefesto, Kimball y SAS Rapid, para determinar la Metodología que más aporte al análisis de datos del Instituto Nacional de Patrimonio Cultural Regional 3, e implementar un Data Warehouse para la toma de decisiones, lo cual ofrecerá una solución tecnológica que permitirá procesar la información de mejor manera para alcanzar los objetivos institucionales y agreguen valor a los servicios prestados.

La implementación de un Data Warehouse para Business Intelligence (BI) logrará generar reportes más específicos y detallados, se acortarán los tiempos de procesamiento de información, se podrá tomar decisiones de una forma más estratégica respaldada en información íntegra.

Declaración de Originalidad y Responsabilidad

Yo, Geovanny Euclides Silva Peñafiel, portador de la cédula de ciudadanía y/o pasaporte No. 0602891764, declaro que los resultados obtenidos en el proyecto de titulación y presentados en el informe final, previo a la obtención del título de Magister en Gerencia Informática, son absolutamente originales y personales. En tal virtud, declaro que el contenido, las conclusiones y los efectos legales y académicos que se desprenden del trabajo propuesto, y luego de la redacción de este documento, son y serán de mi sola y exclusiva responsabilidad legal y académica.



Geovanny Euclides Silva Peñafiel

0602891764



Pontificia Universidad
Católica del Ecuador

BIBLIOTECA

Dedicatoria

Al creador de todas las cosas, el que me ha dado fortaleza para continuar cuando a punto de caer he estado; por ello, con toda la humildad que de mi corazón puede emanar, agradezco primeramente, a Dios.

A mi madre, por ser el pilar más importante y pesar de nuestra distancia física, siento que estás conmigo siempre y aunque nos faltaron muchas cosas por vivir juntos, sé que este momento hubiera sido tan especial para ti como lo es para mí.
A mi padre, por su apoyo incondicional sin importar nuestras diferencias de opiniones.

Son muchas las personas que han formado parte de mi vida profesional a las que me encantaría agradecerles su amistad, consejos, apoyo, ánimo y compañía en los momentos más difíciles de mi vida. Algunas están aquí conmigo y otras en mis recuerdos y en mi corazón, sin importar en donde estén quiero darles las gracias por formar parte de mí, por todo lo que me han brindado y por todas sus bendiciones.

Para ellos: Muchas gracias y que Dios los bendiga

Geovanny Euclides Silva Peñafiel

Reconocimientos

Al Ing. Marcelo Balseca Manzano, por su asesoría, amistad y apoyo durante el desarrollo de mi vida académica en la Maestría y la culminación de este proyecto de Investigación y Desarrollo.

A la Arq. Sara Silva Cajas, Directora del Instituto Nacional de Patrimonio Cultural Regional 3, por su interés en el desarrollo institucional con la aplicación de este trabajo con miras a un servicio más automatizado. Por su amistad y apoyo a lo largo de mi estancia en el instituto.

Resumen

La investigación persigue la implementación de un *Data Warehouse* para la toma de decisiones del Instituto Nacional de Patrimonio Cultural Regional 3; que permita la selección de la metodología adecuada para el manejo de los datos de la institución, la exteriorización de los datos se presentará de forma cualitativa, los cuales servirán de soporte para la identificación de los factores importantes que deben ser valorados. Debido a la naturaleza del estudio de tipo descriptiva comparativa, se apoyará en el desarrollo de prototipos metodológicos con las propuestas realizadas por Ralph Kimball, Hefestos 2.0 y SAS *Rapid Data Warehouse Methodology*; por medio del análisis tanto a nivel de características como de las fases que integran las metodologías; para ello se tomaron parámetros tales como: el grado del cumplimiento de los objetivos, la capacidad de mejoras y del grado de satisfacción de los directivos, empleados en: “Análisis, diseño, construcción e implementación de un *Data Warehouse* para la toma de decisiones y construcción de los KPI, para la Empresa Kronosconsulting Cia Ltda.”; y los indicadores para la evaluación de las fases tomados de “Estudio comparativo de metodologías e implementación de alternativas *Business Intelligence OpenSource* vs. propietarias en entornos tradicionales”; estos indicadores están basados en las fases y actividades o índices involucrados en la construcción de *Data Warehouse*. Del estudio se ha obtenido resultados que permiten brindar una cimentación robusta para la creación de la propuesta con la metodología Hefestos versión 2.0; debido al alto porcentaje adquirido en las 7 fases examinadas, con sus 54 puntos, mientras que, en el análisis de características se obtienen 29 puntos para Hefestos. El uso de la metodología Hefestos, permitió identificar ágilmente y comprender los objetivos que la institución deseaba alcanzar con la implementación del *Data Warehouse*, ayudó a determinar las necesidades al establecer en los requerimientos en base al usuario; involucrándolo en cada etapa del proceso de información y determinar los datos idóneos para la toma de decisiones.

Palabras Clave: *Data Warehouse*, Metodologías, Ralph Kimball, Hefestos 2.0, comparativa.

Abstract

The study follows the implementation of a data warehouse for decision making at the Regional 3 National Institute of Cultural Heritage, which facilitates the selection of a suitable methodology for data management at the institution. The exteriorization of the data will be presented qualitatively; this data will help to support the identification of the important factors that need to be valued. Due to the nature of a comparative descriptive study, the work will be based on the development of methodological prototypes with the proposals made by Ralph Kimball, Hefestos 2.0 and SAS Rapid Data Warehouse Methodology through the analysis of both the features and the stages that the methodologies integrate. To this end, the parameters taken were the fulfilment level of the objectives, the capacity for improvements and the satisfaction level of managers and employees in the “Analysis, Design, Construction and Implementation of a Data Warehouse for Decision-making and the Construction of KPI for the Company Kronosconsulting Co. Ltd.” as well as the indicators for the evaluation of the stages taken from “A Comparative Study of Methodologies and the Implementation of Business Intelligence OpenSource Alternatives vs. One’s Own in Traditional Environments”. These indicators are based on the stages and activities or indexes involved in the construction of a data warehouse. The findings of the study have made it possible to provide a sturdy foundation for the creation of the proposal with Hefestos version 2.0 methodology due to the high percentage obtained in the 7 stages that are examined with its 54 points while in the analysis of features, 29 points are obtained for Hefestos. The use of Hefestos methodology made it possible to easily identify and understand the goals that the institution would like to achieve with the implementation of the data warehouse. It also helped to determine needs as the requirements are established based on the user by involving it in each step of the information process and determining the ideal data for decision-making.

Key words: data warehouse, Hefestos 2.0, Kimball, SAS Rapid

Tabla de Contenidos

Ficha Técnica	iii
Líneas de Investigación, Innovación y Desarrollo.....	iii
Resumen Ejecutivo.....	iii
Declaración de Originalidad y Responsabilidad	iv
Dedicatoria.....	v
Reconocimientos	vi
Resumen.....	vii
Abstract.....	viii
Tabla de Contenidos	ix
Lista de Tablas.....	xii
Lista de Figuras.....	xiii
Lista de Gráficos.....	xv
CAPITULOS	
1. Introducción	1
1.1. Presentación del trabajo	1
1.2. Descripción del documento.....	2
2. Planteamiento de la propuesta de trabajo	3
2.1. Información técnica básica	3
2.2. Descripción del problema	3
2.3. Preguntas básicas.....	3
2.4. Formulación de meta.....	4
2.5. Objetivos.....	4
2.6. Delimitación funcional.....	4
3. Marco Teórico	5
3.1. Introducción	5
3.2. Inteligencia de negocios.....	6
3.2.1. Definiciones	6
3.2.2. Datos, información y conocimiento.....	6
3.2.2.1. Reporteo.....	8
3.2.2.2. Procesamiento analítico en línea (OLAP).....	8
3.2.2.3. Minería de datos (Data Mining)	8
3.2.3. Ventajas de <i>Business Intelligence</i>	8
3.2.4. Desventajas de <i>Business Intelligence</i>	9
3.2.5. Arquitectura típica de soluciones <i>Business Intelligence</i>	10
3.2.5.1. Funciones.....	10
3.3. <i>Data Mining</i> (minería de datos)	11
3.3.1. Definiciones	11

3.3.2.	Etapas de <i>Data Mining</i>	12
3.4.	Data Warehousing.....	12
3.4.1.	Definiciones	13
3.4.2.	Data Warehouse	13
3.4.3.	Definiciones	14
3.4.3.1.	Definición de Bill Inmon	14
3.4.3.2.	Definición de Ralph Kimball	15
3.4.3.3.	Definición de Bernabeu.....	16
3.4.3.4.	Una definición más amplia de almacén de datos	17
3.4.4.	Características y objetivos.....	17
3.4.5.	Arquitectura de un <i>Data Warehouse</i>	18
3.4.5.1.	Arquitectura de flujo de datos	18
3.4.5.2.	Arquitectura del sistema	23
3.4.6.	Herramientas de <i>Data Warehouse</i>	24
3.4.6.1.	Herramientas de almacenamiento	24
3.4.6.2.	Herramientas de extracción, transformación y carga.....	24
3.4.6.3.	Herramientas de análisis intelligence	24
3.5.	Estado Arte	24
3.5.1.	Metodología Ralph Kimball.....	24
3.5.1.1.	Introducción	24
3.5.1.2.	Características	25
3.5.1.3.	Fases.....	25
3.5.2.	Metodología Hefestos.....	39
3.5.2.1.	Introducción	39
3.5.2.2.	Características	40
3.5.2.3.	Fases.....	41
3.5.3.	<i>SAS Rapid Data Warehouse Methodology</i>	49
3.5.3.1.	Introducción	49
3.5.3.2.	Fases.....	49
4.	Metodología	53
4.1.	Diseño de la Investigación	53
4.2.	Métodos, Técnicas e Instrumentos	54
4.2.1.	Métodos	54
4.2.2.	Técnicas	54
4.2.3.	Instrumentos.....	55
4.3.	Procesamiento de la Información.....	55
4.4.	Parámetros a comparación de fases metodológicas	55
4.5.	Población y muestras	56
4.5.1.	Creación de Prototipos.....	57
4.5.1.1.	Metodología Kimball	57

4.5.1.2.	Metodología Hefesto.....	57
4.5.1.3.	SAS Methodology.....	57
5.	Resultados.....	59
5.1.	Evaluación preliminar.....	59
5.1.1.	Determinación de parámetros de comparación fase metodológica.....	59
5.1.1.1.	Fase 1. Requerimentación.....	59
5.1.1.2.	Fase 2: Estrategia de proyecto.....	63
5.1.1.3.	Fase 3: Planificación del proyecto.....	66
5.1.1.4.	Fase 4: Selección de la tecnología	70
5.1.1.5.	Fase 5: Diseño del sistema de información	73
5.1.1.6.	Fase 6: Elaboración del sistema de información	76
5.1.2.	Determinación de parámetros de comparación de características metodológicas.....	89
5.1.2.1.	Características	89
5.2.	Análisis de resultados	100
5.2.1.1.	Resultados de comparativas fases metodológicas.....	100
5.2.1.2.	Resultados de comparativas características metodológicas.....	100
6.	Conclusiones y Recomendaciones.....	124
6.1.	Conclusiones.....	124
6.2.	Recomendaciones	125
	APÉNDICES	126
	Apéndice A. — METODOLOGÍA KIMBALL.....	126
	Apéndice B. — METODOLOGÍA HEFESTO.....	170
	Apéndice C. — SAS RAPID DATA WAREHOUSE METHODOLOGY	209
	REFERENCIAS.....	231

Lista de Tablas

1: Funciones de los componentes del <i>Business Intelligence</i>	10
2: Factores de retorno de la inversión (ROI)	28
3: Consideraciones del diseño arquitectónico	34
4: Parámetros de comparación de fases metodológicas.....	55
5: Escala de Likert.....	59
6: Indicador. Fase 1. Requerimientos.....	59
7: Parámetro. Determinación de requerimientos	60
8: Parámetro. Características de la organización	61
9: Parámetro. Análisis de usuarios.....	62
10: Resumen. Indicador requerimientos	63
11: Indicador. Fase 1. Requerimientos	63
12: Parámetro. Cuantificar el tiempo para cumplir la necesidad	64
13: Parámetro. Definir ventajas y desventajas	65
14: Resumen. Indicador estrategias del proyecto	66
15: Indicador. Fase 3. Planificación del proyecto	66
16: Parámetro. Cronograma de actividades	67
17: Parámetro. Análisis de riesgos	68
18: Parámetro. Definición de responsabilidades y grupos de trabajo	69
19: Resumen. Indicador planificación del proyecto	70
20: Indicador. Fase 4. Selección de la tecnología.....	70
21: Parámetro. Entorno actual tecnológico de la organización.....	71
22: Parámetro. Definición de tecnologías.....	72
23: Resumen. Indicador selección de la tecnología	73
24: Indicador. Fase 5. Diseño del sistema de información	73
25: Parámetro. Determinación del modelo de información	74
26: Parámetro. Diseño de la interfaz de usuario	75
27: Resumen. Indicador diseño del sistema	76
28: Indicador. Fase 6. Elaboración del sistema de información.....	76
29: Parámetro. Análisis de requerimientos	77
30: Parámetro. Modelo conceptual	78
31: Parámetro. Análisis OLTP	79
32: Parámetro. Conformar indicadores	80
33: Parámetro. Nivel de granularidad.....	81
34: Parámetro. Nivel conceptual ampliado.....	82
35: Parámetro. Modelo lógico de la estructura del DW	83
36: Parámetro. Tablas de dimensiones	84
37: Parámetro. Tablas de hechos	85
38: Parámetro. Proceso ETL.....	86
39: Resumen. Indicador elaboración del sistema de información	87
40: Indicador. Características metodológicas.....	89
41: Parámetro. Adaptable sobre cualquier metodología.....	90
42: Parámetro. Afinidad con el sistema actual	91
43: Parámetro. Comunicación con el cliente.....	92
44: Parámetro. Tamaño del proyecto	93
45: Parámetro. Tiempo en el análisis y diseño	94
46: Parámetro. Tiempo en el análisis y diseño	95
47: Parámetro. Fácil entendimiento en principiantes	96
48: Parámetro. Fácil entendimiento en principiantes	97
49: Parámetro. Más usado en el mercado	98

50: Resumen. Indicador características metodológicas	99
51: Indicadores y perspectivas	101
52: Herramientas tecnológicas	105

Lista de Figuras

1. Datos, información y conocimiento.....	7
2. Arquitectura de una solución BI	10
3. Arquitectura de un <i>Data Warehouse</i>	14
4. Arquitectura de un <i>Data Warehouse</i> según Bill Inmon	15
5. Arquitectura de un <i>Data Warehouse</i> según Ralph Kimball.....	16
6. Arquitectura de flujo de datos	18
7. Arquitectura simple con un almacén de datos dimensional	21
8. Arquitectura con un almacén de datos normalizado más un almacén de datos dimensional.....	22
9. Arquitectura con un almacén de datos operacional más un almacén de datos dimensional.....	22
10. Arquitectura física de un <i>Data Warehouse</i>	23
11. Fases Metodología Ralph Kimball.....	26
12. Factores de determinación de un proyecto	27
13. Ejemplo cubo dimensional Kimball.....	31
14. Ejemplo granularidad Kimball.....	32
15. Fases Metodología Hefesto.....	41
16. Ejemplo modelo conceptual Hefesto	42
17. Ejemplo establecimiento de correspondencias en modelo conceptual Hefesto	43
18. Ejemplo modelo conceptual ampliado Hefesto	44
19. Ejemplo transformación de tablas de dimensiones Hefesto	45
20. Ejemplo tabla jerarquía Hefesto	46
21. Ejemplo esquema estrella Hefesto	47
22. Ejemplo uniones de tablas en Hefesto	47
23. Ejemplo del proceso de carga inicial en Hefesto.....	48
24. Ejemplo del proceso de carga dimensión cliente en Hefesto.....	48
25. Fases metodología SAS.....	49
26. Fuentes de datos inpc	102
27. Modelo conceptual para el inpc	103
28. Modelo lógico de los procesos trámites	104
29. Ventana workspace launcher.....	105
30. Ventana de ingreso del nombre del proyecto	106
31. Ventana de creación de una nueva conexión de base de datos	106
32. Ventana de selección del controlador	106
33. Ventana de ingreso de datos de conexión.....	107
34. Ventana de validación del proceso de testeo.....	107
35. Ventana de creación del modelo de datos.....	107
36. Ventana de selección de tablas.....	108
37. Ventana de visualización del modelo	108
38. Ventana de creación de la plantilla.....	109
39. Ventana de selección de la ubicación de la plantilla	109
40. Ventana de edición de la plantilla creada.....	110
41. Ventana de selección de la opción publicar	110
42. Ventana de ingreso y selección del <i>datasource</i>	111
43. Ventana de ingreso a <i>spagobi</i>	111
44. Ventana de bienvenida a <i>spagobi</i>	112

45. Ventana de visualización del reporte generado.....	112
46. Ventana de visualización de un reporte.....	113
47. Menú de reporte.....	113
48. Ventana de filtros de un reporte.....	113
49. Visualización de la organización por filas y columnas de un <i>grid</i>	114
50. Ventana para ordenar un <i>grid</i>	114
51. Despliegue de la ventana guardar.....	115
52. Visualización del esquema de las columnas del reporte.....	115
53. Visualización del esquema invertido filas y columnas.....	116
54. Visualización de las propiedades de un determinado gráfico.....	116
55. Visualización de las propiedades de impresión.....	117
56. Visualización de un reporte seleccionado.....	117

Lista de Gráficos

1. Parámetro. Determinación de requerimientos.....	60
2. Parámetro. Características de la organización	61
3. Parámetro. Análisis de usuarios	62
4. Resumen. Indicador Requerimientos.....	63
5. Parámetro. Cuantificar el tiempo para cumplir la necesidad	64
6. Parámetro. Definir ventajas y desventajas	65
7. Resumen. Indicador estrategias del proyecto	66
8. Parámetro. Cronograma de actividades.....	67
9. Parámetro. Análisis de riesgos	68
10. Parámetro. Definición de responsabilidades y grupos	69
11. Resumen. Indicador planificación del proyecto	70
12. Parámetro. Entorno actual tecnológico de la organización	71
13. Parámetro. Definición de tecnologías	72
14. Resumen. Indicador selección de la tecnología	73
15. Parámetro. Determinación del modelo de información.....	74
16. Parámetro. Diseño de la interfaz de usuario.....	75
17. Resumen. Indicador diseño del sistema.....	76
18. Parámetro. Análisis de requerimientos.....	77
19. Parámetro. Modelo conceptual.....	78
20. Parámetro. Análisis de OLTP.....	79
21. Parámetro. Conformar indicadores.....	80
22. Parámetro. Nivel de granularidad.....	81
23. Parámetro. Nivel conceptual ampliado	82
24. Parámetro. Modelo lógico de la estructura del DW.....	83
25. Parámetro. Tablas de dimensiones	84
26. Parámetro. Tablas de hechos	85
27. Parámetro. Proceso ETL.....	86
28. Resumen. Indicador elaboración del sistema de información	88
29. Parámetro. Adaptable sobre cualquier metodología.....	90
30. Parámetro. Afinidad con el sistema actual.....	91
31. Parámetro. Comunicación con el cliente.....	92
32. Parámetro. Tamaño del proyecto.....	93
33. Parámetro. Tiempo en el análisis y diseño	94
34. Parámetro. Tiempo en construcción.....	95
35. Parámetro. Fácil entendimiento principiantes	96
36. Parámetro. Documentación precisa.....	97
37. Parámetro. Más usada en el mundo	98
38. Parámetro. Características metodológicas	99
39. Título. Trámites en el tiempo	118
40. Título. Trámites por remitentes en el tiempo	118
41. Título. Trámites asignados en el tiempo.....	119
42. Título. Trámites por estados.....	120
43. Título. Trámites por localidad.....	121
44. Título. Trámites por prioridad.....	122
45. Título. Número de días de proceso de trámite.....	122
46. Título. Trámites por medio de entrega.....	123

Capítulo 1

Introducción

En la actualidad la cantidad de datos que se obtienen de las entidades públicas y privadas, son almacenados en diferentes repositorios haciendo imposible el acceso a ellos, para la obtención de resultados debido al desconocimiento de procesos que posibiliten la extracción de datos que sean de ayuda a nivel gerencial; la información útil para el proceso de toma de decisiones debe ser sólida y oportuna que faciliten en la toma de decisiones para cumplir con los objetivos que garanticen la aplicación de mejoras y acciones.

El Instituto Nacional de Patrimonio Cultural Regional 3 actualmente maneja la información por medio de hojas electrónicas en Excel; mismas que presentan el registro, recepción de documentos oficiales a ser tramitados. El volumen de datos que maneja la entidad es alto en la recolección de información para la construcción de reportes, generando tiempos excesivos, en referencia a los intervalos de procesamiento y desfases al momento de la entrega de respuestas de los trámites que se realizan.

Por lo que se ha visto la necesidad con el presente estudio orientar al análisis de metodologías con el propósito de determinar la más óptima para la construcción de un repositorio que permita seguir con el procedimiento de control y tratamiento respectivo de la información a tramitar que llega al Instituto Nacional de Patrimonio Cultural Regional 3 permitiendo el planteamiento y diseño de una colección de datos.

Al analizar las fases y los procesos que se realizan en cada una de las metodologías en el desarrollo del *Data Warehouse*, se acopló a Hefesto versión 2.0 que es una metodología propia para la construcción de *Data Warehouse*, que garantiza mayores prestaciones motivo por el que se aplicó en el INPC R3, certificando los datos para la toma de decisiones.

1.1. Presentación del trabajo

Cada día surgen estudios orientados a analizar la información de una institución sea esta pública o privada como un elemento importante para apoyar a la toma de decisiones. Sin embargo, los sistemas tradicionales no presentan flexibilidad a la hora en que los mandos medios y altos necesitan extraer información sustancial estratégica. Los sistemas tradicionales simplemente no cuentan con

la capacidad de presentar información entendible a dichos niveles de la organización debido a que no están orientados a este objetivo.

Para la toma de decisiones se creó prototipos con las metodologías de Ralph Kimball, Hefestos y *SAS Rapid Data Warehouse Methodology*; por medio del análisis realizado a cada una de las fases de estas metodologías; los resultados obtenidos, han brindado una cimentación robusta para la creación de la propuesta con la metodología Hefestos versión 2.0; siendo la que más se ajusta a las necesidades del Instituto Nacional de Patrimonio Cultural Regional 3 debido al diseño y planificación detallada en la construcción de cada uno de los procesos del *Data Warehouse* que permitan determinar medidas de cambio y mejora basados en el análisis de la información y datos procesados, generando una visión lucida del estado actual de la entidad.

1.2. Descripción del documento

Se presenta una visión global del trabajo, mismo que plantea en cada capítulo. En el primer capítulo se cubrirán algunos conceptos introductorios, la presentación y descripción del documento. En el segundo capítulo constará el planteamiento de la propuesta de trabajo conformado por la información técnica básica, descripción del documento, preguntas básicas, formulación de metas, objetivos y la delimitación funcional. En el tercer capítulo se tratará acerca del estado del arte, algunos conceptos introductorios, inteligencia del negocio, definiciones, datos, información y conocimiento, ventajas de *Business Intelligence*, Arquitectura típica de soluciones *Business Intelligence*, *Data Mining* (minería de datos) con sus respectivas definiciones y etapas, *Data Warehouse* y definiciones, características y objetivos, Arquitectura de un *Data Warehouse*, Metodología de desarrollo del *Data Warehouse*, Metodología de Ralph Kimball. Metodología Hefestos, *SAS Rapid Data Warehouse Methodology*. En el cuarto capítulo está enfocado en el análisis de las metodologías, desarrollo de prototipos, resultados y la respectiva elevación hacia la construcción del *Data Warehouse* para el Instituto Nacional de Patrimonio Cultural Regional 3.

Capítulo 2

Planteamiento de la propuesta de trabajo

2.1. Información técnica básica

Tema: Análisis de metodologías para la implementación de un *Data Warehouse* aplicado a la toma de decisiones del Instituto Nacional de Patrimonio Cultural Regional 3

Tipo de trabajo: Proyecto de Investigación y Desarrollo

Clasificación técnica del trabajo: Aplicación

Líneas de Investigación, Innovación y Desarrollo

Principal: Sistemas de Información y/o Nuevas Tecnologías de la Información y Comunicación y sus aplicaciones

2.2. Descripción del problema

El problema se enfoca en la inadecuada manipulación de la información generada en el Instituto Nacional de Patrimonio Cultural Regional 3 (INPC R3), y en la toma de decisiones a destiempo debido a la intervención de cálculos manuales.

Actualmente se registra de forma manual los trámites del INPC R3, esta acción genera pérdida de tiempo, el procesamiento de la información es empírico sin respaldo de una herramienta tecnológica, lo cual no permite hacer un análisis de información óptima, lo cual dificulta la toma de decisiones y no da paso a la mejora continua del instituto.

Se hace necesaria la implementación de una solución tecnológica que permita automatizar el procesamiento de la información, reducir tiempos respuesta, automatizar cálculos manuales y prevenir inseguridades de información provocadas por la mala manipulación de los datos.

2.3. Preguntas básicas

¿Cómo aparece el problema que se pretende solucionar? Al no contar con una herramienta adecuada de análisis de información, no permite la toma de decisiones adecuadas acorde a la realidad y respaldada en información íntegra.

¿Por qué se origina? Al realizar una manipulación manual de información vital para el funcionamiento del INPC R3, se genera errores, tarda mucho tiempo en procesar y se obtiene poca integridad de la información.

¿Qué lo origina? El no poseer una herramienta tecnológica que procese información íntegra y genere reportes que ayuden a la toma de decisiones estratégicas.

2.4. Formulación de meta

Implementar un *Data Warehouse* para la toma de decisiones basadas en la información generada por el INPC R3.

2.5. Objetivos

Objetivo general.

Implementar un *Data Warehouse* para la toma de decisiones del INPC R3.

Objetivos Específicos.

1. Fundamentar aspectos teóricos de las metodologías de desarrollo de *Data Warehouse*.
2. Seleccionar la metodología que sea más apropiada para los datos del INPC R3 mediante la evaluación de parámetros de las metodologías de desarrollo *Data Warehouse*.
3. Implementar un *Data Warehouse opensource* con información íntegra de los procesos del INPC R3.
4. Generar reportes *Business Intelligence* detallados para una adecuada y oportuna toma de decisiones de la Dirección del INPC R3.

2.6. Delimitación funcional

Pregunta 1. ¿Qué será capaz de hacer el producto final del proyecto de titulación?

- Proporcionar una solución tecnológica para la toma de decisiones.
- Prestar información respaldada en un proceso seguro e íntegro por parte de la herramienta tecnológica.
- Generar reportes para una adecuada toma de decisiones de la Dirección Regional.

Pregunta 2. ¿Qué no será capaz de hacer el producto final del proyecto de titulación?

- No se podrá alojar en el internet para acceso público debido a políticas internas del INPC R3.
- No realizará actualizaciones automáticas de la información que contiene el *Data Warehouse* ante nuevos registros de trámites.

Capítulo 3

Marco Teórico

3.1. Introducción

Tradicionalmente se considera que el éxito o fracaso de cualquier política o plan estratégico aplicado a una empresa depende de que sus actores tengan una total comprensión del sistema al cual tratan de influenciar; pero las organizaciones y empresas financieras y comerciales han enfocado todos su esfuerzos a proveer sistemas de información transaccionales y a la generación de datos sobre todos los aspectos de sus actividades diarias, pero sólo una pequeña fracción de los datos capturados, procesados y almacenados en la empresa está realmente disponible para ejecutivos y tomadores de decisiones; es hasta la actualidad que las compañías han podido explotar esta información en la toma de decisiones, otorgando mayor rentabilidad a todas las acciones realizadas dentro de la empresa.

Debido al tamaño y la complejidad de los datos que los responsables de las decisiones deben manejar en diferentes empresas, es muy común que los efectos de esta toma de decisiones resulte inadecuados o contraproducentes cuando se trata de lograr productividad para la empresa; es por ello que el desarrollo de los métodos de análisis de datos, han permitido la optimización en la obtención de la información, toma de decisiones oportunas en el momento adecuado, lo que genera una minimización de costos, tiempo y recursos, este fenómeno se relaciona de forma directa con la inteligencia de negocio.

Las soluciones *Data Warehousing* han sido diseñadas para que evolucionen y se desarrollen en base a los requerimientos del negocio, donde la información varía de acuerdo a un período de tiempo; por el contrario de lo que nos ofrecen los sistemas operacionales.

Existen diferentes metodologías que permiten la implementación de almacenes de datos, para su posterior análisis actual e histórico, que permita a sus actores conocer la situación de una empresa en base a indicadores, con la finalidad de tomar una decisión acertada.

3.2. Inteligencia de negocios

En la actualidad son diversos los proyectos relacionados con la inteligencia de negocios, las organizaciones hoy en día, se esfuerzan por mejorar sus procesos y la toma de decisiones. *Business Intelligence* (BI) es un conjunto de actividades que la empresa emplea para comprender su situación actual, para la toma de decisiones estratégicas en base al análisis de datos históricos, no obstante, el término inteligencia de negocios no se asocia de ninguna manera con los almacenes de datos *Data Warehousing*.

3.2.1. Definiciones BI

La inteligencia de negocios apoya a la toma de decisiones del negocio: estratégicas, tácticas y operativas; según varios autores:

“Business Intelligence es el proceso de convertir datos en conocimiento, y el conocimiento en acción, para la toma de decisiones” (Bernabeu R. , 2010).

Desde un punto de vista pragmático, y asociándolo directamente con las tecnologías de la información, se define a *Business Intelligence* como “el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa de la compañía) en información estructurada, para su explotación directa (*reporting*, análisis OLTP/OLAP, alertas) o para su análisis y conversión en conocimiento dando así soporte a la toma de decisiones sobre el negocio” (Sinnexus, 2016).

Entre sus características destacan:

- Accesibilidad a la información: se debe garantizar el acceso de forma independiente de los usuarios sin tener en cuenta la procedencia de los datos.
- Apoyo a la toma de decisiones: permite la manipulación de la información por parte de los usuarios de manera que estos los conjuguen con las herramientas de análisis y les permitan realizar una toma de decisiones acertada
- Orientación al usuario final: se busca la independencia entre conocimientos técnicos de usuarios y su capacidad para utilizar estas herramientas.

3.2.2. Datos, información y conocimiento

La forma apropiada para diferenciar los términos; los datos están localizados en el mundo y el conocimiento se encuentra ubicado en agentes de cualquier tipo, mientras que la información adopta un papel mediador entre ambos.

Figura 1. Datos, información y conocimiento



Fuente: (Bernabeu D. , 2009)

Datos

Se consideran elementos primarios de la información cuyo valor por si solos no son de relevancia para la toma de decisiones, son la mínima unidad semántica; los datos pueden ser una colección de hechos almacenados en algún dispositivo; pueden ser internos o externos a la organización, objetivos o subjetivos o de tipo cualitativos y cuantitativos (Turnero, 2015).

Información

Un conjunto de datos procesados con un significado, un propósito y/o contexto; cuya utilidad reside en la toma de decisiones debido a la minimización de la incertidumbre, para la transformación en información de los datos se les añade valor (Computación, 2010):

- Contextualizar: contexto y propósito de su generación
- Categorizar: se conocen las unidades de medida que ayudan a interpretarlos.
- Calcular: procesamiento matemático o estadístico
- Corregir: depuración de los datos; errores e inconsistencias
- Condensar: resumen de datos

Información = Datos + Contexto (añadir valor) + Utilidad (disminuir la incertidumbre)

Conocimiento

El conocimiento es la combinación de experiencias, valores, información y “saber hacer” que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción; para que la información se transforme en conocimiento es necesario realizar (Vicente, 2017):

- Comparación con otros elementos

- Predicción de consecuencias
- Búsqueda de conexiones
- Conversación con otros portadores de conocimiento.

3.2.2.1. Reporteo

Desde los reportes hasta tableros de mando integral (*dashboards*) son las herramientas mediante las cuales se entrega la información de una empresa a los usuarios finales, estos reportes son recuperados desde un *Data Warehouse*, *Data Mart* u otro almacén de datos.

3.2.2.2. Procesamiento analítico en línea (OLAP)

Analiza de forma interactiva las transacciones del negocio, almacenadas en un *Data Warehouse*, *Data Mart* u otros almacenes de datos, algunas de las funciones OLAP incluyen las agregaciones resumidas (*ROLL-UP*), agregaciones detalladas (*DRILL-DOWN*), el corte y la rotación de cubos OLAP (*SLICE AND DICE*), entre otras. El análisis de las bases de datos se realiza según los procesos:

- ROLAP (*Relation Online Analytical Processing*) si se utiliza una base de datos relacional.
- MOLAP (*Multidimensional Online Analytical Processing*) si se utiliza una base de datos multidimensional.
- HOLAP (*Hybrid Online Analytical Processing*) se logra un tipo de análisis híbrido.

3.2.2.3. Minería de datos (*Data Mining*)

Consiste en la exploración y búsqueda de patrones y relaciones que describen los datos prediciendo sus valores futuros o inciertos, predicción de presupuestos, análisis de clientes potenciales, evaluación de análisis de tiempos de entrega a través de la comprensión de las situaciones del negocio actuales o históricas; se denomina análisis descriptivo cuando se encarga de explicar la situación actual o pasada, mientras que el análisis predictiva, se usa para realizar predicciones del futuro de la empresa (Uvidia, 2016).

3.2.3. Ventajas de *Business Intelligence*

“Los beneficios que se pueden obtener a través del uso de BI pueden ser de distintos tipos” (Cano, 2007).

Beneficios tangibles: reducción de costes y tiempos de respuesta, generación de ingresos.

Beneficios intangibles: acceso a la información por los diferentes actores del negocio

Beneficios estratégicos: apoya la formulación de estrategias para acceso y permanencia de la empresa en el mercado.

Algunos de los beneficios orientados a las organizaciones son (Bernabeu D. , 2009):

- Minimización del tiempo de recolección de la información, debido a que la información se encuentra integrada en una única fuente.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones
- Permite a los usuarios la no dependencia de informes escritos o programados, debido a la generación dinámica de los mismos
- Posibilita la formulación de preguntas y respuestas por parte de los usuarios que mejoren el rendimiento de la empresa
- Permite acceder y analizar directamente los indicadores de éxito.
- Se pueden detectar de manera rápida y oportuna los procesos fuera de lo esperado.
- Predicción de los resultados con un alto índice de certeza, esto basado en los datos históricos de la empresa
- Consulta de información, rápida, sencilla e intuitiva.

3.2.4. Desventajas de *Business Intelligence*

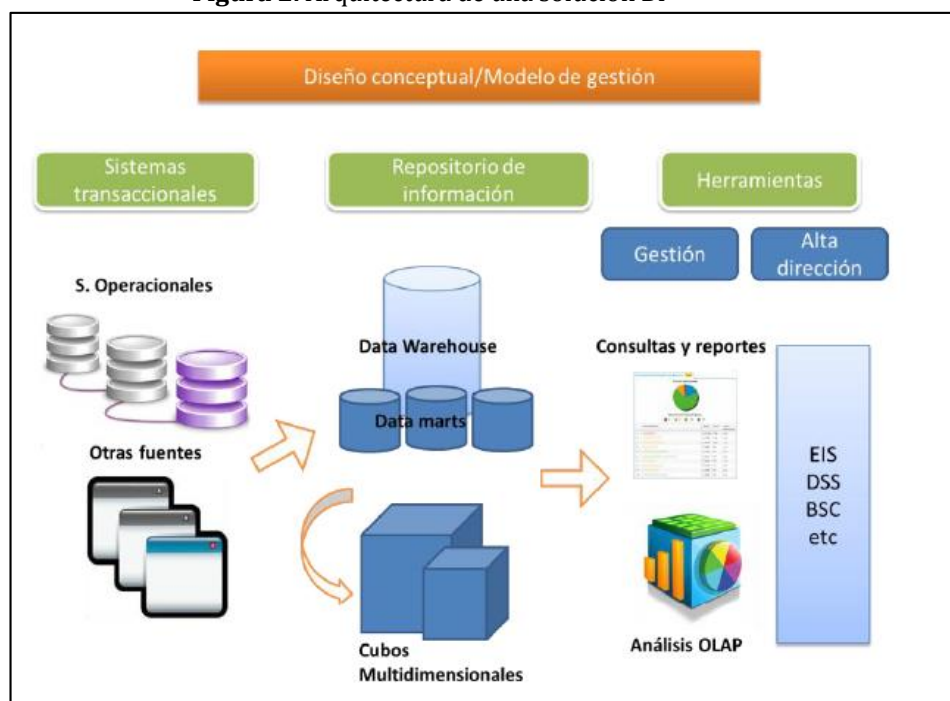
- Los costos se incrementan tanto en adquisición de recursos tecnológicos, así como en capacitación del personal operativo, para alcanzar los resultados esperados.
- Los materiales a emplearse en la implementación de un BI deben estar acorde a la evolución tecnológica, ya que los proveedores de hardware y software procuran la evolución rápida de sus productos, y el deterioro de los mismos es proporcional a su avance.
- La resistencia al cambio por parte de los usuarios es un factor determinante en la implementación de un BI.
- Los cambios en los ámbitos de comunicación e información son en apariencia rápido, pero las complicaciones presentadas en el proceso pueden retardar el proceso.
- Los beneficios de la implementación de un BI, no son apreciados en un periodo corto de tiempo, por el contrario, estos resultados se visualizan en un periodo largo de tiempo.

- Los datos confidenciales de la empresa: datos de clientes y proveedores quedan expuestos a todos los actores que intervienen en el BI
- Incremento continuo de los requerimientos de los usuarios
- Subestimación de las capacidades que brinda la correcta utilización del *Data Warehouse* y las herramientas de BI en general, al final no todos quedan conformes.

3.2.5. Arquitectura típica de soluciones *Business Intelligence*

Las tres capas básicas de este tipo de soluciones son: BDD Relacional, BDD Multidimensional y herramientas de visualización.

Figura 2. Arquitectura de una solución BI



Fuente: (Villareal, 2013)

3.2.5.1. Funciones

Tabla 1: Funciones de los componentes del *Business Intelligence*

Base de datos relacional	Base de datos multidimensional	Visualizador
Depuración y homogeneización Almacenamiento de datos Motos de cálculo, solo de ser necesario	Velocidad de acceso y consulta Capacidad de análisis desde varios puntos de vista (dimensiones de análisis)	Análisis de la información (<i>Drill-down, drill-through</i> y rotación de filas y columnas)

Fuente: Elaboración propia

3.3. *Data Mining* (minería de datos)

En las últimas dos décadas se ha producido un aumento dramático en la cantidad de información o datos almacenados en formato electrónico y se han generado nuevos métodos de procesamiento; es así que los sistemas de gestión de base de datos daban acceso a los datos almacenados, pero esta solo era una pequeña parte de lo que se podía hacer; los sistemas de procesamiento de transacciones en línea (OLTP), son buenos para poner los datos en las bases de forma rápida, segura y eficiente, pero no son buenos en la parte analítica.

El análisis de datos puede proveer más conocimiento sobre el negocio, es aquí donde *Data Mining* proporciona beneficios evidentes para cualquier empresa, la minería de datos es un proceso de descubrimiento, que permite a los usuarios comprender la importancia de los datos, así como la relación existente entre los mismos, a minería de datos descubre patrones y rasgos en el contenido de esta información (Uvidia, 2016).

3.3.1. Definiciones minería de datos

La minería de datos o descubrimiento del conocimiento en base de datos (KDD), como también se conoce, es la extracción no trivial de información implícita, previamente desconocida y potencialmente útil de los datos. Esto abarca una serie de enfoques técnicos diferentes, tales como agrupación, resumen de datos reglas de clasificación de dependencia, análisis de cambios y detección de anomalías (Frawley W & Piatetsky-Shapiro & Matheus, 2013).

La minería de datos es la búsqueda de relaciones y patrones globales que existen en grandes bases de datos para que estén “ocultos” entre la gran cantidad de datos, como una relación entre los datos del paciente y su diagnóstico médico. Estas relaciones representan valiosos conocimientos sobre la base de datos y los objetos de la base de datos y, si la base de datos es un fiel reflejo, del mundo real registrado por la base de datos (Knobbe, 2007).

Básicamente, la minería de datos se refiere al análisis de datos y al uso de técnicas de software para encontrar patrones y regularidades en conjuntos de datos. El análisis de *Data Mining* tiende a trabajar a partir de los datos y las mejores técnicas son aquellas, desarrolladas con una orientación hacia grandes volúmenes de datos, haciendo uso de la mayor cantidad posible de los datos recogidos para llegar a conclusiones y decisiones confiables.

El proceso de análisis comienza con un conjunto de datos, utiliza una metodología para desarrollar representación óptima de la estructura de los datos durante los cuales se adquiere el conocimiento. Una vez que el conocimiento se ha adquirido esto se puede extender a conjuntos más

grandes de datos trabando sobre la suposición de que el conjunto de datos tiene una estructura similar a la muestra de datos. Una vez más, esto es análogo a una operación de minería en la que se tamizan grandes cantidades de materiales de baja calidad para encontrar algo de valor.

3.3.2. Etapas de *Data Mining*

Las etapas comienzan con los datos en bruto y terminan con el conocimiento extraído, que fue adquirido como el resultado de las siguientes etapas:

- Selección: seleccionando o segmentando los datos de acuerdo con algunos criterios.
- Pre procesamiento: esta etapa de limpieza de datos en la que se elimina determinada información que se considera innecesaria y puede ralentizar las consultas. Además, los datos se reconfiguran para garantizar un formato coherente, y que existe la posibilidad de formatos inconsistentes porque los datos se extraen de varias fuentes.
- Transformación: los datos no sólo se transfieren, sino que se transforman; los datos se hacen utilizables y navegables.
- Minería de datos: esta etapa se refiere a la extracción de patrones a partir de los datos.
- Interpretación y evaluación: los patrones identificados por el sistema se interpretan en conocimiento que luego puede usarse para apoyar la toma de decisiones humana.

3.4. *Data Warehousing*

El desarrollo tecnológico existente en la actualidad, ha llevado a la creación de hardware más robusto, que permite soportar software más potente; que posea la capacidad de almacenar, procesar y analizar de forma analítica vastos volúmenes de información, algo que habría sido inaudito hasta hace cinco años.

De la misma manera el potencial de un *Data Warehouse* ha crecido y hoy por hoy se puede ampliar su funcionamiento si los datos apropiados son recopilados y almacenados. Se define un *Data Warehouse* como un sistema de gestión de base de datos relacional (RDBMS) diseñado específicamente para satisfacer las necesidades de sistemas de procesamiento transaccional; por lo tanto, es como cualquier repositorio de datos centralizado, que puede ser consultado para el beneficio del negocio.

La implantación de estos almacenes de datos se han generado diferentes técnicas de gran alcance, que permiten extraer datos operativos desde diferentes formatos, así como la integración de los datos de una empresa, independientemente de su ubicación, formato o requisitos de comunicación;

pudiendo incorporar a la misma información adicional, mediante la adición de métodos de acceso (consulta en línea, generación de informes), se puede convertir un almacén de datos “inactivo” en una fuente de información “dinámica”; a esto lo denominamos *Data Warehousing*.

3.4.1. Definiciones *Data Warehousing*

Algunas de las definiciones del *Data Warehousing* se describen a continuación:

“Un *Data Warehousing* es un almacén único, completo y consistente de datos obtenidos de una variedad de fuentes y puestos a disposición de los usuarios finales de una manera que puedan entender y utilizar en un contexto empresarial” (Cuellar, 2011).

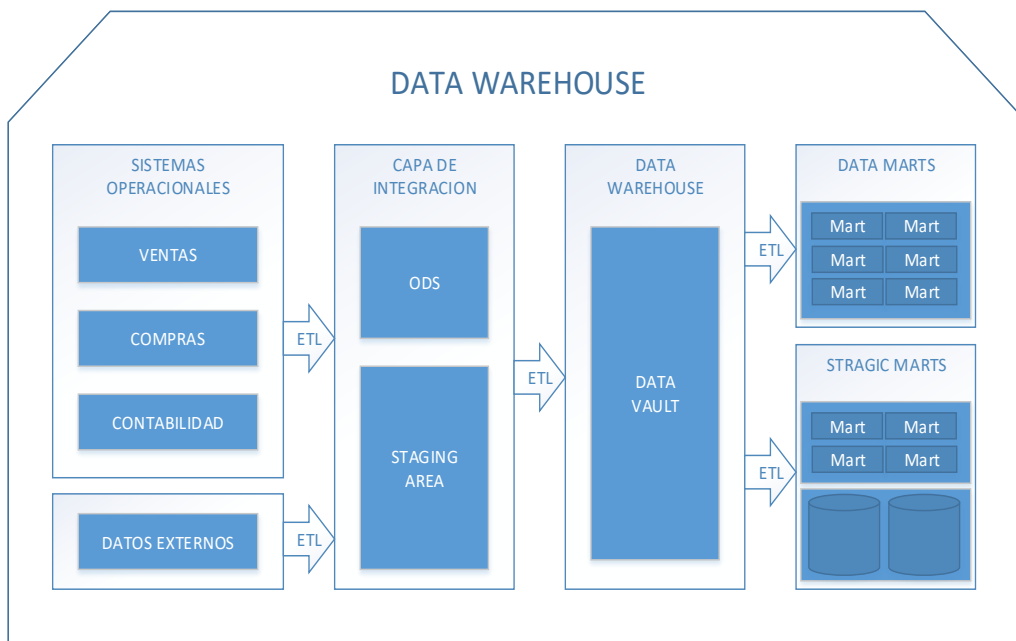
“El *Data Warehousing* es un proceso, no un producto, para ensamblar y administrar datos de diversas fuentes con el fin de obtener una visión única y detallada de una parte o de toda una empresa” (Devlin, 2011).

De lo mencionado por los autores citados anteriormente se puede concluir que el *Data Warehousing* es el proceso mediante el cual los datos son transformados y agrupados dentro de un almacén de datos., permitiendo una fusión de los mismos los cuales provienen de diversas fuentes hacia un solo contexto, lo cual permite un acceso rápido a la información y por consiguiente a la toma de decisiones

3.4.2. *Data Warehouse*

En el contexto informático, *Data Warehouse* en español almacén de datos es una colección de información corporativa, derivada directamente del sistema operativo y algunas fuentes externas de datos, su propósito específico es apoyar las decisiones empresariales. Se trata de un expediente completo de una empresa, más allá de la información transaccional y operacional, almacenada en una base de datos, diseñada para favorecer el análisis y la divulgación eficiente de datos (Peña J & Suárez, 2008).

Figura 3. Arquitectura de un *Data Warehouse*



Fuente: (Villareal, 2013)

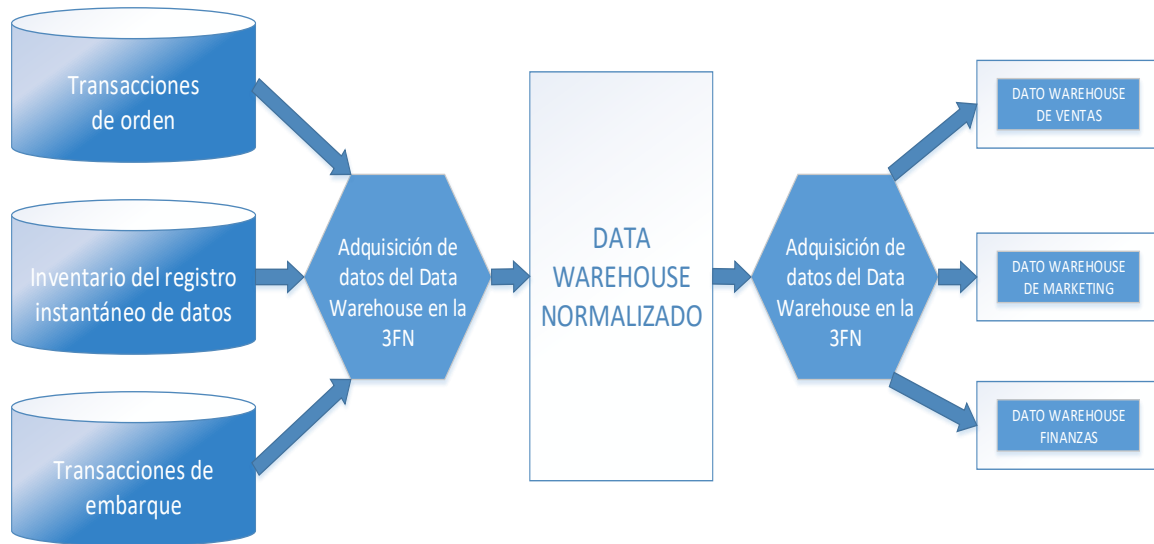
3.4.3. Definiciones

3.4.3.1. Definición de *Bill Inmon*

Bill Inmon fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, para él es necesaria la transferencia de la información de los OLTP de la empresa a un solo lugar, centralizando los datos para su análisis (*CIF o Corporate Information Factory*) el cual debe cumplir con las características de un repositorio de datos:

- Orientado a temas. - Los datos y elementos concernientes a un mismo tema, deben estar organizados de tal manera que ellos queden unidos entre si dentro del almacén de datos
- Variante en el tiempo. - Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar, reflejen esas variaciones.
- No volátil. - La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- Integrado. - La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

Figura 4. Arquitectura de un *Data Warehouse* según Bill Inmon



Fuente: (Espinosa, 2010)

Inmon defiende una metodología descendente (*top-down*) a la hora de diseñar un almacén de datos, ya que de esta forma se considerarán mejor todos los datos corporativos. En esta metodología los *Data Marts* se crearán después de haber terminado el *Data Warehouse* completo de la organización (Kimball, 2016).

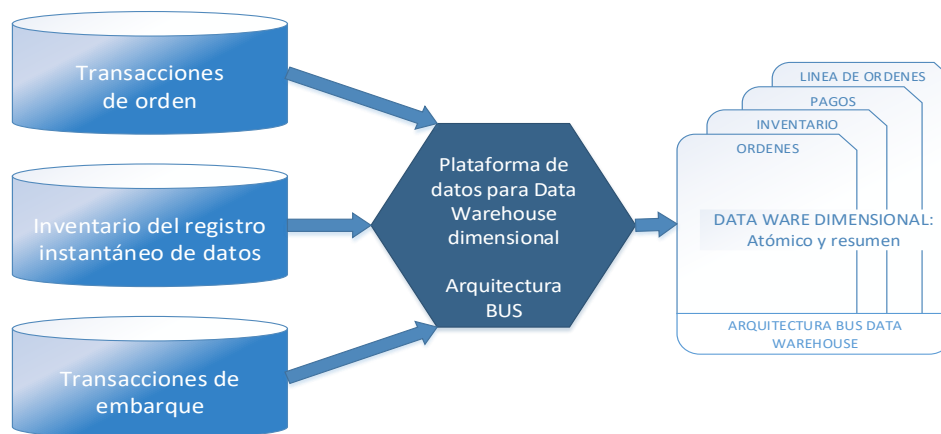
3.4.3.2. Definición de *Ralph Kimball*

Para este autor, un *Data Warehouse* es como:

Un *Data Warehouse* es un conjunto de todos los *Data Marts* que representan la información de una empresa, siendo "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis", de acuerdo al modelo dimensional (no normalizado); la representación de las dimensiones y hechos que son más que las tablas de un diseño relacional de origen.

Cada *Data Mart* se encuentra conectado, formando una estructura de bus, que se comunica entre sí a través de elementos comunes, esto permite una estructura flexible y sencilla de implementar; pues se puede construir un *Data Mart* como un primer elemento del proyecto y luego agregar nuevos *Data Marts* que compartan las mismas dimensiones e incluyan otras.

Figura 5. Arquitectura de un *Data Warehouse* según Ralph Kimball



Fuente: (Espinosa, 2010)

La carga de datos desde su fuente origen se realiza mediante los procesos de extracción, transformación y carga (ETL) los cuales extraen los datos y los procesan en las *stage area*, para luego realizar el proceso de llenado de cada uno de los *Data Mart*.

3.4.3.3. Definición de Bernabeu

“El *Data Warehouse* posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogeneización de los datos de toda la empresa, provee información que ha sido transformada y sumariada, para que ayude en el proceso de toma de decisiones estratégicas y tácticas” (Bernabeu R. , 2010), por ende, el *Data Warehouse* tiene la finalidad de hacer que la información esté integrada y accesible ayudado en la toma de decisiones dentro de una organización.

Entre las características de un *Data Warehouse* según Bernabeu están:

- Orientado a un tema
- Administra grandes cantidades de información
- Guarda información en distintos repositorios
- Condensa y agrega información
- Integra y asocia información
- Ayuda en la decisión estratégica
- Permite explotar la información histórica existente

Las ventajas del *Data Warehouse* según Bernabeu son:

- Posibilita la extracción de datos de sistemas operacionales y fuentes externas.

- Permite la integración y homogeneización de los datos de toda la empresa
- Provee información que ha sido transformada y sumariada, para que ayude en el proceso de toma de decisiones estratégicas y tácticas.

3.4.3.4. Una definición más amplia de almacén de datos

Un *Data Warehouse* proporciona una visión global, común e integrada de los datos de la empresa, independientemente de cómo estos se integren posteriormente; tal como las definiciones anteriores mencionan un *Data Warehouse* se centra en los datos en sí mismo; pero los procesos mediante los cuales se extrae, se cargan y se transforma esta información son componentes esenciales de un almacén de datos.

Es necesario establecer que para esta definición las herramientas para el análisis de la información y toma de decisiones, así como aquellas para la extracción, transformación y carga de datos son una de conceptos que se enmarcan dentro la definición de *Data Warehouse* (Lozada X & Cruz, 2014)

3.4.4. Características y objetivos

Las características que definen un *Data Warehouse* están definidas por:

Orientada al negocio

La primera característica del *Data Warehouse* establece que la información debe ser clasificada de acuerdo a los intereses que la empresa. Esta clasificación determina el diseño y la forma en que los datos serán implementados dentro del almacén.

Integrada

Considerando que la fuente de los datos de una empresa es proporcionada desde diferentes departamentos, secciones y aplicaciones, tanto internas como externas; la integración es la consolidación de la información en una sola instancia antes de ser agregadas al almacén de datos, mediante un proceso que asegure la calidad de los datos, este proceso cuenta con diversas técnicas y subprocesos.

Variante en el tiempo

El volumen de datos que se manejan en un *Data Warehouse*, es directamente proporcional al tiempo de proceso y consulta solicitados. Este espacio de tiempo producido desde el momento de la

solicitud de búsqueda hasta su devolución en del todo normal en estos ambientes y es por ello que la información que se encuentra dentro del almacén de datos de denomina de tiempo variante (Moine J Gordillo S & Haedo, 2011).

No volátil

Para la toma de decisiones es necesario que la información proporcionada sea estable, cuando los datos provienen de la parte operacional de la empresa estos son variantes en el tiempo, por el contrario, en un *Data Warehouse* estos datos no se modifican.

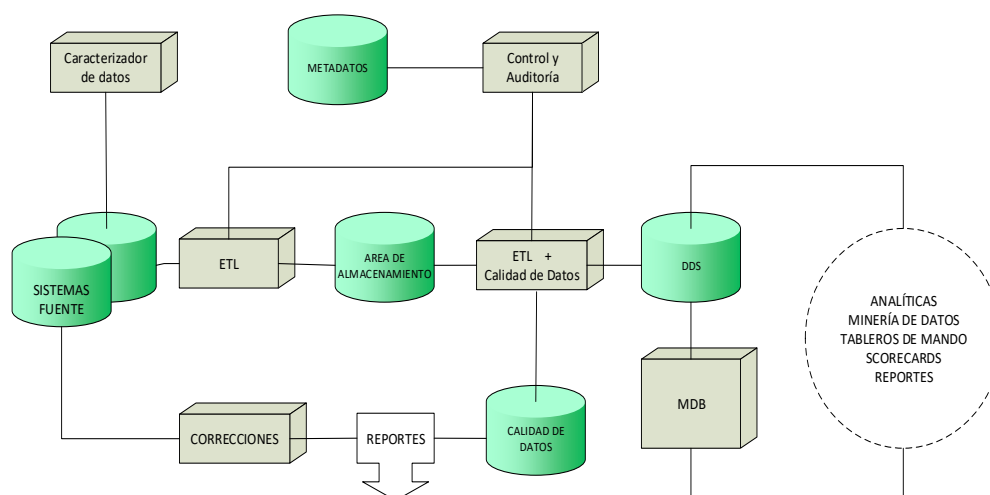
3.4.5. Arquitectura de un *Data Warehouse*

Es necesario considerar a un *Data Warehouse* como un repositorio donde se homogeneiza y conjuga todos los datos de una empresa en una única estructura de donde todos sus actores pueden obtener una visión única de la información; es así que un *Data Warehouse* se encuentra constituido por dos arquitecturas:

3.4.5.1. Arquitectura de flujo de datos

Se define a la arquitectura de flujo de datos como un proceso mediante el cual los datos fluyen a través de los almacenes de datos hasta llegar al usuario final define esta arquitectura; la arquitectura de flujo de datos determina cuales son los componentes a construir dentro del *Data Warehouse*, cumpliendo con requerimientos y calidad del proyecto preestablecidos por los usuarios; esta es una actividad que afecta la planificación de proyecto, los recursos a emplearse y costos que genera (Díaz, 2010).

Figura 6. Arquitectura de flujo de datos



Fuente: elaboración propia

1. Los sistemas de origen integrados por los sistemas OLTP (*Online Transaction Processing*), encargados de la recuperación y el almacenamiento las transacciones generadas por el negocio, estos pueden ser archivos planos u otras fuentes de datos digitalizadas.
2. El caracterizador de datos es una herramienta que tiene la capacidad de analizar los datos; encontrando el número de filas existentes en una tabla, el número de columnas con valores nulos y otras formas que caracterizan a los datos.
3. Los procesos ETL se conectan a los sistemas de origen, para extraer, transformar y cargar los datos en un sistema destino, la cual puede ser un área de almacenamiento o un almacén de datos normalizado o dimensional.
4. Servicio de distribución de datos (DDS) almacena datos integrados de varias fuentes, este almacén de datos dimensional es una base de datos del *Data Warehouse* similar a un sistema OLTP dónde los datos están organizados en un formato dimensional que permiten realizar consultas de análisis.
5. Las reglas de calidad de datos son procesos que se realizan a manera de verificación de la calidad de datos extraídos desde las fuentes de origen. Si los datos no cumplen con estas reglas son recolectados en una base de datos de calidad, los cuales deben ser corregidos para luego ser cargados, este proceso se repite hasta que cumplan ciertos límites de tolerancia.
6. El componente de auditoría y control registra las operaciones del sistema y su uso dentro de la base de datos. Este componente ayuda a comprender los sucesos que ocurrieron durante los procesos ETL.
7. La base de datos multidimensional (MDB) conocida como cubos, es una forma de base de datos dónde los datos se almacenan en celdas y la posición de cada celda está definida por un número de variables llamadas dimensiones. Cada celda del cubo representa un evento del negocio y los valores de las dimensiones muestran ¿cuándo? y ¿Dónde? sucedió un evento.
8. Interfaz de usuario son aplicaciones de inteligencia de negocio que consultan a los almacenes de datos dimensionales (DDS) y a las bases de datos multidimensionales (MDB) mediante: reportes, tableros, analíticas, minería de datos, *scorecards* u otras aplicaciones.

Según su finalidad, los almacenes de datos pueden ser de tres tipos:

- De interfaz de usuario

Los almacenes de datos están disponibles para ser consultados por los usuarios y aplicaciones finales.

- Interno

Con el propósito de integración, limpieza y preparación de datos, los componentes del *Data Warehouse* usan un almacén de datos interno. Estos almacenes no se encuentran disponibles para el usuario ni por aplicaciones finales.

- Híbrido

Un almacén de datos que puede ser usado tanto por mecanismos del sistema *Data Warehouse* como por los usuarios y las aplicaciones finales.

Con respecto al formato de los datos, los almacenes de datos pueden de tipo:

- Áreas de almacenamiento (*stage area*)

Es un almacén de datos interno usado para transformar y preparar los datos obtenidos desde los sistemas fuentes

- Almacén de datos normalizados (NDS, *normalized data store*)

Es un almacén de datos interno que simula una o varias bases de datos relaciones y normalizadas donde se integran los datos obtenidos desde sistemas de origen, se incluyen datos históricos.

- Almacén de datos operacional (ODS, *operational data store*)

Es un almacén interno de composición híbrida puesto que combina una o varias bases de datos relacionales y normalizados con datos transaccionales y las versiones más recientes de los datos maestros. Este tipo de almacenes no posee datos históricos, solamente la versión más reciente de datos.

- Almacén de datos dimensional (DDS, *dimensional data store*)

Es un almacén de datos de interfaz, donde los datos están organizados en un formato dimensional y pueden soportar consultas analíticas.

Las arquitecturas del flujo de datos:

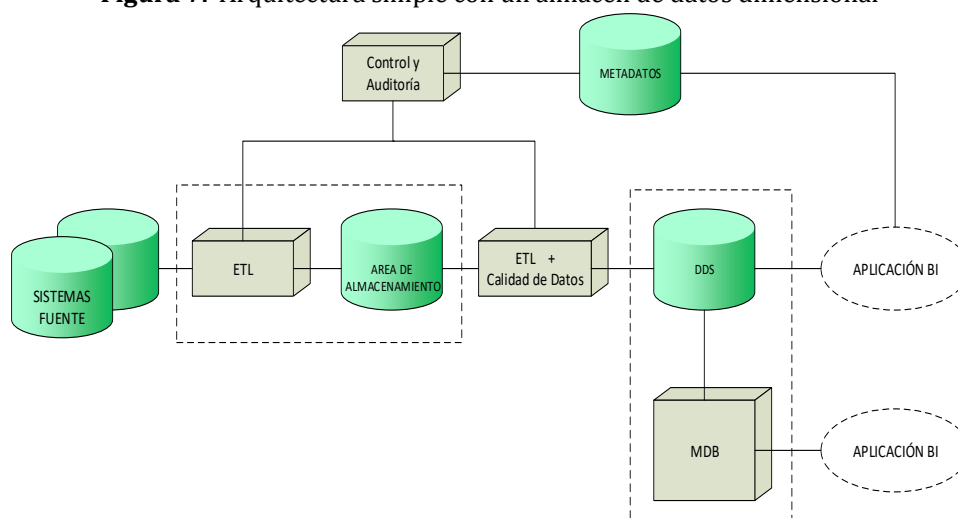
- Almacén de datos dimensional simple (Simple DDS)
- Almacén de datos normalizado más almacén de datos dimensional (NDS+DDS)
- Almacén de datos operacional más almacén de datos dimensional (ODS+DDS)
- *Data Warehouse* Federado (FDW, *Federated Data Warehouse*)

3.4.5.1.1. Arquitectura simple con un almacén de datos dimensional (*Simple DDS*)

Esta arquitectura de datos simple, ha sido diseñada para ambientes pequeños donde los datos provienen de una sola fuente de datos, y no necesitan ser cargados en un almacén de datos normalizados, son trasladados desde un almacén temporal o *stage* área hacia un almacén de datos dimensional (DDS), durante esta transferencia el proceso de ETL aplica las reglas de calidad a los datos y mientras recolecta datos en el paquete de control y auditoría.

Esta arquitectura se relaciona directamente con la propuesta de Kimball donde se establece que para lograr una arquitectura de inteligencia de negocios multidimensional (MD) cada DDS debe tener un formato de datos dimensional y que no es necesario que sus datos provengan de una base normalizada.

Figura 7. Arquitectura simple con un almacén de datos dimensional



Fuente: elaboración propia

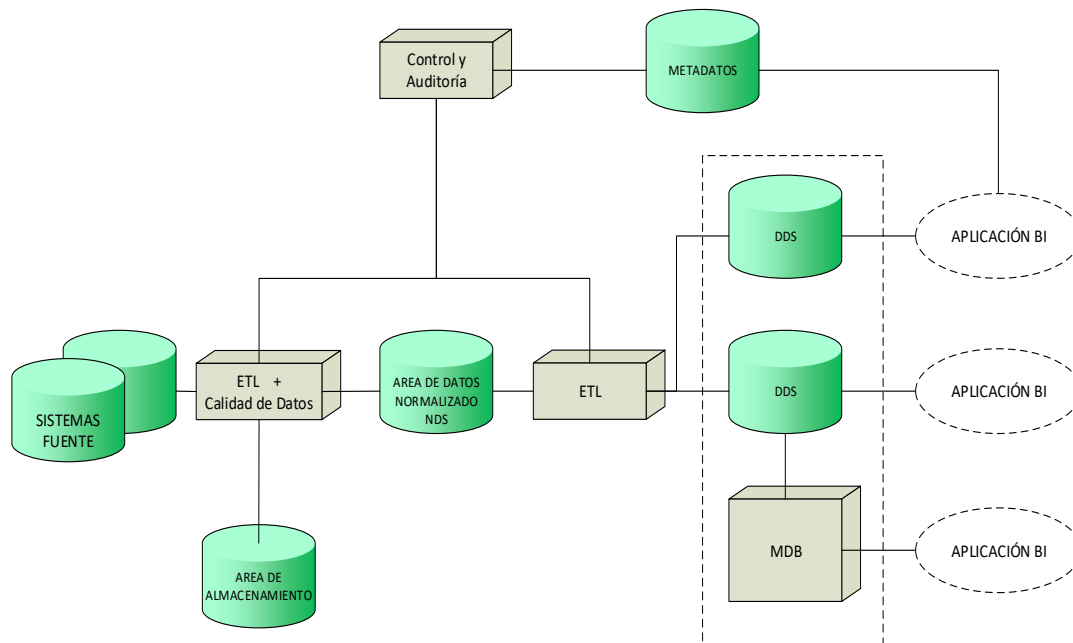
3.4.5.1.2. Arquitectura con un almacén de datos normalizado más un almacén de datos dimensional (NDS-DDS)

En esta arquitectura se establece la necesidad de crear un almacén de datos normalizado (NDS) y almacenes de datos dimensionales (DDS) como derivados del mismo. Esta propuesta coincide con la de Inmon y su fábrica de información corporativa (CIF) en la que se establece la construcción de un *Data Warehouse* normalizado y a partir de esto se derivan los *Data Marts*, tantos como almacenes de datos dimensionales se requieran.

Los procesos ETL usados para extraer los datos de los sistemas fuente NDS aplican reglas de calidad de datos, los datos que no cumplen las reglas aplicadas son recolectados en un área de

almacenamiento temporal con el propósito de cargar datos de calidad en el NDS; existe en esta arquitectura un componente de control y de auditoria y almacenamiento de metadatos.

Figura 8. Arquitectura con un almacén de datos normalizado más un almacén de datos dimensional



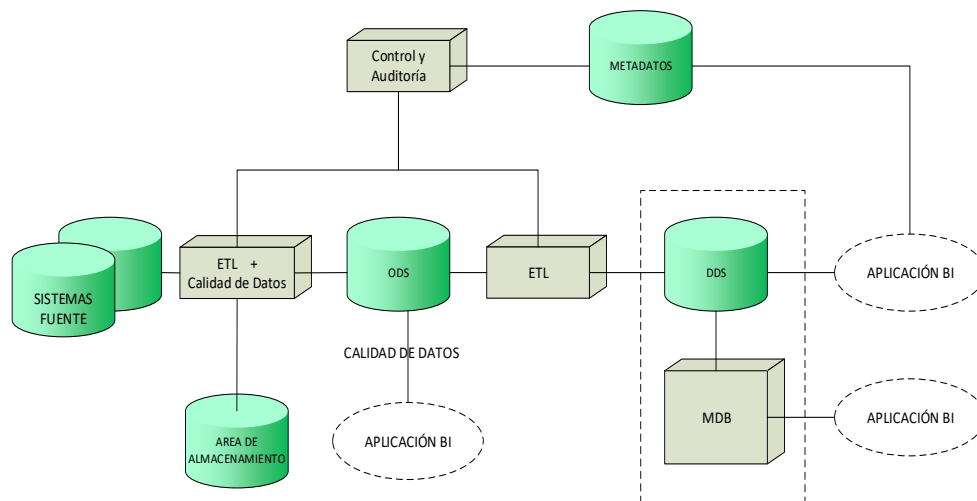
Fuente: elaboración propia

3.4.5.1.3. Arquitectura con un almacén de datos operacional más un almacén de datos dimensional (ODS-DDS)

Esta arquitectura propone la construcción almacenes de datos operacionales (ODS), creando así *Data Marts* operacionales y posteriormente la construcción de almacenes dimensionales; estos contienen únicamente datos recientes, no datos históricos (Gomez, 2010).

“Al igual que en los NDS+DDS, esta arquitectura se relaciona con la propuesta de Immon y su CIF; en esta arquitectura el almacén de datos operacionales contiene datos limpios y de calidad, puedes ser consumidos por las aplicaciones operacionales como sistemas OLTP, reportes operacionales o CRMs (*Customer Relationship Managment*)” (Clark, 2004).

Figura 9. Arquitectura con un almacén de datos operacional más un almacén de datos dimensional



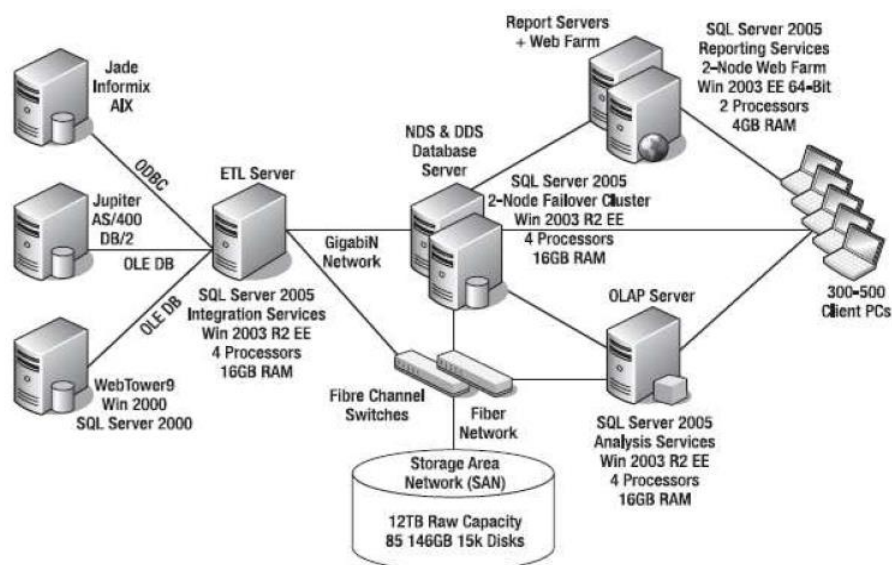
Fuente: elaboración propia

3.4.5.2. Arquitectura del sistema

La infraestructura física mediante la cual se integra un *Data Warehouse* contiene tres elementos:

1. Hardware: Los servidores que se usan para el *Data Warehouse*, pueden ser de base de datos, aplicaciones, ETL, etc.
2. Redes: constituyen el medio de comunicación entre los servidores.
3. Almacenamiento: Discos duros que se usarán, sus arreglos, configuración y tecnologías, así como las soluciones de respaldos que se implementarán.

Figura 10. Arquitectura física de un *Data Warehouse*



Fuente: (Murillo, 2008)

3.4.6. Herramientas de *Data Warehouse*

El sin número de herramientas proporcionadas en la actualidad se ha dividido en las siguientes categorías.

3.4.6.1. Herramientas de almacenamiento

Los procesamientos analíticos en línea (OLAP) proveen a los usuarios respuestas a preguntas ¿Qué sucedió? mediante sus capacidades de “*Slide and Dice*”; esto debido a las limitaciones en la generación de reportes, cuando sus usuarios necesitan más que una vista estática de la información.

3.4.6.2. Herramientas de extracción, transformación y carga

Los datos son extraídos hacia el *Data Warehouse*, esto como el primer paso para la obtención de datos y su transformación en información útil para el usuario; cuando los datos son extraídos es necesario realizar procesos de limpieza, para la selección de aquellos significativos y necesarios para el *Data Warehouse*; combinación de fuentes de datos, y acoplamiento correcto de la información; al finalizar la transformación, los datos son cargados (García, 2012)

3.4.6.3. Herramientas de análisis *intelligence*

Las herramientas usadas para el análisis de información han sido construidas usando inteligencia artificial, las cuales emplean técnicas de *Data Mining*.

3.5. Estado Arte

El análisis de las metodologías para el desarrollo de un *Data Warehouse* conlleva estudio del conjunto de procedimientos que se siguen para construir un sistema, un ciclo de vida de desarrollo de sistema (SDLC, *Systems Development Life Cycle*). Entre las metodologías se analizarán las siguientes.

3.5.1. Metodología Ralph Kimball

Introducción

La metodología de Kimball, llamada Modelo Dimensional (*Dimensional Modeling*), se basa en lo que se denomina Ciclo de Vida Dimensional del Negocio (*Business Dimensional Lifecycle*). Esta metodología es considerada una de las técnicas favoritas a la hora de construir un *Data Warehouse*.

Un almacén de datos (*Data Warehouse*) es una colección o conglomerado de datos o *Data Marts* orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Es una estructura de

datos donde la información contenida está diseñada para favorecer el análisis y la divulgación eficiente de datos.

En el Modelo Dimensional se constituyen modelos de tablas y relaciones con el propósito de optimizar la toma de decisiones, con base en las consultas hechas en una base de datos relacional que están ligadas con la medición o un conjunto de mediciones de los resultados de los procesos de negocio.

El Modelo Dimensional es una técnica de diseño lógico que tiene como objetivo presentar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento (León, Ciclo de vida de Ralph Kimball, 2014).

3.5.1.1. Características

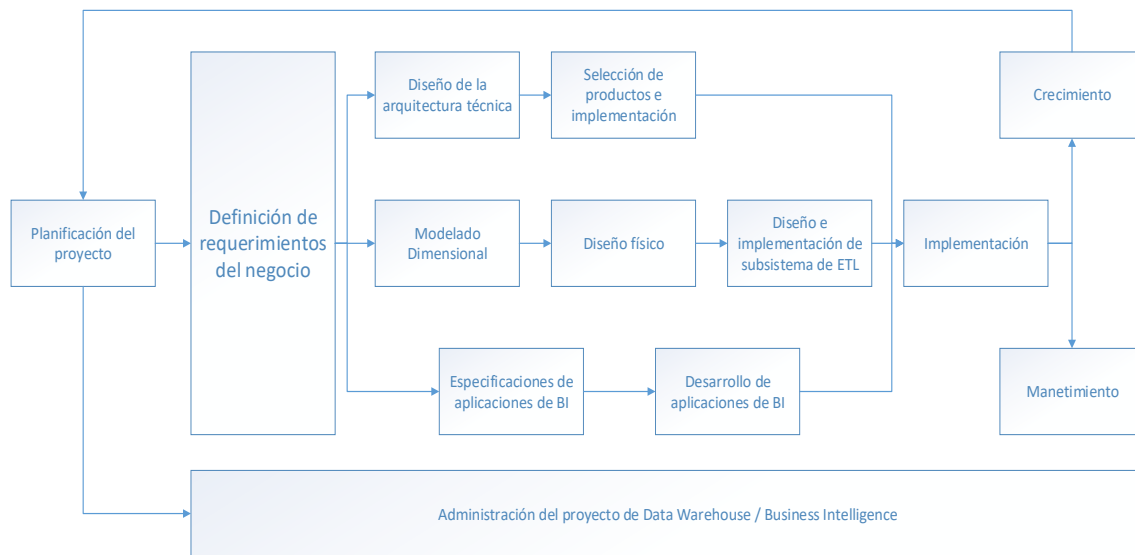
Este ciclo de vida del proyecto de DW, está basado en cuatro principios básicos:

- Centrarse en el negocio: Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- Construir una infraestructura de información adecuada: Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- Realizar entregas en incrementos significativos: crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de *software*.
- Ofrecer la solución completa: proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

3.5.1.2. Fases

Las fases establecidas por Ralph Kimball, han sido diseñadas para que puedan ser desarrolladas en paralelo o en forma secuencial; cada una de las fases planteadas en esta metodología garantiza la calidad en el desarrollo del *Data Warehouse*

Figura 11. Fases Metodología Ralph Kimball



Fuente: (Murillo, 2008)

1. Planificación del Proyecto.

En este proceso se determina el propósito del proyecto *data warehouse*, así como sus objetivos específicos y el alcance del mismo, los riesgos principales y una aproximación inicial a las necesidades de la información.

Definición del Proyecto

La identificación del escenario permite determinar la existencia de demanda para el proyecto y su alcance; los cuales podrían darse de la siguiente manera:

- Demanda de un sector del negocio: donde un ejecutivo necesita acceso a información relevante; caso ideal, con objetivos definidos y un alcance determinado.
- Demasiada demanda de información: donde múltiples ejecutivos necesitan acceso a información relevante; caso riesgoso, puesto que los requerimientos provienen de diferentes áreas de la empresa, maximizando costos y recursos.
- En busca de demanda: donde no se identifican las necesidades del almacén de datos, pero se desea incorporar diversos requerimientos; caso abierto, se establece los requerimientos y se desarrolla la solución en base a estos.

- Luego de la identificación del escenario es necesario la determinar sponsors o patrocinadores internos involucrados en la demanda.

Determinación de la preparación de la empresa

Existen cinco factores a tener en cuenta antes de iniciar un proyecto de *Data Warehouse*

Figura 12. Factores de determinación de un proyecto



Fuente: elaboración propia

- Patrocinio de la gerencia en el negocio: es necesario identificar más de un patrocinador, para generar una mejor estabilidad en el desarrollo del proyecto; los gerentes son sponsors ideales, puesto que determinarán el apoyo y soporte de los demás miembros de la empresa.
- Motivación del negocio: se busca encontrar un sentido para el desarrollo del proyecto, entre los motivadores que influyen se encuentran la competencia y la visión competitiva y hasta una crisis empresarial.
- Acompañamiento del departamento de tecnología y de negocio: el apoyo del departamento tecnológico ayudará en con futuras responsabilidades.
- Presencia de cultura analítica: donde la actual toma de decisiones no se base en la intuición, sino en los resultados obtenidos del negocio.
- Factibilidad: análisis tanto de la infraestructura como de los datos a cargarse.

Desarrollo del enfoque preliminar

La definición del enfoque inicial, debe estar asociada directamente con los requerimientos del negocio, todo esto debe ser documentado, para ello es importante seguir los parámetros:

- La definición del enfoque es realizada por el departamento de tecnología y de negocio, donde se establecen requerimientos específicos y tiempos de entrega.
- El enfoque inicial debe ser factible y manejable, así como sus demás iteraciones.
- Enfoque inicial en un solo requerimiento del negocio soportado por una sola fuente de origen de datos.

- Límite de usuarios de acceso para la primera entrega
- Establecer criterios de éxito, para establecer exactamente lo requerido por los sponsors

Desarrollar la justificación del negocio

Determinación del factor retorno de la inversión (ROI), la comparación del retorno financiero esperado (beneficios del negocio) y la inversión esperada (costos) para ellos se considera:

Tabla 2: Factores de retorno de la inversión (ROI)

COSTOS	RETORNOS Y BENEFICIOS
• Compras de licencias • Mantenimiento (soporte proyecto y capacitación usuarios) • Recursos internos • Recursos externos • Costos de capacitación • Costos de crecimiento.	• Incremento de ventas • Incremento de ganancias • Incremento de niveles de calidad • Exploración de nuevas oportunidades

Fuente: elaboración propia

Planeación del proyecto

La planificación inicial del proyecto comprende la preparación y asignación de roles a desempeñarse dentro de la ejecución del proyecto:

- Patrocinadores
- Gerente del proyecto
- Líder de negocios del proyecto
- Analista del sistema de negocios
- Modelador de datos
- Diseñador de proceso
- Desarrolladores de aplicación
- Educador de proyecto

Desarrollo del plan del proyecto

Se debe proporcionar suficiente detalle para el seguimiento del proyecto, determinando actividades, recursos y tiempos para el desarrollo.

Administración del proyecto

Establecer agendas de reuniones para la comunicación y monitoreo de los avances del proyecto

2. Definición de requerimientos

Acercamiento a la definición de requerimientos

Se debe tratar con los usuarios sobre su trabajo, objetivos retos e intentar establecer el proceso para la toma de decisiones, actual y futura; no se debe preguntar sobre los datos a obtener.

Preparación de la entrevista

Se deben determinar los roles para la ejecución de la entrevista; establecer la persona entrevistadora y un secretario, los cuales están encargados de obtener el mayor detalle posible de las necesidades de la empresa.

Investigación previa a entrevistas

Análisis previo de la empresa, su visión y misión, así como sus fortalezas y debilidades frente a sus competidores; revisión de los reportes anuales de la empresa, para determinar estrategias en la toma de decisiones.

Selección de entrevistados

El organigrama de la empresa, ayudará a determinar a los ejecutivos que permitan comprender las estrategias del negocio, y a los analistas quienes conocen el manejo de los datos y a información.

Desarrollo de cuestionario

Es necesario el desarrollo de los cuestionarios de acuerdo al nivel jerárquico de los entrevistados, así se logrará un nivel de detalle alto que determinará la granularidad del *Data Warehouse*.

- **Inicio y desarrollo de la entrevista**

Es necesario establecer los objetivos de la entrevista, de cada uno de los entrevistados se debe obtener una visión amplia de sus departamentos; se debe entender los campos de la base, granularidad, volumen de datos y otros detalles de estas fuentes de información. Por último, se deben establecer criterios, actitudes y expectativas sobre el proyecto.

- **Análisis de las entrevistas**

Es importante analizar y repasar la documentación adquirida durante el transcurso de las entrevistas, la cual nos proporcionará una aproximación de las dimensiones del modelo y la generación de un documento con los requerimientos.

3. Modelado dimensional

Modelo entidad relación

Es una técnica de diseño lógico para el procesamiento de transacciones; que permiten el almacenamiento de grandes cantidades de datos mediante el proceso de normalización; permitiendo insertar, adicionar, borrar o modificar sobre la base de datos desde un solo lugar y no en múltiples, eliminando así la redundancia.

Sin embargo, este tipo de modelado presenta algunos inconvenientes con respecto a los usuarios y el manejo técnico de un modelo relacional, considerando que ellos no pueden navegar sobre el modelo, incumpliendo de esta manera uno de los principales objetivos de un *Data Warehouse*, el cual establece que se deben proporcionar datos de forma intuitiva minimizando los tiempos de respuesta.

Modelo dimensional

Es una técnica de diseño lógico que presenta los datos para una manipulación intuitiva de alto desempeño. Para el diseño dimensional es posible emplear los modelos de estrella y copo de nieve, donde el modelo estrella tiene una tabla de hechos y en ella se integran llaves foráneas de cada una de las dimensiones relacionadas con esta tabla; mientras que para el modelo en copo de nieve los campos de baja cardinalidad de la dimensión han sido incorporados a tablas separas y enlazados a la principal a través de llaves foráneas ; estos modelos están conformados por: tablas de hechos y dimensiones, llaves foráneas.

Diseño de dimensiones y hechos

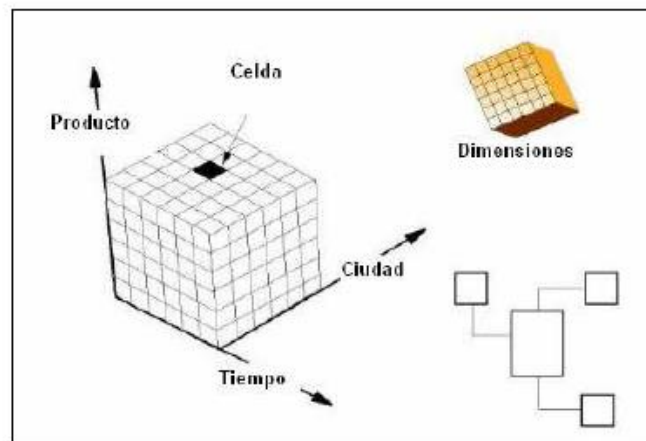
Dada una compleja estructura de un *Data Warehouse*, es necesario establecer en su mayoría las mismas tablas de dimensiones, para la creación de una arquitectura en bus de *Data Marts*; estas tablas deben cumplir con las siguientes condiciones

1. Una tabla de dimensión puede ser usada con cualquier tabla de hechos dentro de una misma base

2. Las interfaces de usuario y contenido de datos son consistentes para cualquier uso de la dimensión
3. La interpretación de atributos es consistente de la tala en cualquier *Data Mart*.

Para el diseño de tablas de hechos y dimensiones se debe tener en cuenta que cada uno de los datos de la empresa deben ser representados en forma de cubo, en esta representación los bordes del cubo son expresadas por las dimensiones, mientras que en las celdas encontramos los datos.

Figura 13. Ejemplo cubo dimensional Kimball



Fuente: (Peña J & Suárez, 2008)

- **Hechos**

Los hechos son aquellos valores que varían durante el desarrollo del negocio, por ejemplo, en un proceso de facturación, se considera un hecho la factura correspondiente a un producto generada a un cliente, y que fue emitida desde una sucursal, por lo tanto, podemos intuir que cada valor numérico y de tipo decimal es un posible hecho.

- **Dimensiones**

Una dimensión es un atributo que describe u organiza los hechos, estos atributos son una fuente para encabezados o restricciones que serán usados en reportes finales.

Cuando se agregan atributos a un reporte final o a una búsqueda para obtener mayor detalle se dice que es un *drilling down*.

- **Llaves subrogadas**

Todas las claves de una tabla debe ser una llave subrogada, estas deben ser diferentes a la de las tablas de las que se originaron, generalmente son de tipo numérico.

Método de diseño de una tabla de hechos

Para Ralph Kimball es necesario cumplir con los siguientes pasos para conseguir un modelado

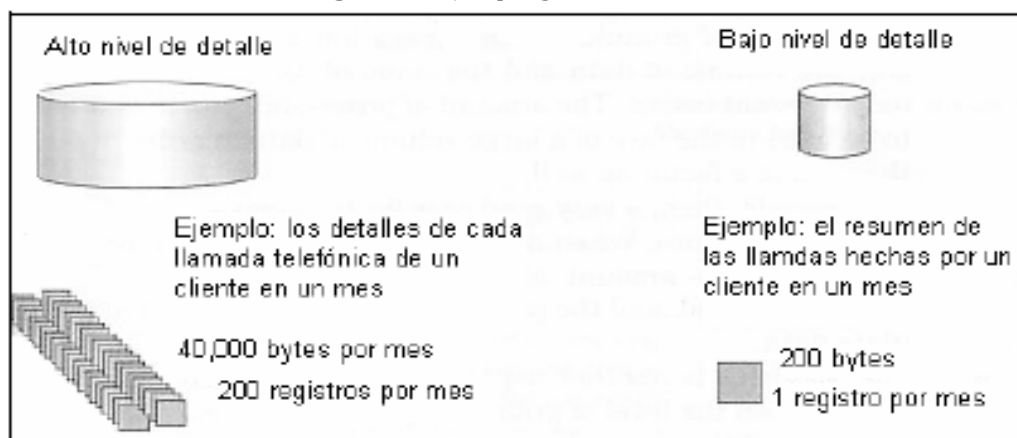
1. Selección del *Data Mart*

Se debe empezar con uno o más *Data Marts* que dependan de una única fuente de datos, para luego continuar con los que extraen datos de fuentes múltiples.

2. Declaración de granularidad de la tabla de hechos

Se selecciona el nivel de detalle al cual se requiere que sean expresados los datos, entre más alto el nivel de detalle más alta será la granularidad y viceversa, a la vez determinará el volumen de datos almacenados.

Figura 14. Ejemplo granularidad Kimball



Fuente: (Bustos S & Mosquera, 2013)

3. Selección de dimensiones

Por lo general las dimensiones están determinadas por el nivel de granularidad, un nivel de granularidad no puede ser menor a la granularidad de la tabla de hechos

4. Selección de los hechos

El nivel de granularidad seleccionado, permite la selección de los hechos, para los hechos transaccionales se tendrá una sola tabla de hechos, pero para los *snapshot* podemos contener varios resúmenes (precio, cantidad)

4. Diseño técnico de la arquitectura

Aspectos de arquitectura

El diseño de la arquitectura parte de un análisis de los sistemas *legacy* que se encuentran a disposición, para realizar una evaluación de capacidad de almacenamiento y rendimiento para el desarrollo del *Data Warehouse*.

Tabla 3: Consideraciones del diseño arquitectónico

Nivel de detalle	Datos ¿Qué?	Back Room ¿Cómo?	From Room ¿Cómo?	Infraestructura ¿Dónde?
Auditoría para los requerimientos del negocio	Qué información necesaria para la toma de decisiones Conexiones de componentes en arquitectura en bus	¿Cómo se Implementarán de procesos ETL? ¿Cómo estarán disponibles los procesos para los usuarios? ¿Evaluación de procesos actuales?	¿Cuáles son los retos que enfrenta el negocio ¿Cómo se quieren analizar los datos	Qué niveles del <i>Data Warehouse</i> se requiere que tengan éxito Cuáles existen en la actualidad
Modelos y documentos de arquitectura	Modelo dimensional ¿Cuáles son las principales entidades y relaciones? ¿Cómo se deben estructurar estas entidades?	¿Qué características son necesarias para tener los datos en forma útil en los lugares apropiados? ¿Cuáles son los principales almacenes de datos y donde deben estar localizados	¿Qué requerirán los usuarios para cargar información de una forma útil? ¿Qué clases de análisis y reportes se deben proveer y cuáles son las prioridades?	¿Cuál es el origen y destino de los datos? ¿Se tiene suficientes capacidades de procesamiento y almacenamiento?
Especificaciones y modelos	Modelos lógicos y físicos ¿Cuáles son los elementos individuales, sus derivaciones y reglas para la derivación? ¿Cuáles son los recursos y cómo se mapean a los destinos?	¿Qué estándares y productos proveen las características requeridas? ¿Cómo se relacionarán? ¿Cuáles son los estándares para desarrollo, administración de código y nombres?	¿Cuáles son las especificaciones de reportes incluyendo filas, columnas, secciones, encabezados, etc.? ¿Quién las necesita, que tan a menudo? ¿Cómo se	¿Cómo se interactúa con estas características? ¿Qué utilidades existen en el sistema, APIs, etc.?
Implementación	Crear bases de datos, índices, back up, documentación	Escritura de extracciones y cargas, automatización de procesos, documentación	Implementación del ambiente de reporte y análisis Construcción del conjunto de reportes inicial, entrenamiento a usuario Documentación	Instalación y pruebas de nuevos componentes de la infraestructura Conexión de las fuentes de los destinos y los clientes. Documentación

Fuente: (Peña J & Suárez, 2008)

5. Proceso de extracción, transformación y carga

Las principales sub-etapas de esta zona del ciclo de vida son: la extracción, la transformación y la carga (*ETL process*). Se definen como procesos de extracción a aquellos requeridos para obtener los datos que permitirán efectuar la carga del modelo físico acordado. Los procesos de transformación sirven para convertir o recodificar los datos fuente para cargar el modelo físico. Los procesos de carga de datos sirven para poblar el *Data Warehouse*. Ralph Kimball propone en esta etapa un PLAN

Paso 1. Plan de alto nivel

Identificación de los orígenes de los datos, transformación y carga a destinos; con el esquema es posible comunicar la complejidad del proyecto a gerencia y miembros del equipo de desarrollo.

Paso 2. Herramientas ETL

La selección de herramientas para el proceso de extracción, transformación y carga de los datos implican costos ya sea de licencias de herramientas desarrolladas o la implementación de procesos que realicen estas actividades.

Paso 3. Plan detallado

La selección del orden de carga de las tablas de hechos debe establecerse en este plan; en el cual se debe empezar con la dimensión más simple y continuar con las demás hasta llegar a la tabla de hechos.

Paso 4. Poblar una tabla de dimensión simple

- Extracción de una dimensión

Es necesario reportes de extracción, los cuales determinarán los campos requeridos para la carga, a la vez se debe considerar la hora de carga de las dimensiones relevantes. Se debe considerar en el proceso de carga el tipo de fuentes de datos origen: archivos y base de datos

- Transformación de una dimensión

Las transformaciones de datos van desde lo más sencillo a lo más complejo: Cambio de tipo de dato, asignación de llaves subrogantes y combinaciones de diferentes fuentes de datos

- Carga

El primer proceso de carga debe realizarse con las características de población masiva de datos del motor de base del modelo dimensiona, esto mejorará su rendimiento.

Paso 5. Implementación de la lógica del cambio de una dimensión

Para el cambio de datos de una dimensión es necesario tener en cuenta los siguientes tipos de respuesta al cambio

- Sobre escritura del dato existente en la dimensión por el de origen
- Insertar registro nuevo con una nueva llave subrogada
- Cambiar valor a un campo en la misma dimensión y colocar en su lugar en valor nuevo.

Paso 6. Poblar las dimensiones restantes

Las demás dimensiones pueden ser cargadas como el paso 4

Paso 7. Cargas históricas de hechos

En el proceso ETL debe existir un paso para reemplazar las llaves de las fuentes por las llaves subrogas que se han asignado a cada dimensión, para ello tenemos dos búsquedas.

1. Se realiza la técnica 2 del paso 5 y sea creado un nuevo registro con una nueva llave subrogada y un campo modificado con el resto del contenido original
2. Cuando se procesa los registros de la tabla de hechos, aquí se debe buscar en todas las dimensiones los registros que coincidan con el dato que se está buscando relacionar con la tabla de hechos, encontrado el dato debe ser reemplazada su llave subrogada por la nueva.

Paso 8. ETL de una tabla de hechos incremental

Es preciso cargar las bases de datos a partir de la última carga realizada; de esta manera solo se extraen y procesan únicamente las transacciones que han ocurrido; para ello se pueden usar *logs* que registren los cambios que son importantes para la bodega, también se pueden utilizar *triggers*.

Paso 9. Operación y automatización de la bodega de datos

La automatización de los procesos ETL es indispensable; para ello se debe tener en cuenta:

- Definición de tareas - Flujos y dependencias
- Horarios de tareas – basadas en tiempo y hechos
- Monitoreo
- *Logging*
- Manejo de excepciones

- Manejo de errores
- Notificaciones

6. Selección e instalación de productos

Selección de productos

Se debe tener en cuenta dos componentes fundamentales para la selección del producto

- Requerimientos técnicos
- Requerimientos del negocio

Mantener el negocio enfocado

Se debe hacer un estudio de cada una de las herramientas, considerando sus ventajas y desventajas; observar los requerimientos de los usuarios, así como las limitaciones de los mismos para con la herramienta; observando la arquitectura y los alcances técnicos pre definidos.

Áreas de evaluación

Se debe analizar algunas áreas para la selección de una herramienta adecuada para la implementación del *Data Warehouse*

- Plataforma de hardware
- Plataforma de base de datos
- Herramienta de *Data Staging*
- Herramienta de acceso a los datos

7. Características de aplicaciones para usuarios finales

El objetivo del *front room* es proporcionar la interfaz para el usuario: reportes, *dashboard*; tomando en cuenta el diseño y la estructura del *Data Warehouse*

Determinar el conjunto inicial de plantillas de reportes

Se deben realizar tres tareas fundamentales:

- Identificar los posibles reportes
- Consolidar la lista de candidatos

- Priorizar la lista de reportes

Definir la navegación para los reportes

Se debe desarrollar un modelo que permita a los usuarios encontrar la información necesaria de una manera rápida y eficiente.

Definir los estándares de las plantillas de reportes

La estandarización permite entender la naturaleza de los reportes

Detalles de especificación de las plantillas de reportes

Se debe establecer un formato de especificaciones que contenga el nombre, descripción o propósito, frecuencia con que debe ser actualizado, parámetros de entrada, ente otros

- Definición que provee información básica acerca de la plantilla del reporte
- El diseño, que proveer información visual de lo que debe mostrar el reporte

Tipos de aplicaciones para usuarios finales

- Basados en web: Los usuarios pueden acceder vía internet
- Herramienta independiente: Se diseñan plantillas, para su acceso desde otra aplicación.
- Herramienta de interfaz ejecutiva: Proporcionan una serie de *screens* para una mejor navegación.
- Interfaz por código: Permite el diseño de una interfaz.

8. Mantenimiento y crecimiento de un *Data Warehouse*

Administración del entorno del *Data Warehouse*

El mantenimiento posterior del *Data Warehouse* posterior a la implementación tiene gran importancia, tanto por el crecimiento tecnológico, así como por el crecimiento de la información.

Diagnóstico de resultados incorrectos

La revisión de pasos para la transferencia de archivos, para lograr resultados consistentes, es aquí donde surgen preguntas tales como

- ¿El equipo de desarrollo atendió correctamente los requerimientos del usuario?
- ¿Los usuarios finales entienden el modelo dimensional o tienen que recurrir a las complicadas tablas del *Data Warehouse*?

- ¿Se consideraron a los usuarios finales para la selección de las herramientas de análisis de datos?
- ¿Los datos utilizados para alimentar el *Data Warehouse* son correctos?
- ¿Los datos en el *Data Warehouse* son actualizados de manera oportuna?
- ¿EL tiempo de respuesta en las consultas es el adecuado?
- ¿Alguien responde cuando ellos necesitan consultoría para resolver problemas?

Enfoque en usuarios del negocio

Durante el desarrollo del proyecto se debe tener en cuenta la proximidad con el usuario, para prevenir deficiencias en el mismo, así como decepciones para con el producto final

Administración de operaciones

Para el proceso de carga de volúmenes extensos de datos se debe realizar el mantenimiento a la infraestructura técnica, y base de datos; se consideran los siguientes puntos para la realización de estas acciones

- Manejo de la infraestructura técnica
Monitoreo de las áreas críticas: Servidores de aplicaciones, servidores de base de datos entre otros
- Desempeño de la base de datos
Establecer indicadores de carga para medir el desempeño de la plataforma
- Mantenimiento de los datos y meta datos
Verificar los procesos de carga de datos y la transaccionabilidad de las operaciones diarias.

3.5.2. Metodología Hefestos

Introducción

Esta metodología se fundamenta en una amplia investigación y comparación entre metodologías existentes, además de expectativas propias del autor en procesos de desarrollo de almacenes de datos, la cual está en continua evolución y que toma en cuenta, como valor agregado, todas las aportaciones de la gran comunidad que la utiliza (Bernabeu R. , 2010).

Para la implementación de un *Data Warehouse* se debe tener en cuenta que las fases de la metodología no sean muy extensas, ni compliquen el desarrollo del mismo; es por ello que para Hefestos se requiere de cuatro pasos: Análisis de requerimientos, Análisis de los OLTP, Modelo lógico del *Data Warehouse* e Integración de datos; con esto logramos establecer las necesidades primordiales de la información por parte de los actores del sistema, e identificar las fuentes de datos de forma concreta y sus indicadores, y la creación de modelo de datos.

3.5.2.1. Características

La metodología de Hefestos posee las características listadas a continuación:

- En cada una de las fases se puede distinguir fácilmente los objetivos que se persiguen, así como los resultados esperados; adicionalmente son de fácil comprensión
- La estructura del *Data Warehouse* es de fácil y rápida adaptación, esto se debe a que fue construida en base a los requerimientos de los usuarios.
- La resistencia que presentan los usuarios finales al cambio se ve reducido gracias a que en cada etapa se los considera para determinar el comportamiento y las funciones que se incorporan al diseño del DW
- Los modelos conceptuales y lógicos que se implementan son de sencilla comprensión y análisis.
- El tipo de ciclo de vida que contenga a la metodología marcan independencia el uno del otro.
- Las herramientas que se utilicen para la construcción del DW son independientes de la metodología
- La metodología es independiente de las estructuras físicas y su distribución, que contengan el DW.
- Los resultados obtenidos al finalizar una fase se convierten en un nuevo punto inicial para el siguiente paso.
- Se aplica tanto para el *Data Warehouse* como para los *Data Mart*

3.5.2.2. Fases

Figura 15. Fases Metodología Hefesto



Fuente: elaboración propia

1. Análisis de requerimientos

El análisis de requerimientos se centra en la identificación de las necesidades del cliente, esto puede desarrollarse mediante preguntas que ayuden a la identificación de los objetivos que persigue la empresa; para luego tomar sus respuestas para determinar los indicadores y perspectivas en las cuales se basarán el modelo del DW; por último, se construirá un modelo conceptual, para la visualización de los resultados obtenidos en este primer paso.

Identificar preguntas

Los objetivos que se persiguen con la identificación de preguntas son: establecer la información clave de alto nivel, aquella que enmarcan los datos que llevarán a la empresa al cumplimiento de metas y desarrollo de estrategias para la toma de decisiones; proveer el soporte para el desarrollo de los pasos siguientes, debido a que la información proporcionada servirá a futuro.

Lo fundamental en este paso es formular preguntas complejas sobre la empresa, las cuales integran variables de análisis que sean relevantes para el estudio de la información desde diferentes perspectivas.

Hay que tener presente, que la información obtenida debe estar soportada de alguna manera por algún OLTP, ya que de otra manera no se podría implementar el *Data Warehouse*.

Las preguntas pueden formularse de la siguiente manera

- Se desea conocer cuántas unidades de cada producto fueron vendidas a sus clientes en un periodo determinado; en otras palabras: “Unidades vendidas de cada producto a cada cliente en un tiempo determinado”.
- Se desea conocer cuál fue el monto total de ventas de productos a cada cliente en un periodo determinado; en otras palabras: “Monto total de ventas de cada producto a cada cliente en un tiempo determinado”.

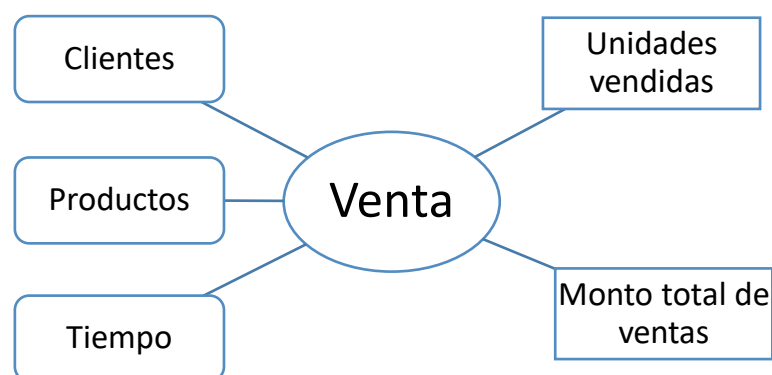
Identificar indicadores y perspectivas

La identificación de los indicadores y perspectivas se deriva de una descomposición de las preguntas obtenidas anteriormente, entre los indicadores efectivos se encuentran aquellos valores numéricos que representan lo que se desea analizar, por ejemplo: en una empresa de producción y ventas; los saldos, promedio de ventas, cantidades vendidas; mientras que las perspectivas se refieren a los objetos mediante los cuales se necesitan analizar los indicadores, por ejemplo: clientes, proveedores, sucursales, productos, entre estos identificamos también el tiempo.

Modelo Conceptual

Un modelo conceptual consiste en la descripción de alto nivel de la estructura de la base de datos, en la cual la información se presenta a través de objetos, relaciones y atributos. La construcción de este modelo se realiza en base a los indicadores y perspectivas, las cuales nos permitirán observar de una forma clara y precisa los alcances del proyecto.

Figura 16. Ejemplo modelo conceptual Hefesto



Fuente: (Bernabeu R. , 2010)

2. Análisis de los OLTP

El análisis de las fuentes de datos OLTP permite establecer cómo se obtendrán los indicadores de acuerdo al modelo conceptual formulado anteriormente; luego se incluirán en los campos para cada perspectiva y finalmente se ampliará el modelo conceptual antes definido

Conformar indicadores

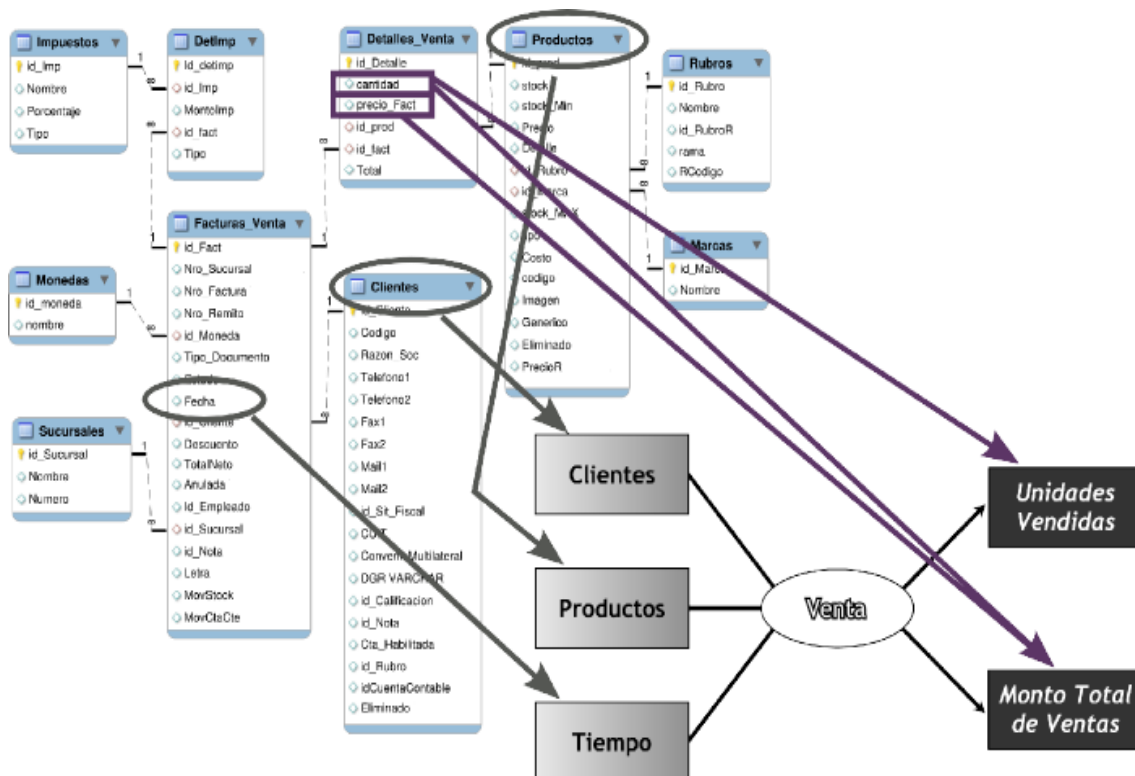
El cálculo de los indicadores debe ser detallado en este paso, de acuerdo a los conceptos siguientes:

- Los hechos que componen los indicadores y su fórmula para el cálculo, como por ejemplo $\text{Indicador} = \text{Hecho1} + \text{Hecho2}$
- En el caso de agregaciones de debe colocar las funciones de sumariación, tales como SUM, AVG, COUNT.

Establecer correspondencias

Se requiere identificar las correspondencias entre el modelo conceptual formulado y las fuentes de datos obtenidas de los OLTP disponibles, los cuales contengan la información que se requiere.

Figura 17. Ejemplo establecimiento de correspondencias en modelo conceptual Hefesto



Fuente: (Bernabeu R., 2010)

Nivel de granularidad

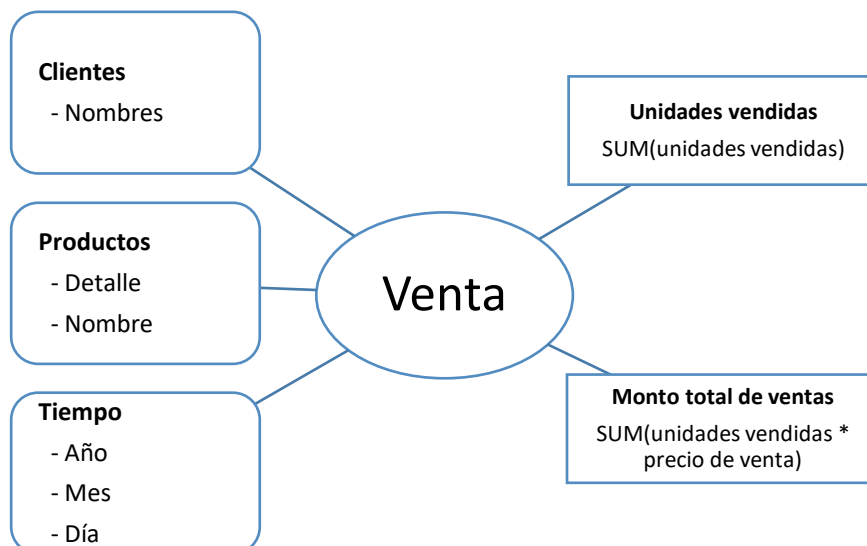
Selección de los campos que contendrán cada perspectiva, para su posterior examinación y filtrado de indicadores; estos son expuestos a los usuarios para que sean estos quienes determinen la relevancia o no de los datos encontrados en los OLTP. La perspectiva “Tiempo” debe ser detallada, puesto que entre sus posibles campos tenemos: día, mes, año, trimestre, semestre, año, etc.

- Con respecto a la perspectiva “Tiempo”, que es la que determinará la granularidad del depósito de datos, los datos más típicos que pueden emplearse son los siguientes:
 - Año.
 - Semestre.
 - Cuatrimestre.
 - Trimestre.
 - Número de mes.
 - Nombre del mes.
 - Quincena.
 - Semana.
 - Número de día.

Modelo conceptual ampliado

Se amplía el modelo conceptual generado con anterioridad colocando cada perspectiva y campos seleccionados, y bajo cada indicador la fórmula del cálculo mediante el cual se obtendrán dichos valores

Figura 18. Ejemplo modelo conceptual ampliado Hefesto



Fuente: elaboración propia

3. Modelo lógico del *Data Warehouse*

El modelo lógico es una representación de la estructura de base de datos, que puede procesarse y almacenarse en algún SGBD. El modelo conceptual formulado en los pasos anteriores nos permitirá la construcción de este modelo lógico; para ello se selecciona el tipo de modelo, se establecen las tablas tanto de dimensiones como de hechos y por último se establecen las uniones entre las tablas

Tipos de modelo lógico

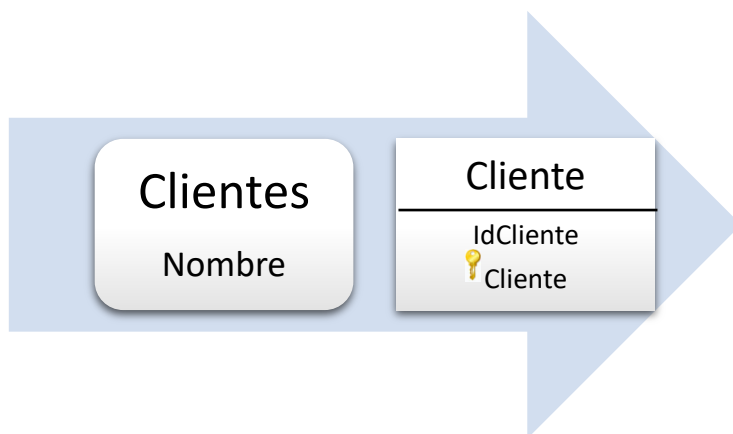
Los modelos lógicos pueden formarse de acuerdo con esquemas en estrellas, constelaciones o copos de nieve

Tablas de dimensiones

Las perspectivas y campos definidos anteriormente deben seguir el proceso:

- Elección de nombre que identifique la dimensión
- Integración de campo clave
- Redefinición de los campos poco intuitivos

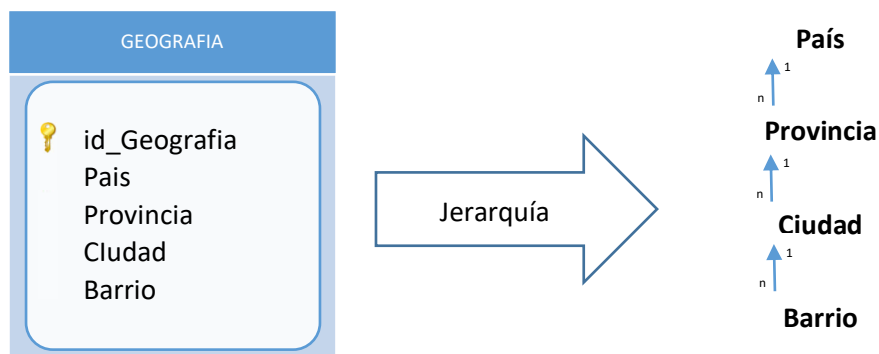
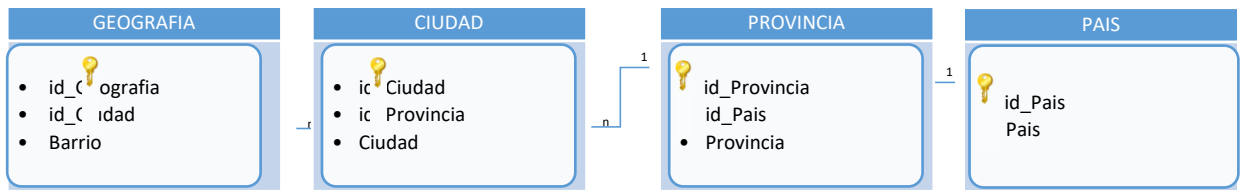
Figura 19. Ejemplo transformación de tablas de dimensiones Hefesto



Fuente: elaboración propia

Para las tablas que generen jerarquías, se debe realizar una normalización, la cual consiste en la creación de varias tablas a partir de una tabla única que contenga las relaciones padres – hijos entre sus campos.

Figura 20. Ejemplo tabla jerarquía Hefesto



Fuente: (Bernabeu R. , 2010)

Tablas de hechos

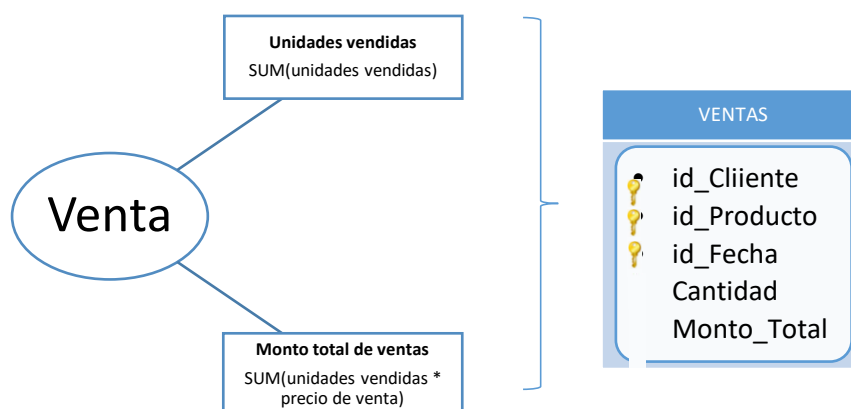
Definición de las tablas de hechos será determinada por el tipo de modelo lógico seleccionado; es así que para el modelo en estrella y copo de nieve se realiza lo siguiente;

- Nombrar a la tabla de hechos.
- Definir una clave primaria, la cual está compuesta de la asociación de las claves primaras de las tablas de dimensión relacionada.
- Crear campos de hemos como indicadores se defina en el modelo conceptual

Para los esquemas de constelación de realiza lo siguiente:

- Las tablas de hechos deben ser construidas acorde a las preguntas, indicadores y perspectivas.
- Nombrar las tablas de hechos, incorporar una clave formada por la combinación de las claves de las tablas de dimensiones relacionadas.

Figura 21. Ejemplo esquema estrella Hefesto

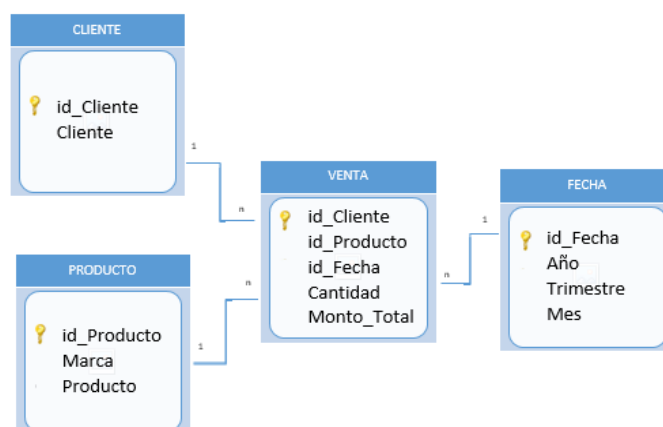


Fuente: (Bernabeu R. , 2010)

Uniones

Para las uniones entre las tablas ya sean de cualquier tipo de esquemas se pueden resumir en el gráfico adjunto

Figura 22. Ejemplo uniones de tablas en Hefesto



Fuente: (Bernabeu R. , 2010)

4. Integración de datos

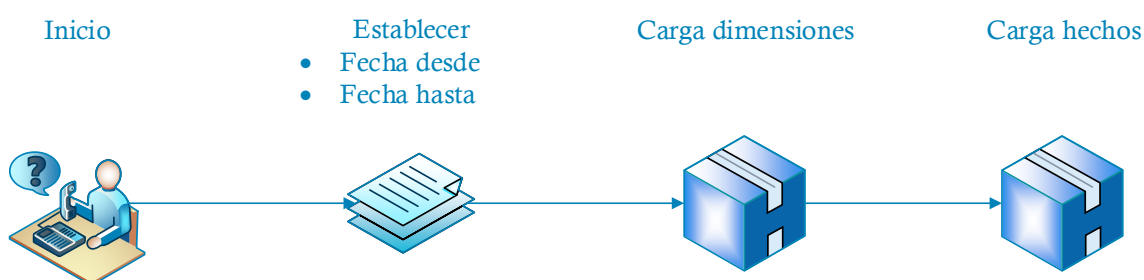
La carga de datos al modelo lógico anteriormente definido es el objetivo de este paso, es necesario emplear técnica de limpieza y calidad de datos, procesos ETL, etc.; así como las reglas y políticas para las continuas actualizaciones.

Carga inicial

Se deben establecer condiciones y restricciones para la carga de los datos hacia el *Data Warehouse*, se debe evitar la subida de datos con valores faltantes o absurdos; el orden de la carga de estos datos debe darse de la siguiente manera: primero las tablas de dimensiones, desde el nivel más general al detallado en el caso de tablas jerárquicas que hayan sido normalizadas, continuando con las tablas de hechos, respetando las correspondencias entre cada elemento.

Aquí se debe detallar paso a paso el proceso de carga de información hacia el *Data Warehouse*

Figura 23. Ejemplo del proceso de carga inicial en Hefesto

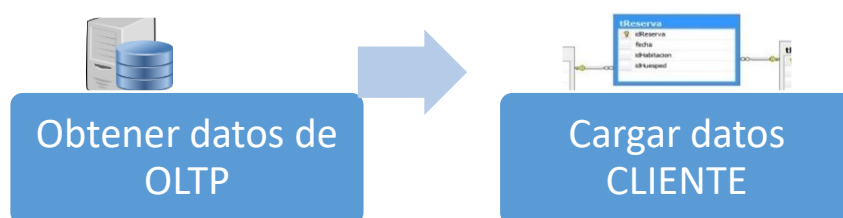


Fuente: elaboración propia

A continuación, se especificarán las tareas llevadas a cabo por "**Carga de Dimensión CLIENTE**".

Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

Figura 24. Ejemplo del proceso de carga dimensión cliente en Hefesto



Fuente: elaboración propia

Actualización

Para la actualización de datos del *Data Warehouse* se debe procurar el establecimiento de políticas y estrategias; consiguientemente con las actividades

- Especificación de las tareas de limpieza, calidad de datos, procesos ETL, etc.
- Especificación general y detallada de las acciones que cada software empleado para la carga de datos deberá realizar.

3.5.3. SAS Rapid Data Warehouse Methodology

Introducción

Siendo la construcción de un *Data Warehouse* un proceso evolutivo, este tiene que apoyarse en una metodología, la cual establezca sus fases en el alcance del proyecto; el seguir los pasos de la metodología y comenzar el desarrollo del *Data Warehouse* desde un área específica de una empresa, nos permite obtener resultados tangibles y minimiza el tiempo de entrega.

3.5.3.1. Fases

Figura 25. Fases metodología SAS



Fuente: (Fernández, 2015)

1. Definición de los objetivos

En esta fase se realiza una evaluación general del proyecto, un informe de las necesidades previas en el que se establece la situación actual de los datos, su fuente de origen, calidad y fiabilidad de los datos; esto debe hacerse a nivel de dirección, gerencia y área de desarrollo para determinar el alcance y funciones del *Data Warehouse*, también se definirán aquellos parámetros que permitan la evaluación de proyecto a su culminación. Esto debe ser entregado en un documento formal, determinando factibilidades y responsables de la toma de decisiones.

Definición del proyecto

En esta etapa se toman en cuenta los siguientes factores

- Elección de *sponsors* o patrocinadores del proyecto
- Elección de un equipo de alto nivel tanto para el negocio como para el desarrollo
- Identificar los recursos comprometidos al proyecto
- Determinar el propietario del proyecto
- Determinar los criterios de evaluación del proyecto
- Asegurar la unidad entre los miembros del equipo
- Determinar el módulo del proyecto a implementar

Evaluación de la infraestructura

Verificación de las capacidades la infraestructura tecnológica de la empresa para el levantamiento del *Data Warehouse*

2. Definición de requerimientos

En esta fase se mantienen entrevistas con los responsables de las áreas del negocio, se realizan estudios de los sistemas existentes; así como un proceso de entrevistas para validar y reforzar la orientación del negocio. Al finaliza esta fase se obtendrá un documento en que constan no solo las necesidades de los usuarios, sino también estrategias y arquitectura de implementación. Para ello se establece lo siguiente:

Reuniones de requerimentación

Mediante métodos como la entrevista, encuestas, etc., se examinan las necesidades de las unidades organizativas de la empresa.

Análisis

Se realiza un análisis entre los requerimientos de los usuarios de la organización y la capacidad de la infraestructura

3. Diseño y Modelización

Identificación de las fuentes de origen de los datos y los procesos de transformación para la obtención del modelo lógico de datos, este modelo lógico se convertirá en el modelo físico y se definirá la arquitectura del *Data Warehouse*. En esta fase se consideran los siguientes puntos para el documento:

- Diseño lógico y físico para el modelado de los datos
- Especificación detallada del modelo del proceso para la extracción, transformación y carga
- Creación del modelo de la aplicación o herramientas de explotación
- Diseño de aspectos adicionales tales como: modelado de los metadatos y la seguridad.

4. Construcción o implementación

Para esta etapa se deben entregar los cubos dimensionales y mecanismos de acceso. La implementación de un *Data Warehouse* se lleva a cabo mediante tres etapas:

Administración

Los datos transaccionales son extraídos y transformados desde las fuentes de origen hacia el *Data Warehouse*, considerando los procesos:

- Validación: se identifican datos inválidos, perdidos, fuera de rango y duplicados.
- Depuración: corrige los problemas de los datos identificados en el proceso anterior
- Integración: extrae de las variables de datos, significados consistentes valores y medidas.

Organización

Los datos son cargados, las estructuras indexadas mientras los metadatos son capturados y creados.

Explotación

La construcción de los mecanismos de acceso al *Data Warehouse*: reportes, gráficos, *dashboards*

5. Pruebas finales

Se realizan pruebas de calidad del proyecto antes de ser entregados al cliente, verificando que los requerimientos hayan sido totalmente incorporados.

6. Despliegue

Instalación y puesta en marcha del proyecto; para esta etapa se consideran los siguientes puntos:

- Instalación inicial, incluyendo las conexiones del *Data Warehouse* con las fuentes de datos, actualización y sincronización.
- Capacitación a usuarios
- Planificación e implementación de actualizaciones y mantenimiento del *Data Warehouse*.

7. Revisión

La evaluación del *Data Warehouse* permitirá definir los aspectos a mejorar o potenciar en función de la utilización del mismo, esto puede darse en un tiempo determinado posterior a su uso. En esta fase se conducen tres ejecuciones con cada construcción

- Seguir la fase de construcción para evaluar el proceso de implementación y aprendizaje de satisfacciones y contratiempos
- A partir de la tercera a sexta semana verificar la fase de despliegue y asegurar que la transición no ha tenido inconvenientes, y que los usuarios pueden acceder al *Data Warehouse*
- A partir de los dieciocho a veinte y cuatro meses se debe medir el beneficio tangible, calcular el Retorno de Inversión (ROI).

Capítulo 4

Metodología

4.1. Diseño de la Investigación

El estudio a realizarse es de tipo descriptiva comparativa, ya que involucra la evaluación de las fases y actividades de cada metodología a analizar. Su meta no se limita a la recolección de datos, sino a la predicción e identificación de las relaciones que existen entre los criterios de evaluación.

La exteriorización de los datos se presentará de forma cualitativo lo cual ayudarán a la identificación de los factores importantes que deben ser valorados, para lo cual se manejará la escala de Likert en la ejecución del estudio comparativo (REDIFP, 2013).

En esta investigación se tendrá como ambiente operativo del *Data Warehouse* para el Instituto Nacional de Patrimonio Cultural Regional 3 denominado (INPC R3), esto es: almacenamiento, consolidación y extracción de información de diversas fuentes, para tratarla, permitiendo su exploración e investigación, desde diferentes perspectivas, facilitando la toma de decisiones.

Según el Estudio Comparativo de Metodologías e implementación de alternativas BI *open source* vs propietarias en entornos tradicionales de la Ingeniera Magdalena Garcés en el cual se ha determinado que las metodologías Hefestos, Kimbal y SAS *Methodology* son las más óptimas para la implementación de un proyecto BI. Por lo cual se ha visto la necesidad de determinar la metodología que respalden el alcance de las metas planteadas para el INPC R3.

Los procedimientos tomados para esta investigación:

- Se realiza la investigación de las metodologías que apoyan la creación de *Data Warehouse*.
- Se determinan los objetivos mediante que permitirán delimitar el estudio comparativo.
- Se justifican los motivos por los cuales se propone realizar la siguiente investigación
- Se elabora un marco teórico que ayude a crear una idea general para la realización del estudio comparativo.
- Se plantea los lineamientos de evaluación para el estudio comparativo que posee una íntima relación entre la problemática a tratarse y el objetivo.

- Se propone la operacionalización de las variables para el estudio comparativo, recolección de datos y se observa el comportamiento del ambiente operativo del *Data Warehouse* del INPC R3.
- Se formula la creación de tres prototipos utilizando las metodologías: Hefestos, Kimball y *SAS Methodology*.
- Se establecen las métricas y los resultados obtenidos de las mismas.
- Se elabora las conclusiones y recomendaciones, del producto de la comparativa realizada.

4.2. Métodos, Técnicas e Instrumentos

4.2.1. Métodos

En el estudio comparativo se utilizarán los siguientes métodos de investigación:

Método Descriptivo: Se utiliza este tipo de investigación el cual ayudará a conocer las situaciones predominantes a través de la descripción exacta de los procesos, actividades y tareas que pertenecen a las metodologías Hefestos, Kimball y *SAS Methodology*. Se recoge los datos sobre la una serie de variables comparativas, se expone la información de una manera cuidadosa y luego se analiza minuciosamente los resultados, a fin de extraer generalizaciones significativas que contribuyan al conocimiento.

Método Comparativo: Se analizará y estudiará cada uno de los procesos que corresponden a cada una de las metodologías en base a parámetros de evaluación definidos que permitan establecer similitudes y diferencias; y de ello obtener como resultado la metodología idónea que respalden el alcance de las metas planteadas para el INPC R3.

4.2.2. Técnicas

Se utilizarán ciertas técnicas para la recolección de información tales como:

- **Observación Directa:** Técnica que consiste en observar detenidamente los indicadores que formarán parte del análisis de los resultados.
- **Entrevista:** Técnica utilizada que facilita la recopilación información de parte del usuario para la requerimentación del proyecto *Data Warehouse*.
- **Encuesta:** Técnica utilizada para la obtención de información subjetiva del usuario acerca de su entorno, que permita conocer aspectos relevantes del proyecto *Data Warehouse*.

- **Recopilación de Información:** Se revisará los documentos generados por el sistema transaccional y reportes generados de forma manual.
- **Método Científico:** Estudia los procesos del ciclo de vida que corresponden a las metodologías y cada una de sus fases.
- **Análisis:** Se analizará y estudiará las fases y actividades propias de las metodologías.
- **Pruebas:** Se utilizará parámetros de comparación para la determinación de la metodología idónea que respalden el alcance de las metas planteadas para el INPC R3.

4.2.3. Instrumentos

Debido al entorno de la investigación, las herramientas que facilitan para la recolección de los datos son las fichas técnicas; con las cuales se recopilarán información de los parámetros de comparación de acuerdo a la escala de Likert y se presentarán los resultados.

4.3. Procesamiento de la Información

Para la realización del estudio comparativo de las metodologías seleccionadas que permita establecer la metodología acorde a los procesos del INPC R3, se tomará en cuenta los indicadores para el estudio comparativo.

Los indicadores han sido tomados del Estudio Comparativo de Metodologías e Implementación de alternativas *Business Intelligence OpenSource* Vs. Propietarias en entornos tradicionales; estos indicadores están basados en las fases y actividades o índices involucradas en la construcción de *Data Warehouse*, los cuales permitirán determinar que metodología tiene mayor porcentaje de utilidad para la gestión e implementación del *Data Warehouse*, esta evaluación se realizara de forma cuantitativa y cualitativa.

Finalmente se elaborará una tabla resumen a la cual se le asignará porcentajes mediante una escala de valoración cualitativa para escoger la metodología con la cual se realizará el desarrollo del *Data Warehouse*.

4.4. Parámetros a comparación de fases metodológicas

Tabla 4: Parámetros de comparación de fases metodológicas

Indicadores	Índices
Fase 1: Requerimentación	Determinación de requerimientos
	Características de la organización

	Análisis de usuarios
Fase 2: Estrategia de proyecto	Cuantificar el tiempo para cumplir la necesidad (corto-mediano-largo plazo)
	Definir ventajas y desventajas
Fase 3: Planificación del proyecto	Cronograma de actividades
	Análisis de riesgo
	Definición de responsabilidades y grupo de trabajo
Fase 4: Selección de la tecnología	Entorno actual tecnológico de la organización
	Definición de tecnologías
Fase 5: Diseño del sistema de información	Determinación del modelo de información
	Diseño de la interfaz de usuario
Fase 6: Elaboración del sistema de información	Análisis de requerimientos (indicadores)
	Modelo conceptual
	Análisis OLTP
	Conformar indicadores
	Nivel de granularidad
	Nivel conceptual ampliado
	Modelo lógico de la estructura del DW
	Tablas de dimensiones
	Tablas de hechos
	Procesos ETL

Fuente: (Garcés, 2015)

4.5. Población y muestras

La población es el conjunto de elementos a ser evaluados, la población constituye las metodologías de desarrollo del *Data Warehouse* y los estándares para el desarrollo del mismo.

La muestra se encuentra determinada por una selección no probabilística de las metodologías de desarrollo para *Data Warehouse* el Estudio Comparativo de Metodologías e implementación de alternativas BI *open source* vs propietarias en entornos tradicionales de la Ingeniera Magdalena Garcés en el cual se ha determinado que las metodologías Hefestos, Kimbal y *SAS Methodology* son las óptimas para la implementación de un proyecto BI.

- Metodología Kimball
- Metodología Hefesto
- *SAS Methodology*

4.5.1. Creación de Prototipos

La formulación de los prototipos, consiste en el desarrollo metodológico del *Data Warehouse* mediante los cuales se establecerán las métricas necesarias para determinar la metodología que respalden el alcance de las metas planteadas para el INPC R3 (INPC, 2015).

En la actualidad se ha evidenciado la dificultad al momento de acceder a la información crucial de las colecciones de datos almacenados en los sistemas tradicionales para la cual aparecen los *Data Warehouse* que están orientados a temáticas que ofrecen una solución eficaz y eficiente que para solventarla.

La *Data Warehouse* consolida la información de diversas fuentes, para tratarla permitiendo su exploración, investigación desde perspectivas diferentes, facilitando los tiempos de respuestas al utilizarla de manera informacional o analítica para la toma decisiones.

Data Warehouse “es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia” (Bernabeu D. , 2009)

4.5.1.1. Metodología Kimball

La metodología Kimball es un proceso que garantiza la creación exitosa de una fuente de datos, integrada, no volátil y variable en el tiempo que ayuda a la toma de decisiones; se basa en el ciclo de vida dimensional del negocio que facilita la construcción de DW.

Sus fases se encuentran desarrolladas en el Apéndice A.

4.5.1.2. Metodología Hefesto

Hefesto se ha convertido con el paso de los años en una metodología robusta para la creación de *Data Warehouse* (DW) desde cero, su pragmatismo facilita la adaptación con cualquier ciclo de vida al enfocarse en el análisis de requerimientos, análisis de fuentes de datos para la elaboración e implementación de DW.

Sus fases se encuentran desarrolladas en el Apéndice B.

4.5.1.3. SAS Methodology

El almacenamiento de datos es un tema técnico y metodológico que determinan el éxito o el fracaso de los sistemas, por lo que SAS proporciona un almacén de datos que abarca varios ámbitos como son: Gestión de almacenes, organización y explotación, consta de una arquitectura de motor

con paso para MySQL ya que su aporte facilita el desarrollo de sistemas transaccionales legados racionales y otros; su base esta cimentada en la experiencia adquirida en el desarrollo de las diferentes fases lógicas e hitos para verificar que las soluciones tengan bases sólidas.

Para esta metodología se han desarrollado sus fases en el Apéndice C.

Capítulo 5

Resultados

5.1. Evaluación preliminar

5.1.1. Determinación de parámetros de comparación fase metodológica

Para la identificación de la metodología que respalde el alcance de las metas planteadas para el INPC R3 es necesario realizar un estudio descriptivo-comparativo de sus fases y actividades involucradas, para ello se elabora una lista de todas las tareas que las metodologías elegidas efectúan, y se agruparon sus criterios en 6 fases.

Usando la escala de Likert para la respectiva demostración, tomando los valores previamente estudiados, siendo:

Tabla 5: Escala de Likert

Escala	Valoración
Baja	1
Media	2
Alta	3

Fuente: elaboración propia

5.1.1.1. Fase 1. Requerimentación

Tabla 6: Indicador. Fase 1. Requerimientos

Indicador	Parámetros de evaluación
Fase 1 Requerimentación	Determinación de requerimientos
	Característica de la organización
	Análisis de usuarios

Fuente: elaboración propia

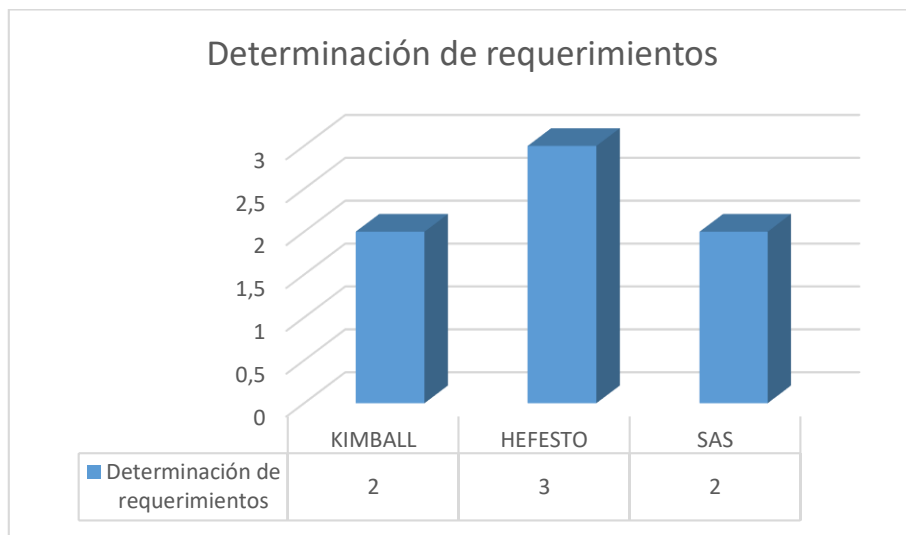
a. Determinación de requerimientos

Tabla 7: Parámetro. Determinación de requerimientos

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Determinación de requerimientos	2	3	2	7

Fuente: elaboración propia

Gráfico 1. Parámetro. Determinación de requerimientos



Fuente: elaboración propia

Interpretación

La metodología SAS cumple con los criterios de valoración expuestos de una forma básica, debido a que realiza una requerimentación muy pobre, la cual se basa en una recolección de datos funcionales y no funcionales para el *Data Warehouse*; mientras que, la metodología Kimball realiza una requerimentación amplia ya que recolecta datos mediante entrevistas, análisis de documentos y sistemas fuentes, por lo cual recibe una puntuación media; obteniendo el primer lugar la metodología Hefesto, debido que cumple con los parámetros anteriormente descritos y adicionalmente realiza un análisis del modelo conceptual de los sistemas fuente.

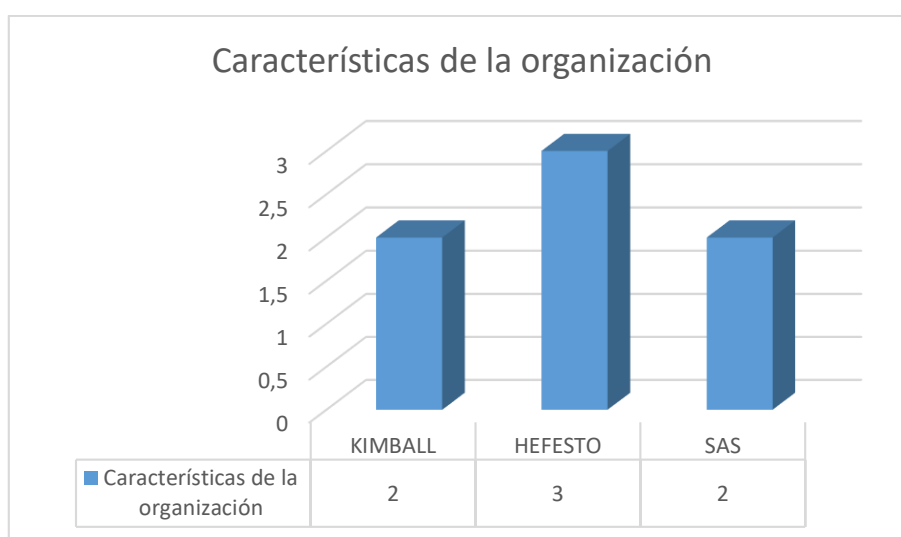
b. Características de la organización

Tabla 8: Parámetro. Características de la organización

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Características de la organización	2	3	2	7

Fuente: elaboración propia

Gráfico 2. Parámetro. Características de la organización



Fuente: elaboración propia

Interpretación

La evaluación de la visión, misión, objetivos y metas de la empresa; con las cuales se podrá tener como base un ámbito predefinido que ayude a la comprensión y a la toma de decisiones con respecto a la implementación y diseño del *Data Warehouse*, por consiguiente la metodología de Kimball y SAS reciben un criterio de evaluación medio debido a que realizan una inspección rápida de la empresa, mientras que Hefesto, permite realizar una exploración detallada de la estructura funcional y administrativa de la misma.

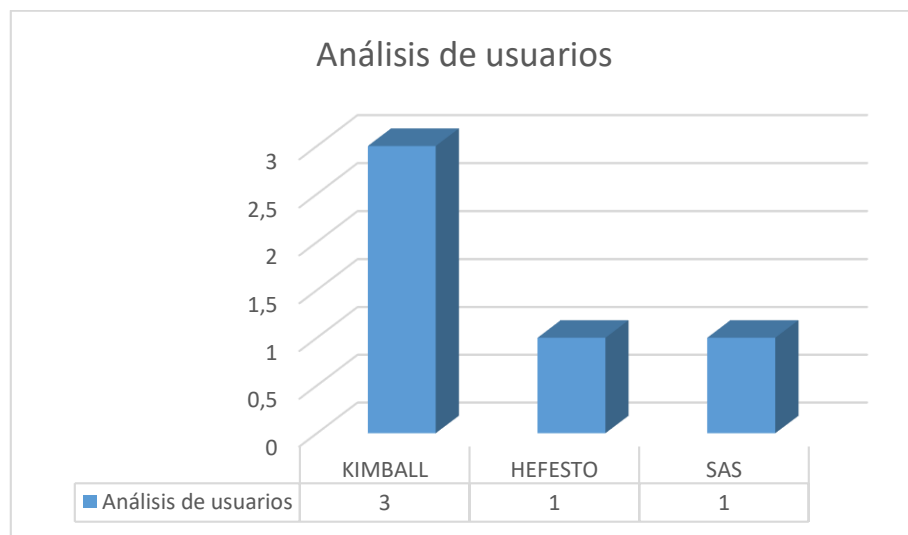
c. Análisis de usuarios

Tabla 9: Parámetro. Análisis de usuarios

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Análisis de usuarios	3	1	1	5

Fuente: elaboración propia

Gráfico 3. Parámetro. Análisis de usuarios



Fuente: elaboración propia

Interpretación

La identificación de sponsors determinará el apoyo y soporte de los demás miembros de la empresa, por lo cual Kimball recibe un valor alto en este indicador; para Hefesto y SAS esta identificación de usuarios ayuda al establecimiento de los requerimientos para el *Data Warehouse*.

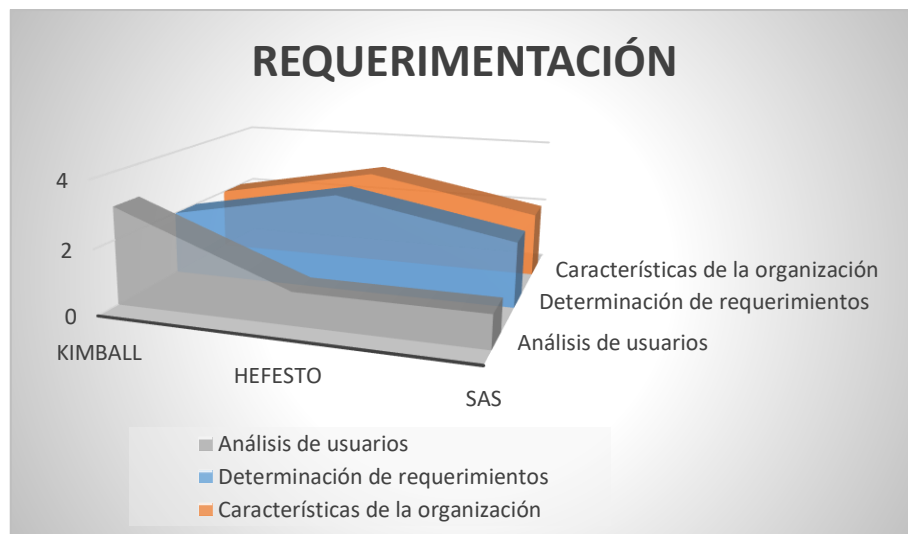
A continuación, se presenta el resumen de la valoración del Indicador Requerimientos

Tabla 10: Resumen. Indicador requerimientos

Indicador	Metodologías desarrollo de <i>Data Warehouse</i>		
	Kimball	Hefesto	SAS
Determinación de requerimientos	2	3	2
Características de la organización	2	3	2
Análisis de usuarios	3	1	1
TOTAL	7	7	5

Fuente: elaboración propia

Gráfico 4. Resumen. Indicador Requerimientos



Fuente: Elaboración propia

5.1.1.2. Fase 2: Estrategia de proyecto

Tabla 11: Indicador. Fase 1. Requerimientos

Indicador	Parámetros de evaluación
Fase 2 Estrategia del proyecto	Cuantificar el tiempo para cumplir la necesidad
	Definir ventajas y desventajas

Fuente: elaboración propia

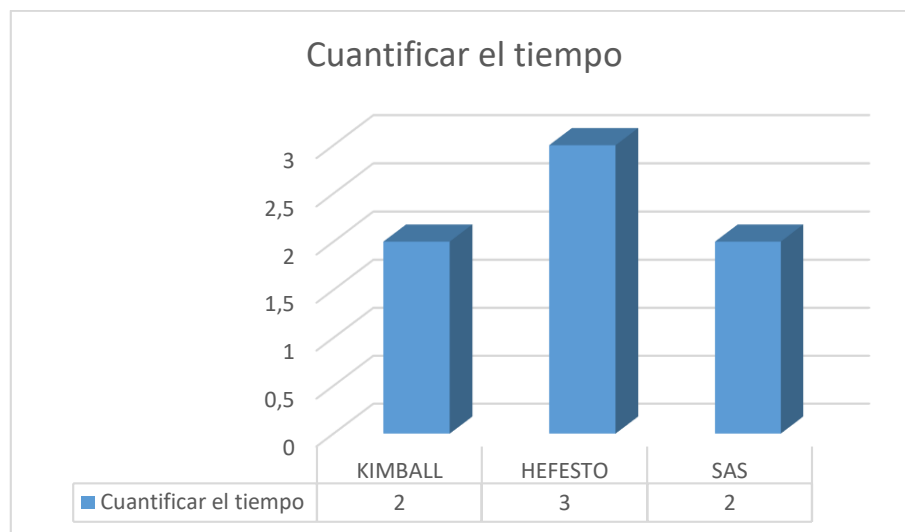
- a. Cuantificar el tiempo para cumplir la necesidad

Tabla 12: Parámetro. Cuantificar el tiempo para cumplir la necesidad

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Cuantificar el tiempo para cumplir la necesidad	2	3	2	7

Fuente: elaboración propia

Gráfico 5. Parámetro. Cuantificar el tiempo para cumplir la necesidad



Fuente: elaboración propia

Interpretación

La metodología de Hefesto cumple satisfactoriamente con el parámetro a considerarse, debido a que se realiza un estimado del tiempo en el cual se realizará la implementación del proyecto, mientras que Kimball y SAS realizan un cronograma de actividades y entrega de proyecto.

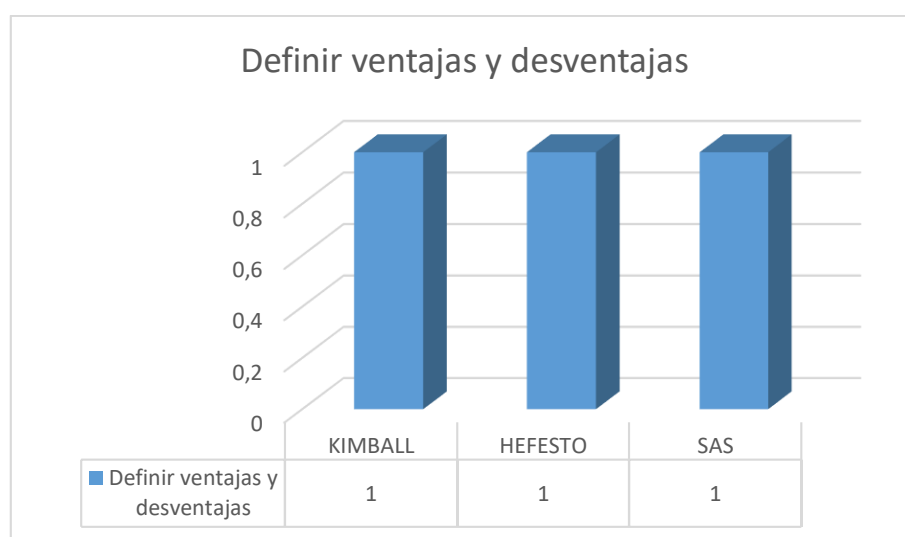
b. Definir ventajas y desventajas

Tabla 13: Parámetro. Definir ventajas y desventajas

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Definir ventajas y desventajas	1	1	1	3

Fuente: elaboración propia

Gráfico 6. Parámetro. Definir ventajas y desventajas



Fuente: elaboración propia

Interpretación

Ninguna de las metodologías expuestas para evaluación, contienen un estudio previo relacionado con las ventajas y desventajas que el *Data Warehouse* proporcionaría a la empresa; se construye una proyección enfocada utilidad que el proyecto brindará.

A continuación, se presenta el resumen de la valoración del Indicador Estrategias del proyecto

Tabla 14: Resumen. Indicador estrategias del proyecto

Indicador	Metodologías desarrollo de <i>Data Warehouse</i>		
	Kimball	Hefesto	SAS
Cuantificar el tiempo	2	3	2
Definir ventajas y desventajas	1	1	1
TOTAL	3	4	3

Fuente: elaboración propia

Gráfico 7. Resumen. Indicador estrategias del proyecto



Fuente: elaboración propia

5.1.1.3. Fase 3: Planificación del proyecto

Tabla 15: Indicador. Fase 3. Planificación del proyecto

Indicador	Parámetros de evaluación
Fase 3 Planificación del proyecto	Cronograma de actividades
	Análisis de riesgos
	Definición de responsabilidades y grupo de trabajo

Fuente: elaboración propia

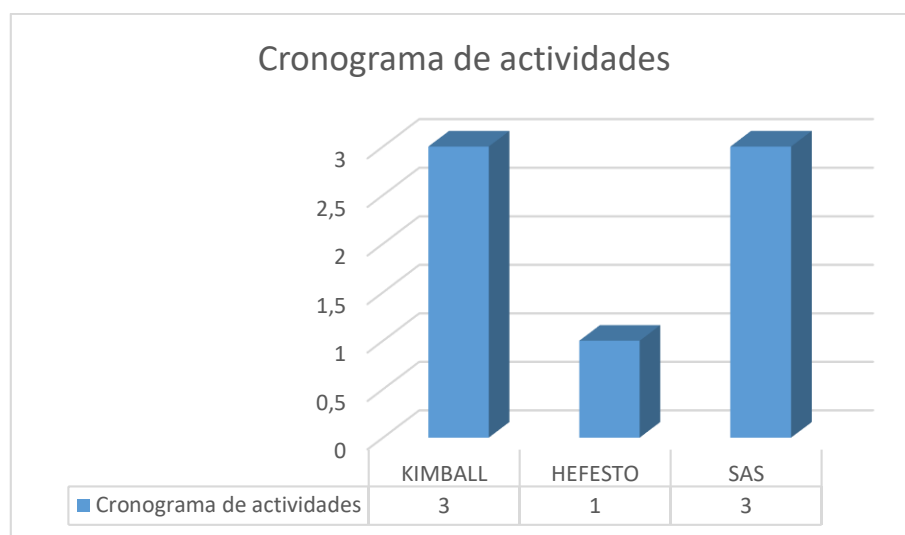
a. Cronograma de actividades

Tabla 16: Parámetro. Cronograma de actividades

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Cronograma de actividades	3	1	3	7

Fuente: elaboración propia

Gráfico 8. Parámetro. Cronograma de actividades



Fuente: elaboración propia

Interpretación

La metodología de Kimball y SAS realizan el establecimiento de un cronograma, cada uno basado en las fases, por lo tanto, reciben la valoración más alta; Hefesto plantea una entrega del proyecto sobre estimados de tiempo

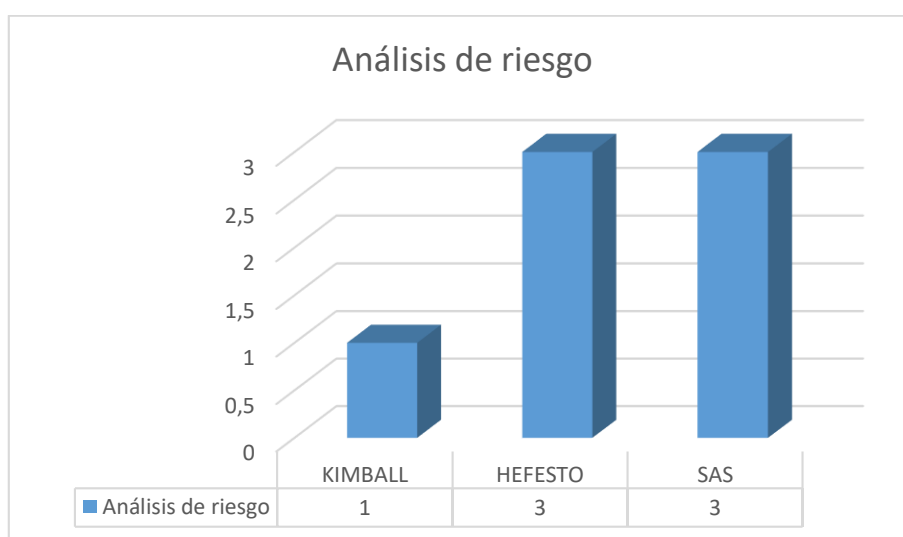
b. Análisis de riesgos

Tabla 17: Parámetro. Análisis de riesgos

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Análisis de riesgo	1	3	3	7

Fuente: elaboración propia

Gráfico 9. Parámetro. Análisis de riesgos



Fuente: elaboración propia

Interpretación

La metodología SAS realiza un análisis de riesgos enfocado a identificar posibles problemas y mitigar riesgos en el desarrollo del proyecto, esta metodología recibe un puntaje alto; al igual que Hefesto realiza una evaluación de riesgos en cada fase para minimizarlos y producir entregas en tiempos razonables, Kimball no provee al proyecto de una fase específica en la cual se realice un análisis de estos riesgos implicados en el desarrollo del *Data Warehouse*, por ello se lo califica con un puntaje bajo.

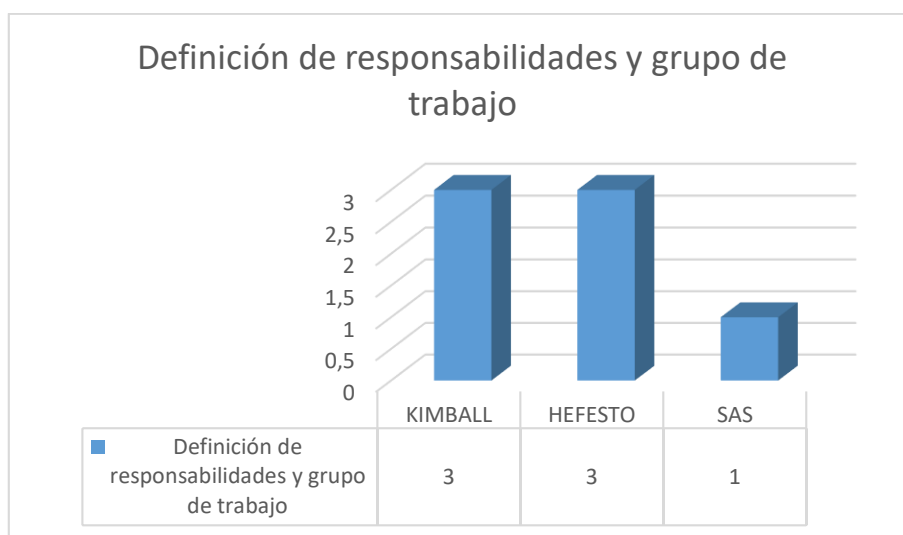
c. Definición de responsabilidades y grupo de trabajo

Tabla 18: Parámetro. Definición de responsabilidades y grupos de trabajo

Parámetro de evaluación	Metodologías desarrollo de Data Warehouse			Total
	Kimball	Hefesto	SAS	
Definición de responsabilidades y grupo de trabajo	3	3	1	7

Fuente: elaboración propia

Gráfico 10. Parámetro. Definición de responsabilidades y grupos



Fuente: elaboración propia

Interpretación

El establecimiento de los grupos de trabajo en SAS, se realizan de forma interna, por lo cual reciben un puntaje bajo, por el contrario, Hefestos y Kimball identifican miembros y roles para el desempeño de cada uno de los requerimientos, donde a cada miembro se le puede asignar varios requerimientos, dependiendo del tamaño del proyecto.

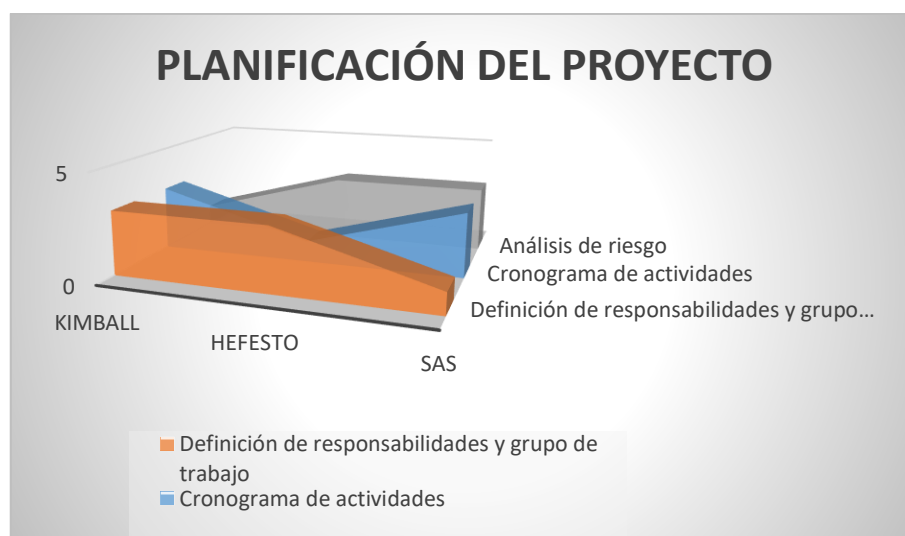
A continuación, se presenta el resumen de la valoración del Indicador Planificación del proyecto

Tabla 19: Resumen. Indicador planificación del proyecto

Indicador	Metodologías desarrollo de <i>Data Warehouse</i>		
	Kimball	Hefesto	SAS
Cronograma de actividades	3	1	3
Análisis de riesgo	1	3	3
Definición de responsabilidades y grupo de trabajo	3	3	1
TOTAL	7	7	7

Fuente: Elaboración propia

Gráfico 11. Resumen. Indicador planificación del proyecto



Fuente: elaboración propia

5.1.1.4. Fase 4: Selección de la tecnología

Tabla 20: Indicador. Fase 4. Selección de la tecnología

Indicador	Parámetros de evaluación
Fase 4 Selección de la tecnología	Entono actual tecnológico de la organización
	Definición de tecnologías

Fuente: elaboración propia

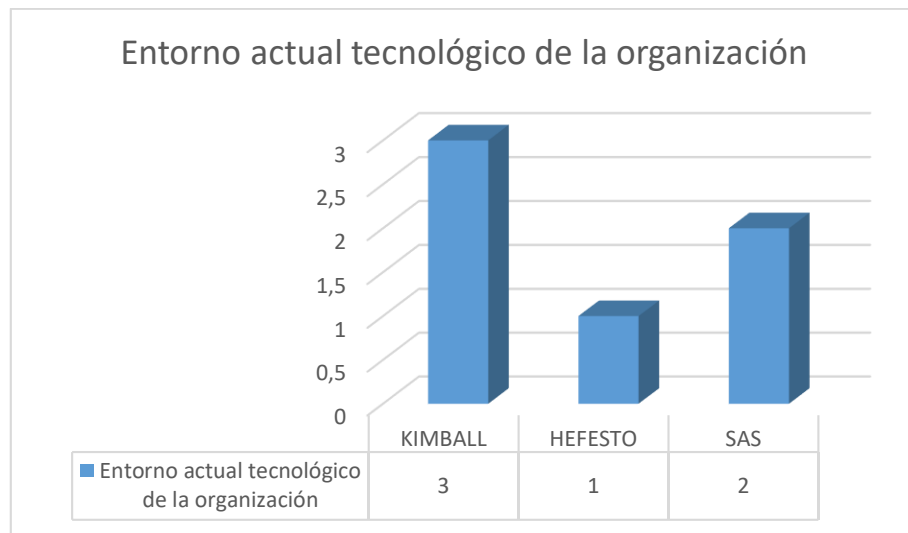
a. Entorno actual tecnológico de la organización

Tabla 21: Parámetro. Entorno actual tecnológico de la organización

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Entorno actual tecnológico de la organización	3	1	2	6

Fuente: elaboración propia

Gráfico 12. Parámetro. Entorno actual tecnológico de la organización



Fuente: elaboración propia

Interpretación

El análisis del entorno tecnológico actual o sistemas *legacy* actuales, es importante, que los sistemas sean consistentes y manejen de forma correcta sus transacciones, debido a que en el desarrollo de del *Data Warehouse* se toma como hecho que estos sistemas son confiables, para lo cual, Kimball establece una matriz que apoya este análisis, obteniendo el puntaje más alto, mientras que SAS se encuentra en un nivel medio, debido a su sencilla verificación de equipos y sistemas, y por último Hefestos, el cual da por supuesto que la arquitectura actual es correcta, y solo se enfoca la implementación de nuevos equipos.

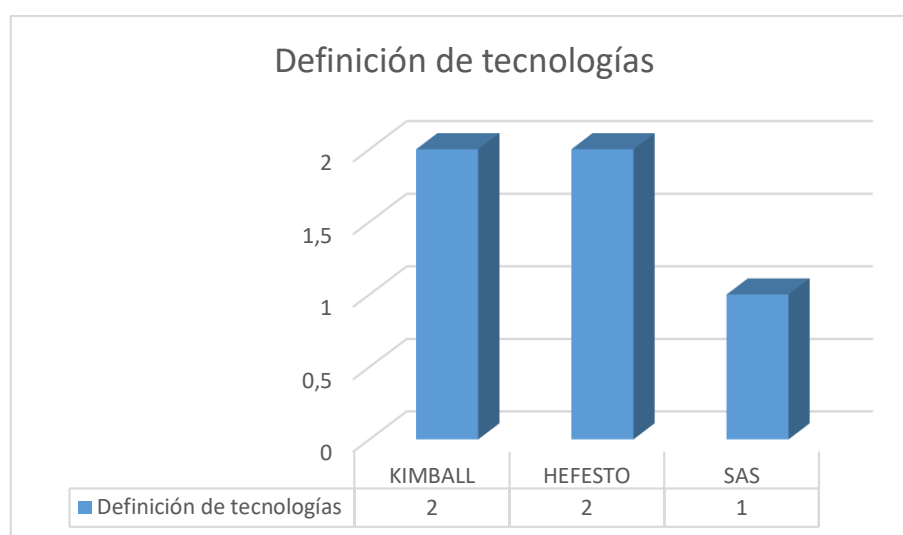
b. Definición de tecnologías

Tabla 22: Parámetro. Definición de tecnologías

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Definición de tecnologías	2	2	1	5

Fuente: elaboración propia

Gráfico 13. Parámetro. Definición de tecnologías



Fuente: elaboración propia

Interpretación

El análisis de las tecnologías y definición de las mismas han sido evidenciados en las metodologías Kimball, Hefestos las cuales realizan un estudio previo de las necesidades de la empresa entorno a la capacidad económica con que cuenta la empresa, debido a este factor que reciben una calificación media, mientras que SAS define de forma rápida y concisa las tecnologías a emplearse, recibiendo una valoración baja.

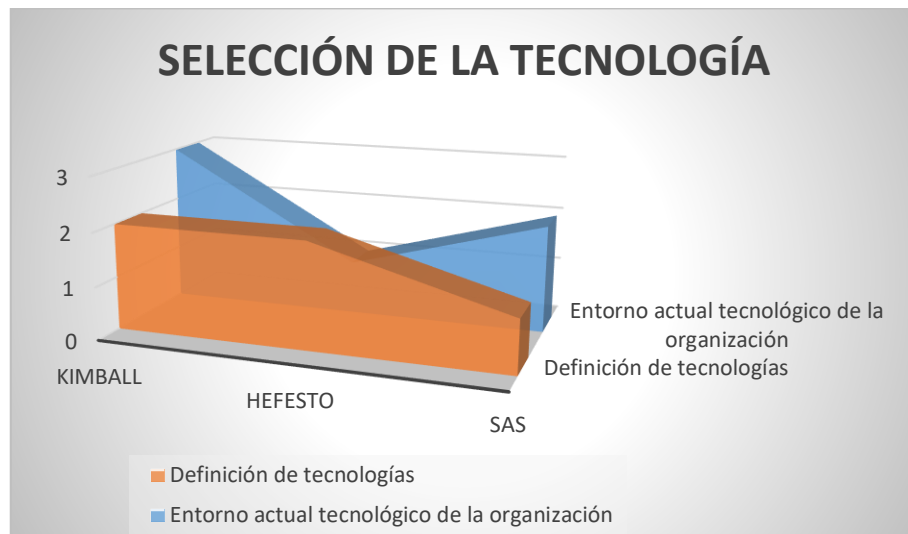
A continuación, se presenta el resumen de la valoración del Indicador Selección de la tecnología.

Tabla 23: Resumen. Indicador selección de la tecnología

Indicador	Metodologías desarrollo de <i>Data Warehouse</i>		
	Kimball	Hefesto	SAS
Entorno actual tecnológico de la organización	3	1	2
Definición de tecnologías	2	2	1
TOTAL	5	3	3

Fuente: elaboración propia

Gráfico 14. Resumen. Indicador selección de la tecnología



Fuente: elaboración propia

5.1.1.5. Fase 5: Diseño del sistema de información

Tabla 24: Indicador. Fase 5. Diseño del sistema de información

Indicador	Parámetros de evaluación
Fase 5 Diseño del sistema de información	Determinación del modelo de información
	Diseño de la interfaz de usuario

Fuente: elaboración propia

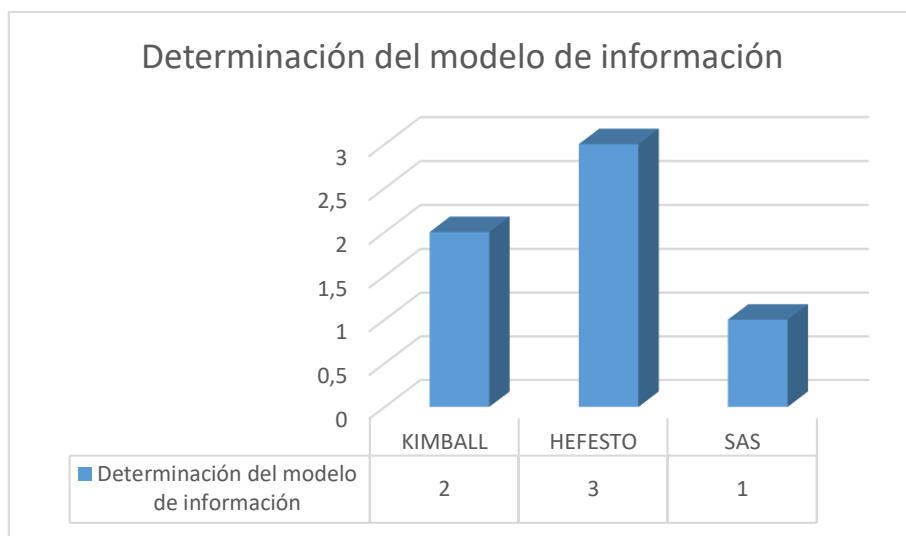
a. Determinación del modelo de información

Tabla 25: Parámetro. Determinación del modelo de información

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	KIMBALL	HEFESTO	SAS	
Determinación del modelo de información	2	3	1	6

Fuente: elaboración propia

Gráfico 15. Parámetro. Determinación del modelo de información



Fuente: elaboración propia

Interpretación

La identificación de las tablas de dimensiones y hechos se enmarca dentro este indicador, para Hefesto esta identificación es clara, y se basa en la definición de requerimientos y la sobre posición de las nuevas tablas sobre el modelo relacional del cual partimos, para Kimball esta definición está basada en la requerimentación, para estas metodologías el establecimiento de estándares de tablas y columnas es bastante amplio, permitiendo un mejor desarrollo, mientras que SAS cuenta con una requerimentación simple y pobre de datos.

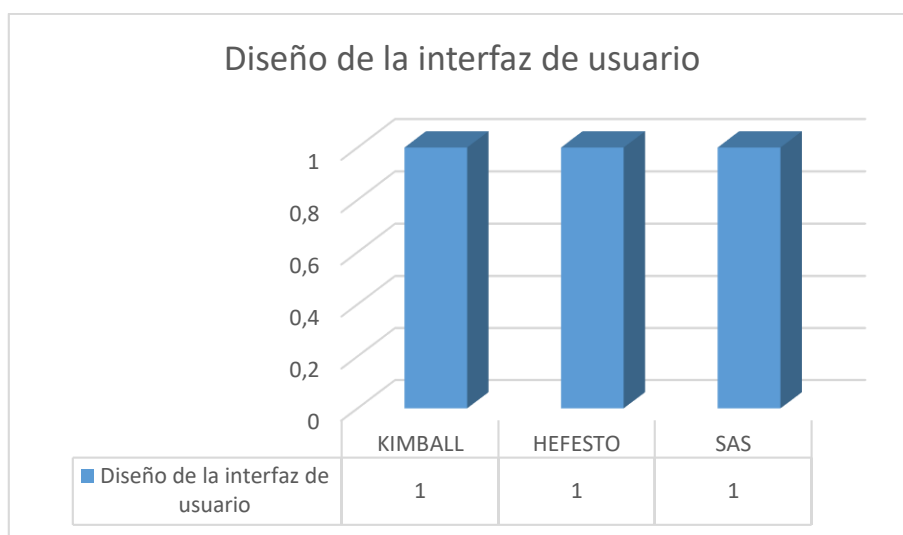
b. Diseño de la interfaz de usuario

Tabla 26: Parámetro. Diseño de la interfaz de usuario

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Diseño de la interfaz de usuario	1	1	1	3

Fuente: elaboración propia

Gráfico 16. Parámetro. Diseño de la interfaz de usuario



Fuente: elaboración propia

Interpretación

El diseño de la interfaz de usuario para las metodologías Kimball, Hefestos y SAS tiene una valoración baja debido a que en ninguna de ellas existe un estándar para el desarrollo de las mismas.

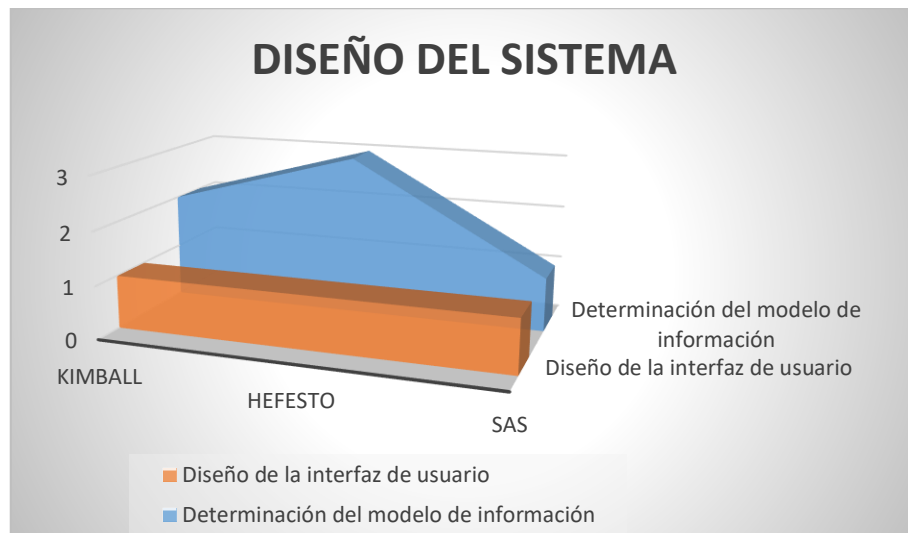
A continuación, se presenta el resumen de la valoración del Indicador Diseño del Sistema.

Tabla 27: Resumen. Indicador diseño del sistema

Indicador	Metodologías desarrollo de <i>Data warehouse</i>		
	Kimball	Hefesto	SAS
Determinación del modelo de información	2	3	1
Diseño de la interfaz de usuario	1	1	1
TOTAL	3	4	2

Fuente: elaboración propia

Gráfico 17. Resumen. Indicador diseño del sistema



Fuente: elaboración propia

5.1.1.6. Fase 6: Elaboración del sistema de información

Tabla 28: Indicador. Fase 6. Elaboración del sistema de información

Indicador	Parámetros de evaluación
Fase 6 Elaboración del sistema de información	Análisis de requerimientos (indicadores)
	Modelo conceptual
	Análisis OLTP
	Conformar indicadores
	Nivel de granularidad
	Nivel conceptual ampliado
	Modelo lógico de la estructura del DW
	Tablas de dimensiones
	Tabla de hechos
	Proceso ETL

Fuente: elaboración propia

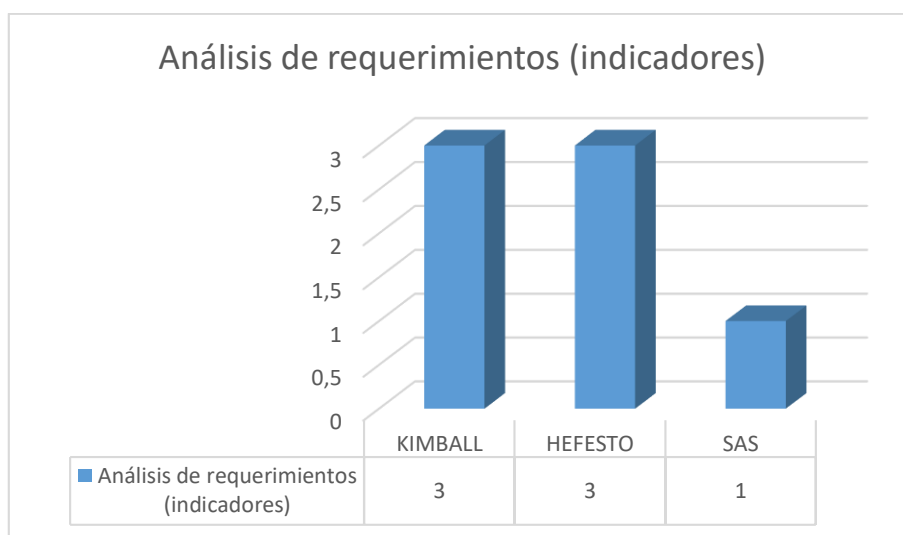
a. Análisis de requerimientos

Tabla 29: Parámetro. Análisis de requerimientos

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Análisis de requerimientos (indicadores)	3	3	1	7

Fuente: elaboración propia

Gráfico 18. Parámetro. Análisis de requerimientos



Fuente: elaboración propia

Interpretación

En la metodología Hefestos el análisis de requerimientos es una de las fases fundamentales para el establecimiento de indicadores y perspectivas, por lo cual recibe la valoración más alta; seguido por Kimball y su definición de indicadores con una valoración media; por último, se encuentra SAS cuya definición de indicadores no se evidencia de una forma clara y concisa.

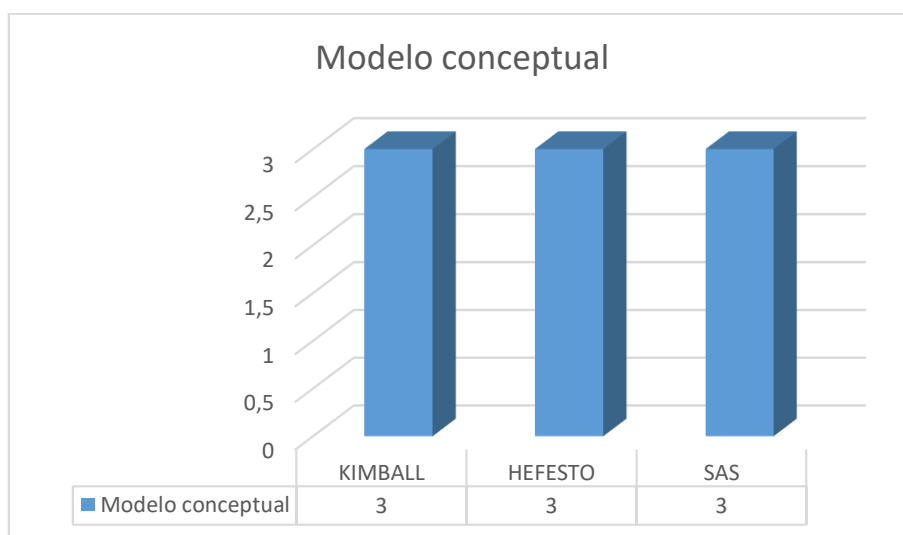
b. Modelo conceptual

Tabla 30: Parámetro. Modelo conceptual

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Modelo conceptual	3	3	3	9

Fuente: elaboración propia

Gráfico 19. Parámetro. Modelo conceptual



Fuente: elaboración propia

Interpretación

Las tres metodologías: Kimball, Hefesto, SAS; cumplen con la definición y modelamiento conceptual, con un amplio diseño de tablas de hechos y dimensiones, diagramas cumpliendo con las especificaciones para tipos de datos y relaciones, los cuales permitirán un mejor establecimiento del *Data Warehouse*; por lo que reciben una valoración alta.

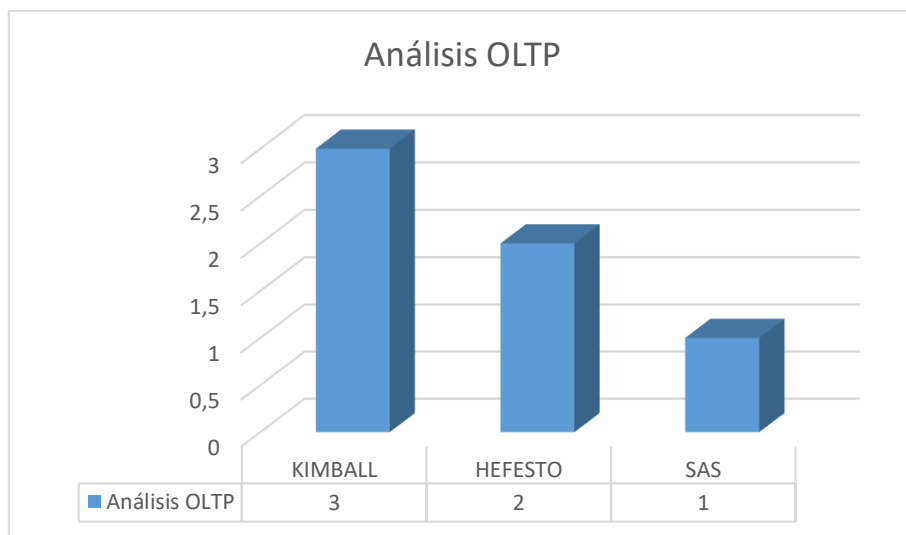
c. Análisis OLTP

Tabla 31: Parámetro. Análisis OLTP

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Análisis OLTP	3	2	1	6

Fuente: elaboración propia

Gráfico 20. Parámetro. Análisis de OLTP



Fuente: elaboración propia

Interpretación

Kimball realiza el análisis del modelo entidad relación la cual es una poderosa técnica de diseño lógico de sistemas, que contribuyen de forma efectiva en el procesamiento de transacciones OLTP, motivo por el que se le ha valorado con una puntuación media, por el contrario, al análisis OLTP que realiza Hefestos determinando los indicadores y estableciendo correspondencias entre el modelo conceptual permitiendo una definición limpia y concreta por lo cual se le asigna una valoración alta. Mientras que para SAS el análisis de OLTP corresponde a un diagrama de infraestructura de la extracción de datos, recibiendo por ello una valoración baja

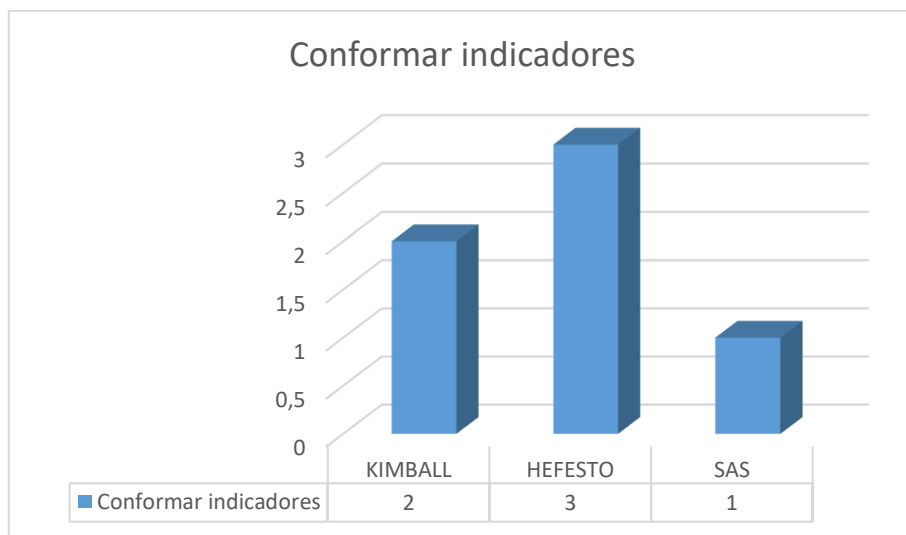
d. Conformar indicadores

Tabla 32: Parámetro. Conformar indicadores

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	KIMBALL	HEFESTO	SAS	
Conformar indicadores	2	3	1	6

Fuente: elaboración propia

Gráfico 21. Parámetro. Conformar indicadores



Fuente: elaboración propia

Interpretación

La conformación de indicadores en la metodología de Hefestos es fundamental para el desarrollo del *Data Warehouse*, Hefestos emplea como parte de su requerimentación la identificación de estos indicadores, proporcionando un análisis preciso de los datos sin la necesidad de reevaluarlos, dándole una valoración alta; Kimball por el contrario primero realiza un análisis de la requerimentación para posteriormente determinar los indicadores por lo que su valoración media; para SAS, los indicadores deben ser analizados justo en la recopilación de la información, haciendo que el proceso de requerimentación sea extenuante, al tratar de incorporar necesidades y proyecciones de resultados, recibiendo así una valoración baja.

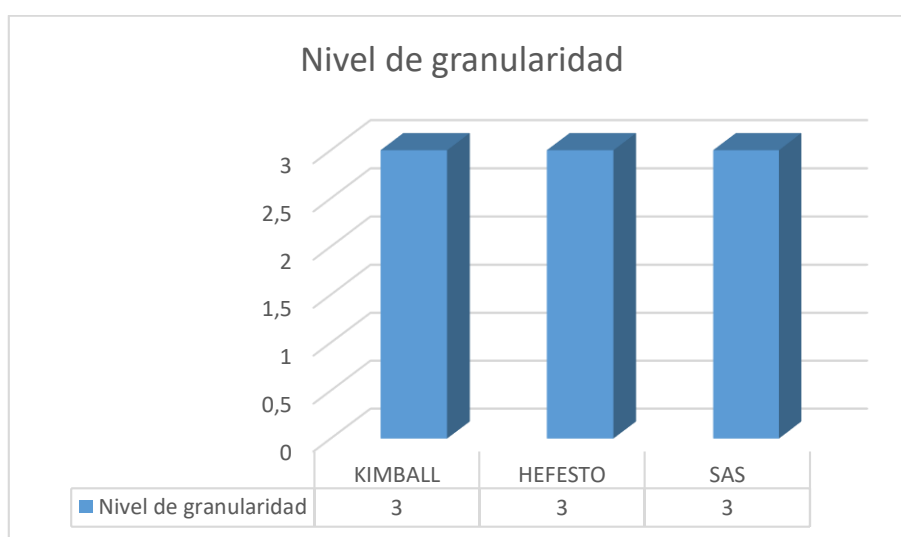
e. Nivel de granularidad

Tabla 33: Parámetro. Nivel de granularidad

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Nivel de granularidad	3	3	3	9

Fuente: elaboración propia

Gráfico 22. Parámetro. Nivel de granularidad



Fuente: elaboración propia

Interpretación

La granularidad es un factor que se debe tener en cuenta para el diseño de las tablas, las dimensiones deben estar basadas en datos con la mayor especificación posible. Es por ello que todas las metodologías realizan un análisis exhaustivo para el establecimiento de este factor, lo que les brinda una valoración alta a cada una de ellas.

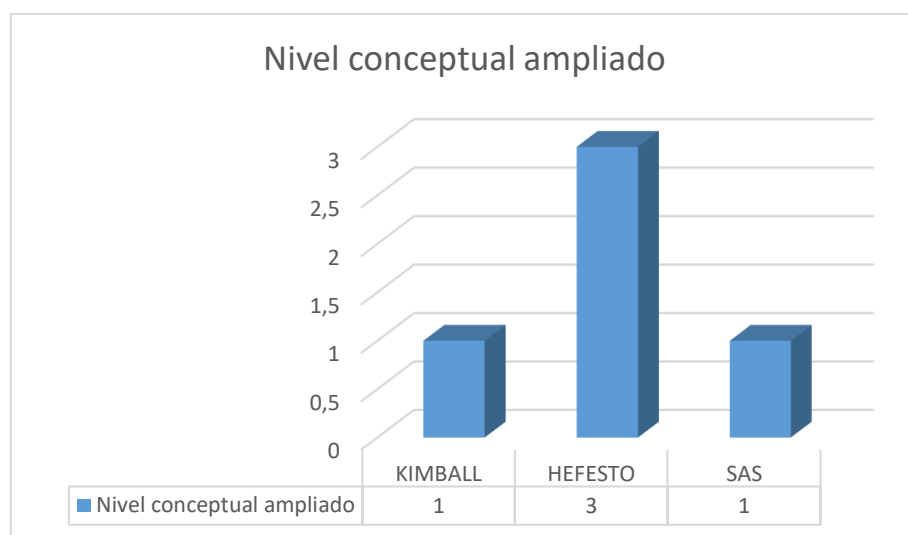
f. Nivel conceptual ampliado

Tabla 34: Parámetro. Nivel conceptual ampliado

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Nivel conceptual ampliado	1	3	1	5

Fuente: elaboración propia

Gráfico 23. Parámetro. Nivel conceptual ampliado



Fuente: elaboración propia

Interpretación

Para Kimball al igual que SAS el modelo conceptual ampliado es sustituido por una definición de un diccionario de datos especializado en el cual se describen las tablas, dimensiones y su relación con el modelo entidad-relación, por lo que reciben en este parámetro una valoración baja, en el caso de Hefestos el modelo conceptual ampliado es necesario para establecer las dimensiones y hechos para el diseño del *Data Warehouse*, lo cual provee a esta metodología una valoración alta.

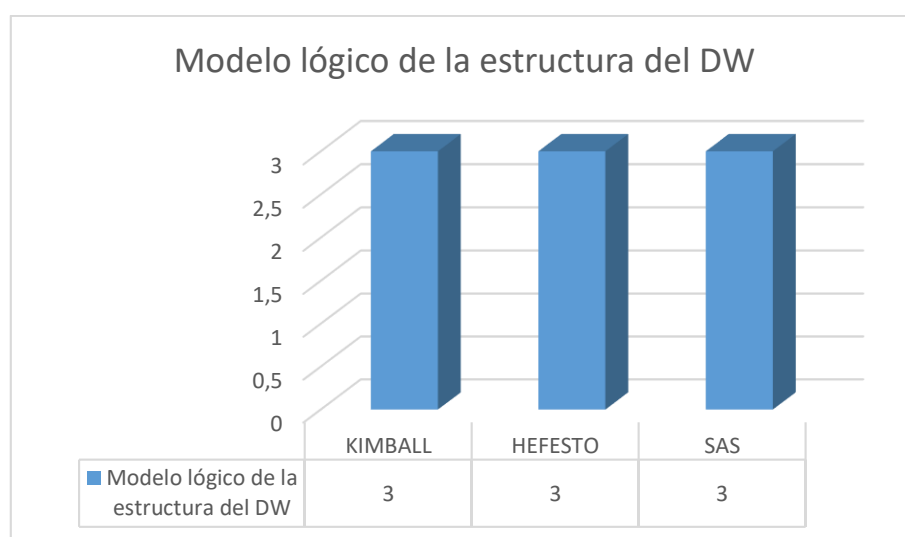
g. Modelo lógico de la estructura del Data Warehouse

Tabla 35: Parámetro. Modelo lógico de la estructura del DW

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	KIMBALL	HEFESTO	SAS	
Modelo lógico de la estructura del DW	3	3	3	9

Fuente: elaboración propia

Gráfico 24. Parámetro. Modelo lógico de la estructura del DW



Fuente: elaboración propia

Interpretación

Las metodologías presentadas describen de manera detallada el modelo lógico del *Data Warehouse*, definiendo claramente tipos de datos, columnas, tablas y relaciones que deben ser considerados para la definición de la base de datos.

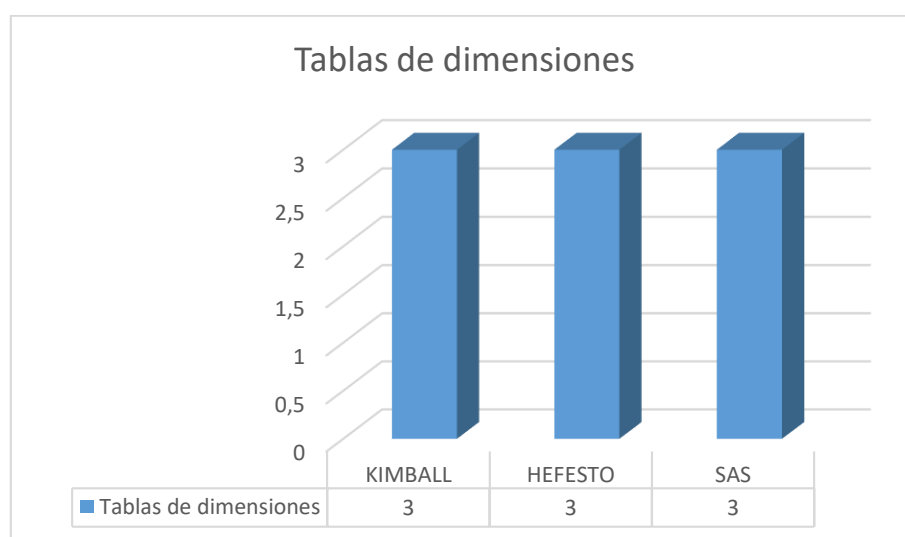
h. Tablas de dimensiones

Tabla 36: Parámetro. Tablas de dimensiones

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Tablas de dimensiones	3	3	3	9

Fuente: elaboración propia

Gráfico 25. Parámetro. Tablas de dimensiones



Fuente: elaboración propia

Interpretación

La definición de tablas de dimensiones para las diferentes metodologías es bastante amplia, cada uno de ellas describe de diferente manera las especificaciones para el establecimiento de las dimensiones. Por lo cual reciben una valoración alta.

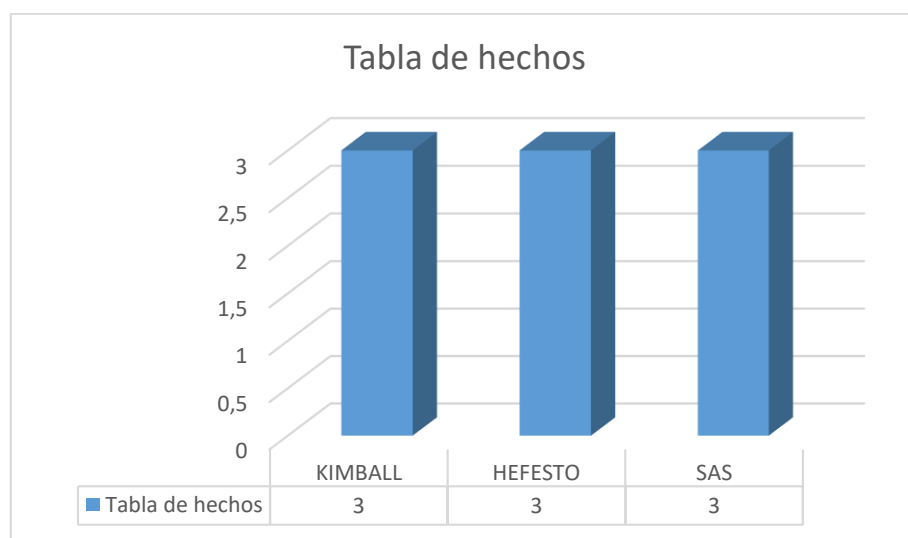
i. Tablas de hechos

Tabla 37: Parámetro. Tablas de hechos

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Tabla de hechos	3	3	3	9

Fuente: elaboración propia

Gráfico 26. Parámetro. Tablas de hechos



Fuente: elaboración propia

Interpretación

La definición de las tablas de hechos para las metodologías: Kimball, Hefesto y SAS son detalladas, su proceso de obtención es diferente a pesar de estas diferencias la valoración para este indicador es alta.

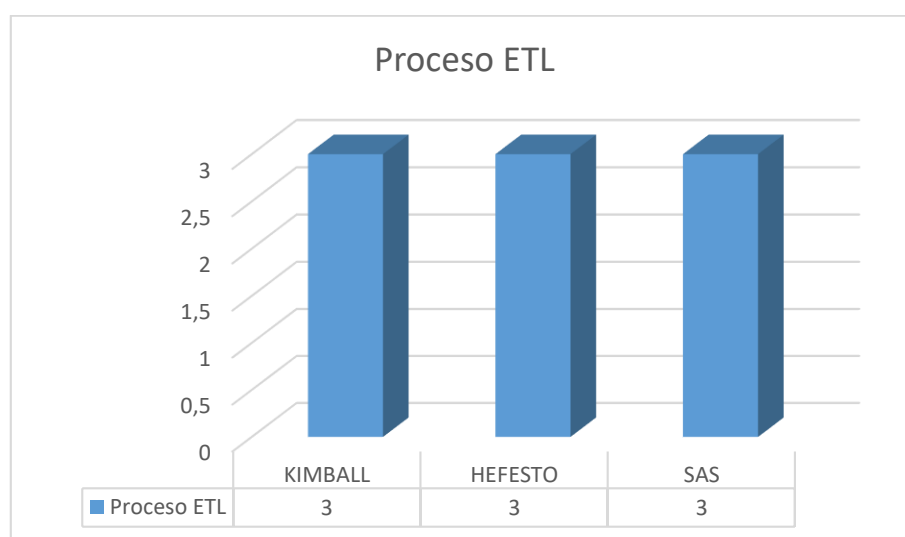
j. Proceso ETL

Tabla 38: Parámetro. Proceso ETL

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Proceso ETL	3	3	3	9

Fuente: elaboración propia

Gráfico 27. Parámetro. Proceso ETL



Fuente: elaboración propia

Interpretación

Para la determinación del proceso ETL en cada una de las metodologías planteadas se realiza un minucioso análisis para determinar los procesos adecuados para la extracción, transferencia y carga de datos, este indicador recibe una valoración alta.

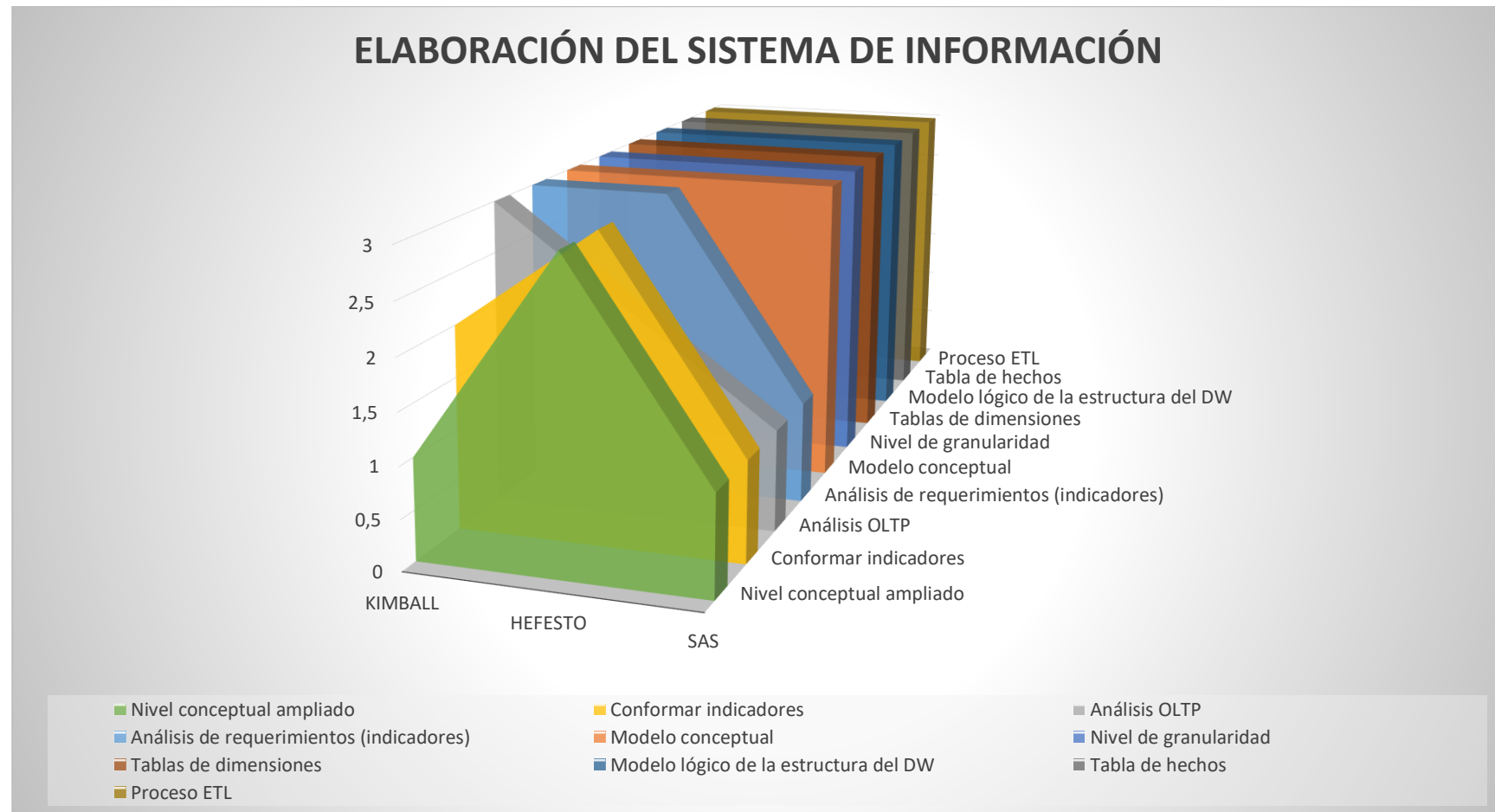
A continuación, se presenta el resumen de la valoración del Indicador Elaboración del sistema de información.

Tabla 39: Resumen. Indicador elaboración del sistema de información

Indicador	Metodologías desarrollo de <i>Data Warehouse</i>		
	Kimball	Hefesto	SAS
Análisis de requerimientos (indicadores)	3	3	1
Modelo conceptual	3	3	3
Análisis OLTP	3	2	1
Conformar indicadores	2	3	1
Nivel de granularidad	3	3	3
Nivel conceptual ampliado	1	3	1
Modelo lógico de la estructura del DW	3	3	3
Tablas de dimensiones	3	3	3
Tabla de hechos	3	3	3
Proceso ETL	3	3	3
TOTAL	27	29	22

Fuente: Elaboración propia

Gráfico 28. Resumen. Indicador elaboración del sistema de información



Fuente: elaboración propia

5.1.2. Determinación de parámetros de comparación de características metodológicas

5.1.2.1. Características

El análisis de las metodologías de desarrollo de un *Data Warehouse* debe ser complementada mediante un estudio de sus características generales, tomadas de Análisis, Diseño, Construcción e Implementación de un *Data Warehouse* para Toma de Decisiones y Construcción De Los KPI, para la Empresa Kronosconsulting Cia Ltda.; de acuerdo al desarrollo del proyecto sobre el cual es implementado, estas mediciones facilitan una valoración estandarizada por lo que nos permite su análisis, de acuerdo a los parámetros de evaluación pre establecidos; para la valoración de las diferentes magnitudes tales como: el grado del cumplimiento de los objetivos, la capacidad de mejoras y del grado de satisfacción de los directivos.

Los indicadores serán utilizados de acuerdo con las necesidades y posibilidades de la institución; de esta forma se realiza la evaluación sobre los siguientes:

Tabla 40: Indicador. Características metodológicas

Indicador	Parámetros de evaluación
Características	Adaptable sobre cualquier tecnología
	Afinidad con el sistema actual
	Comunicación con el cliente
	Tamaño del proyecto
	Tiempo en análisis y diseño
	Tiempo en construcción
	Fácil entendimiento principiantes
	Documentación precisa
	Más usada en el mundo

Fuente: (Lozada X & Cruz, 2014)

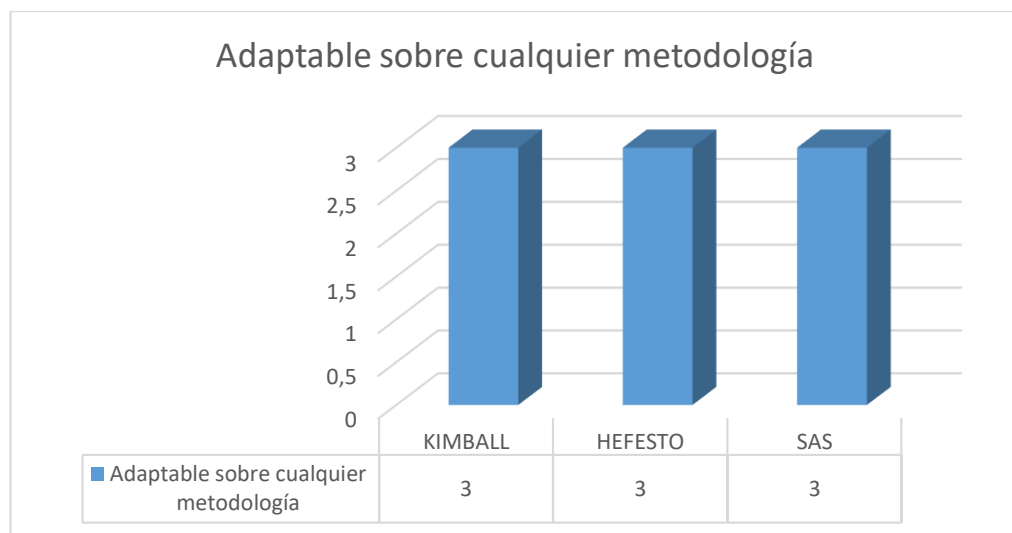
- a. Adaptable sobre cualquier metodología

Tabla 41: Parámetro. Adaptable sobre cualquier metodología

Parámetro de evaluación	Metodologías desarrollo de Data Warehouse			Total
	Kimball	Hefesto	SAS	
Adaptable sobre cualquier metodología	3	3	3	9

Fuente: elaboración propia

Gráfico 29. Parámetro. Adaptable sobre cualquier metodología



Fuente: elaboración propia

Interpretación

Debido a la simplicidad del proyecto y la construcción de las colecciones de datos podemos observar que la solución es de fácil adaptabilidad para cualquiera de las metodologías analizadas en este estudio comparativo.

Es por ello que, Kimball, Hefesto y SAS obtienen una valoración alta sobre este indicador.

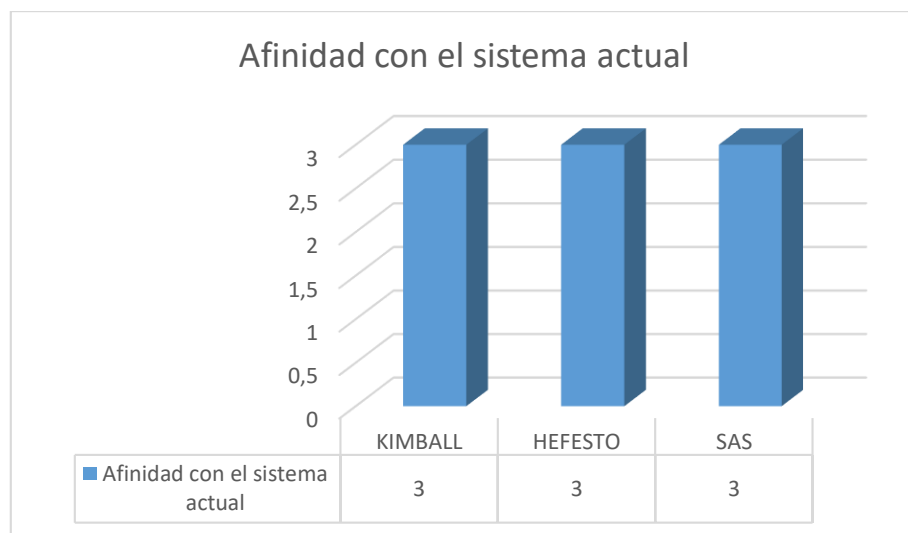
b. Afinidad con el sistema actual

Tabla 42: Parámetro. Afinidad con el sistema actual

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Afinidad con el sistema actual	3	3	3	9

Fuente: elaboración propia

Gráfico 30. Parámetro. Afinidad con el sistema actual



Fuente: elaboración propia

Interpretación

El INPC R3 actualmente maneja su sistema de información por medio de un sistema relación en una base de datos MySQL y hojas electrónicas en Excel; por lo que se ha tomado en cuenta el desarrollo de la solución en el mismo gestor de base de datos, por lo cual su afinidad en cuanto al sistema actual recibe una valoración alta.

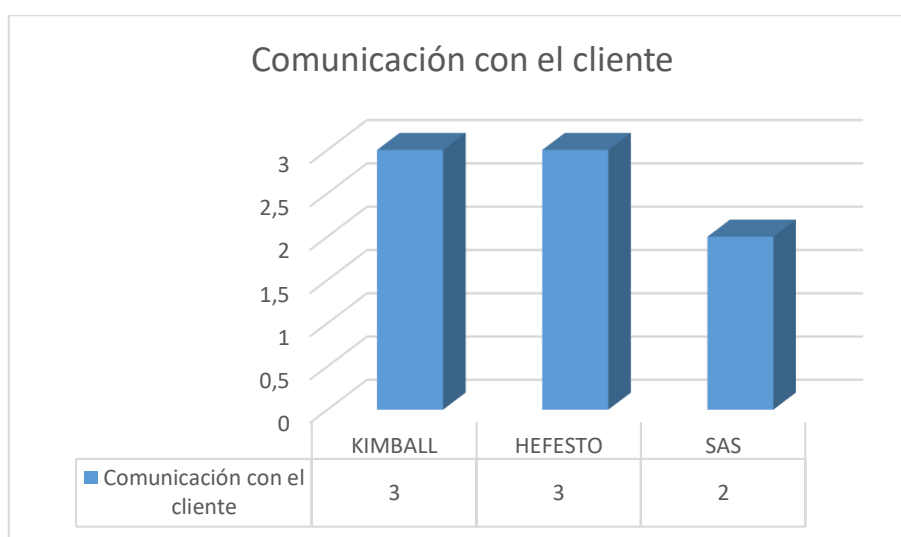
c. Comunicación con el cliente

Tabla 43: Parámetro. Comunicación con el cliente

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Comunicación con el cliente	3	3	2	8

Fuente: elaboración propia

Gráfico 31. Parámetro. Comunicación con el cliente



Fuente: elaboración propia

Interpretación

Las metodologías de Kimball y Hefestos proveen un alto nivel de comunicación entre el cliente y el desarrollador de la solución, puesto que en la fase de requerimentación existe una interacción directa con el cliente, a través de encuestas, entrevistas y dialogo directo con los directivos de la organización, recibiendo una valoración alta; mientras SAS realiza varias intervenciones con el personal operativo de la solución, dejando de lado los objetivos que persigue la entidad, recibiendo una valoración media.

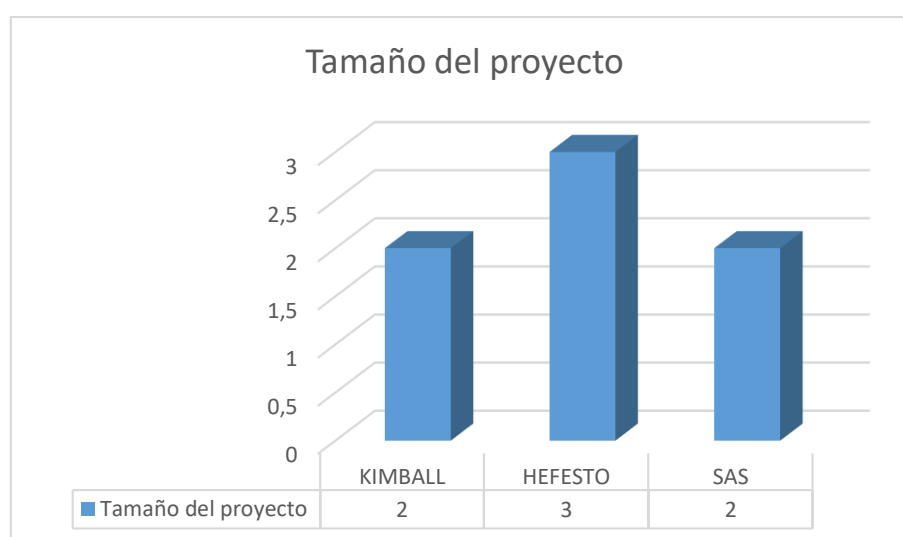
d. Tamaño del proyecto

Tabla 44: Parámetro. Tamaño del proyecto

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Tamaño del proyecto	2	3	2	7

Fuente: elaboración propia

Gráfico 32. Parámetro. Tamaño del proyecto



Fuente: elaboración propia

Interpretación

Debido a la definición de Hefestos sobre el tamaño del proyecto, es el más apto para el desarrollo de la solución planteada para el INPC R3, puesto que abarca el desarrollo de soluciones pequeñas y medianas; por lo cual recibe una valoración alta; mientras que Kimball y SAS están dedicadas para proyectos grandes, por lo cual se descartan para la implementación del mismo, debido al desperdicio de recursos, recibiendo así una valoración media.

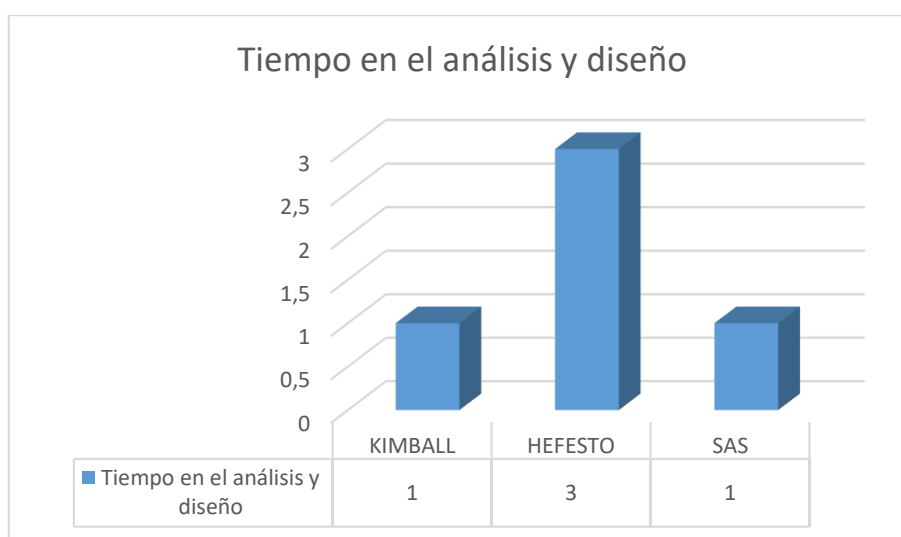
- e. Tiempo en el análisis y diseño

Tabla 45: Parámetro. Tiempo en el análisis y diseño

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Tiempo en el análisis y diseño	1	3	1	5

Fuente: elaboración propia

Gráfico 33. Parámetro. Tiempo en el análisis y diseño



Fuente: elaboración propia

Interpretación

Debido al costo excesivo que genera el mantener los recursos por ser metodologías iterativas, Kimball Y SAS reciben una valoración baja, mientras que Hefestos para esta solución es la más indicada debido que se realiza una sola iteración en la cual se abarcan las necesidades de directivos y operativos sobre el proyecto, recibiendo así una valoración alta.

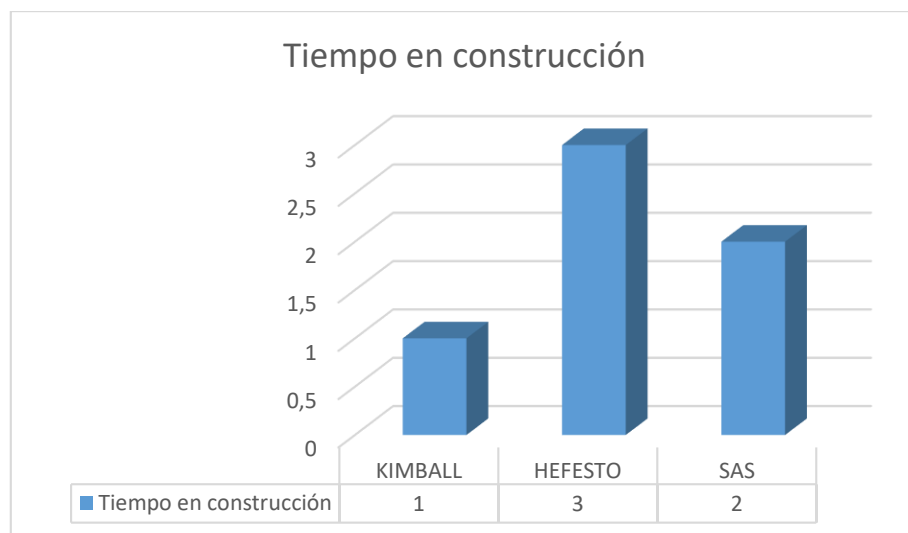
f. Tiempo en construcción

Tabla 46: Parámetro. Tiempo en el análisis y diseño

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Tiempo en construcción	1	3	2	6

Fuente: elaboración propia

Gráfico 34. Parámetro. Tiempo en construcción



Fuente: elaboración propia

Interpretación

Para la construcción de la solución se debe considerar los costos que cada metodología genera, es por ello que siendo Kimball una metodología costosa recibe una puntuación baja; mientras SAS por su sencillez recibe una valoración media, por las diferentes iteraciones que deben ser realizadas para la generación del proyecto; siento la metodología de Hefestos la más adecuada ya que permite la reducción de costos debido a su única iteración, con una valoración alta.

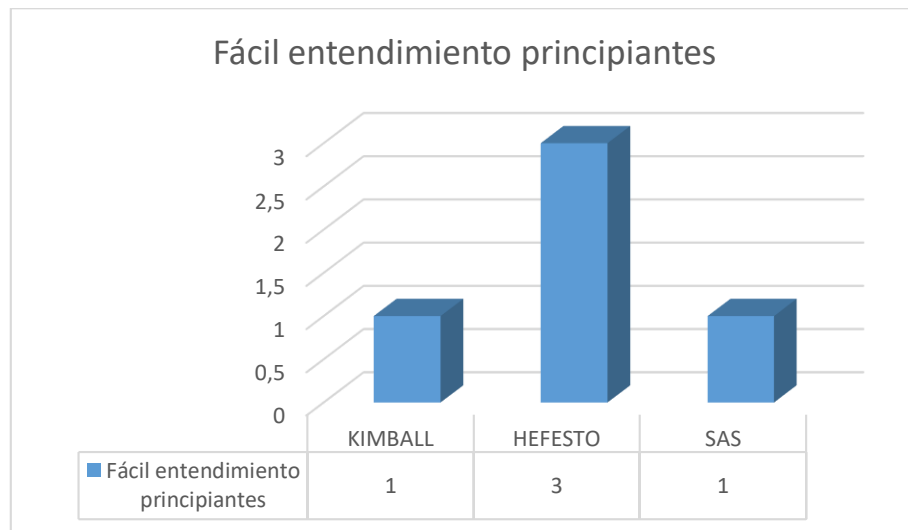
g. Fácil entendimiento en principiantes

Tabla 47: Parámetro. Fácil entendimiento en principiantes

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Fácil entendimiento principiantes	1	3	1	5

Fuente: elaboración propia

Gráfico 35. Parámetro. Fácil entendimiento principiantes



Fuente: elaboración propia

Interpretación

Debido a la facilidad de comprensión por parte de los desarrolladores sobre la metodología Hefestos y las guías prácticas que esta proporciona permite asignarle una valoración alta en este indicador; mientras que la curva de aprendizaje para SAS y Kimball es elevada, por lo que su entendimiento requiere de un mayor tiempo, por lo tanto, recibe una valoración baja

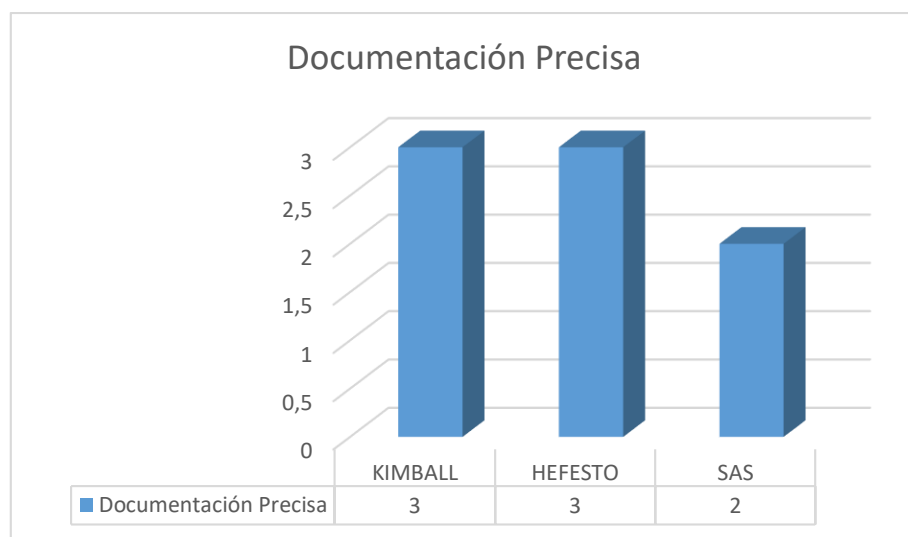
h. Documentación precisa

Tabla 48: Parámetro. Fácil entendimiento en principiantes

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Documentación Precisa	3	3	2	8

Fuente: elaboración propia

Gráfico 36. Parámetro. Documentación precisa



Fuente: elaboración propia

Interpretación

La documentación proporcionada por la metodología SAS presenta un grado intermedio de especificación debido a misma definición de una metodología ágil, por lo cual recibe una valoración media, ya que el proyecto necesita de una guía técnica precisa para su desarrollo, implementación y uso, al momento de la construcción de la solución; por otro lado, Kimball y Hefestos mantienen un alto estándar, detallando paso a paso la construcción de la solución

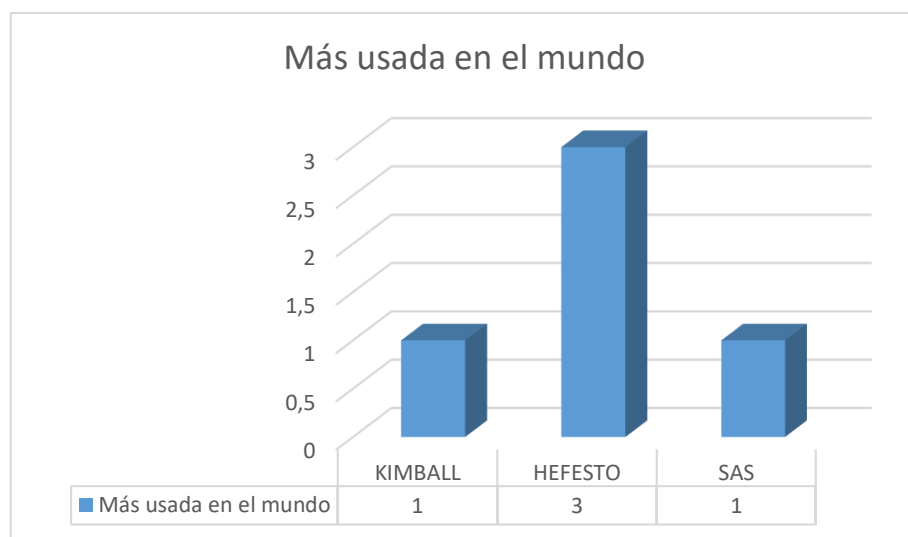
- i. Más usado en el mercado

Tabla 49: Parámetro. Más usado en el mercado

Parámetro de evaluación	Metodologías desarrollo de <i>Data Warehouse</i>			Total
	Kimball	Hefesto	SAS	
Más usada en el mundo	1	3	1	5

Fuente: elaboración propia

Gráfico 37. Parámetro. Más usada en el mundo



Fuente: elaboración propia

Interpretación

La metodología de Kimball ha perdido seguidores en el mercado debido a que no existe una retroalimentación que esté acorde al enfoque tecnológico basado a futuros proyectos; recibiendo una valoración baja, al igual que SAS debido a su desconocimiento de las prestaciones por su modo ágil de desarrollo. Por lo cual ha tomado la posta Hefestos, recibiendo una valoración alta.

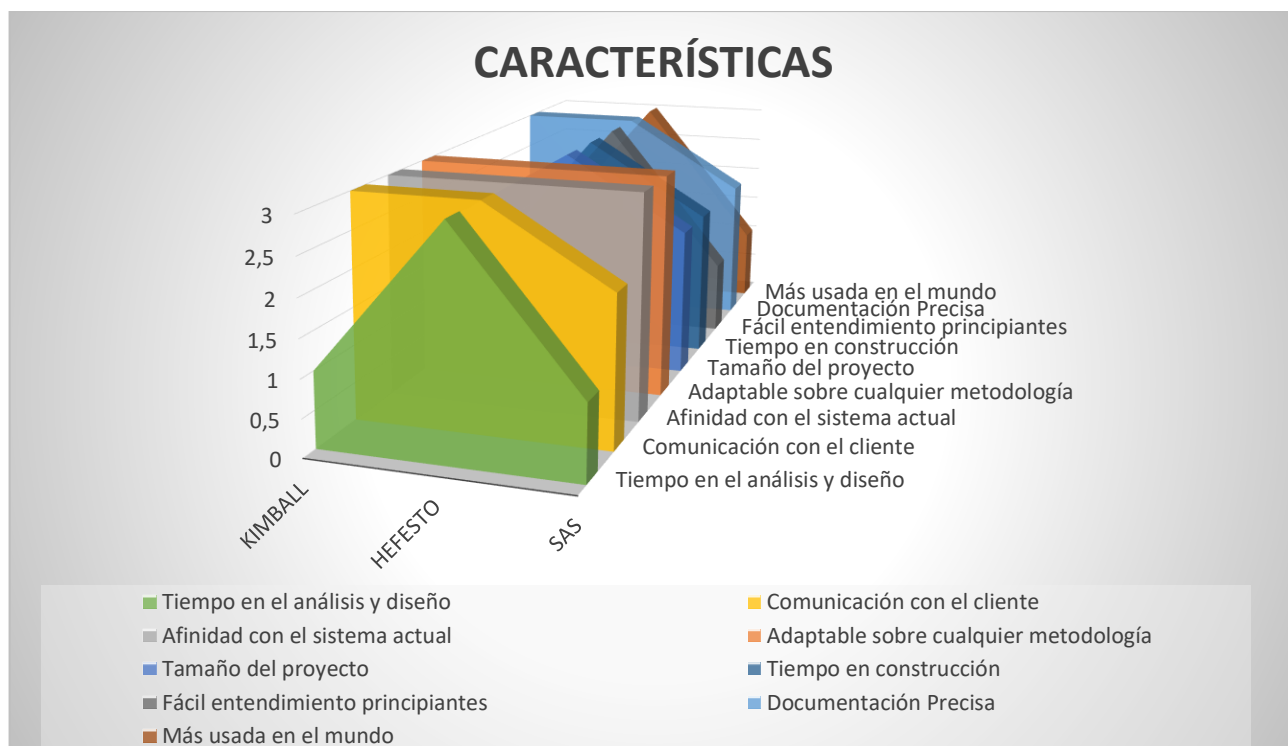
A continuación, se presenta el resumen de la valoración del Indicador Características metodológicas

Tabla 50: Resumen. Indicador características metodológicas

Indicador	Metodologías desarrollo de <i>Data Warehouse</i>		
	Kimball	Hefesto	SAS
Adaptable sobre cualquier metodología	3	3	3
Afinidad con el sistema actual	3	3	3
Comunicación con el cliente	3	3	2
Tamaño del proyecto	2	3	2
Tiempo en el análisis y diseño	1	3	1
Tiempo en construcción	1	3	2
Fácil entendimiento principiantes	1	3	1
Documentación Precisa	3	3	2
Más usada en el mundo	1	3	1
TOTAL	18	27	17

Fuente: elaboración propia

Gráfico 38. Parámetro. Características metodológicas



Fuente: elaboración propia

5.2. Análisis de resultados

5.2.1. Resultados de comparativas fases metodológicas

Al analizar las metodologías Kimball, Hefestos y SAS con cada uno de los parámetros de comparación de las fases metodológicas se observa que la metodología Hefestos cumple con la mayoría de los índices tomados en cuenta de cada uno de los indicadores expuestos en las 7 fases examinadas obteniendo el primer lugar con un total de 54 puntos en la evaluación; mientras que la metodología Kimball cumple parcialmente con los indicadores obteniendo 52 puntos; por otro lado SAS obtiene una puntuación baja con 42 puntos, por lo que la metodología con la menor posibilidad de uso para el desarrollo.

5.2.2. Resultados de comparativas características metodológicas

Al analizar las características con las métricas correspondientes a las metodologías Kimball, Hefestos y SAS se observa que la metodología Hefestos brinda una mayor cohesión en la construcción del *Data Warehouse* cumpliendo con el mayor porcentaje de los indicadores establecidos para dicha evaluación haciéndola la metodología con mayor valoración al obtener 29 puntos; en segundo lugar, tenemos la metodología Kimball con 27 puntos la cual no cumple con algunas de las características valoradas; mientras que la metodología SAS se encuentra en el último lugar debido al poco entendimiento al crear este tipo de soluciones.

5.3. Desarrollo e implementación del *Data Warehouse* con la metodología Hefestos

En 2017 el Instituto Nacional de Patrimonio Cultural se consolidará como un centro generador de conocimiento sobre los elementos que constituyen el sustento de nuestra identidad nacional multiétnica y pluricultural. Se fundamentará en la investigación formal, en la conservación y en la difusión multifacética del patrimonio cultural y de la memoria histórica del Ecuador. Este caudal será difundido para concienciar a los actores involucrados y a la ciudadanía sobre la importancia de las raíces y de la identidad. Así, mediante el uso y desarrollo de teorías, métodos y técnicas tanto inter como transdisciplinarias, el INPC R3 facilitará el acceso, el uso social y la preservación de todo el acervo cultural de la Nación.

La aplicación de la metodología Hefesto para el diseño del *data warehouse* se realizó en el INPC R3, su conjunto documental está constituido por diversos soportes digitalizados y bases de datos con los que cuenta la entidad.

5.3.1. Análisis de requerimientos

5.3.1.1. Identificar preguntas

Mediante las entrevistas realizadas a los usuarios se ha determinado la creación de un OLTP, que permita el tratamiento del proceso de control de trámites.

Al identificar las variables o perspectivas a utilizar para la toma de decisiones mediante las respuestas de los usuarios se determinando cuales indicadores representan mejor este proceso de trámites.

Las preguntas obtenidas fueron las siguientes:

- Se requiere conocer el número de trámites receptados en un periodo de tiempo.
- Se requiere conocer el número de trámites por remitente y dependencia en un periodo de tiempo.
- Se requiere conocer el personal asignado durante el proceso de duración del trámite.
- Se requiere conocer el número de trámites por estado en un periodo de tiempo.
- Se requiere conocer el número de trámites por procedencia geográfica a ser procesados.
- Se requiere conocer el número de trámites por prioridad.
- Se requiere conocer el número de días transcurridos hasta la finalización del trámite.
- Se requiere conocer el número de respuestas por medio de entrega en un periodo de tiempo

5.3.1.2. Identificar indicadores y perspectivas

Tabla 51: Indicadores y perspectivas

Indicadores	Perspectivas
Trámites	Tiempo
Respuestas	Localidad
Número de días	Remitente
	Persona asignada
	Área
	Medio de entrega
	Estado
	Tipo de trámite

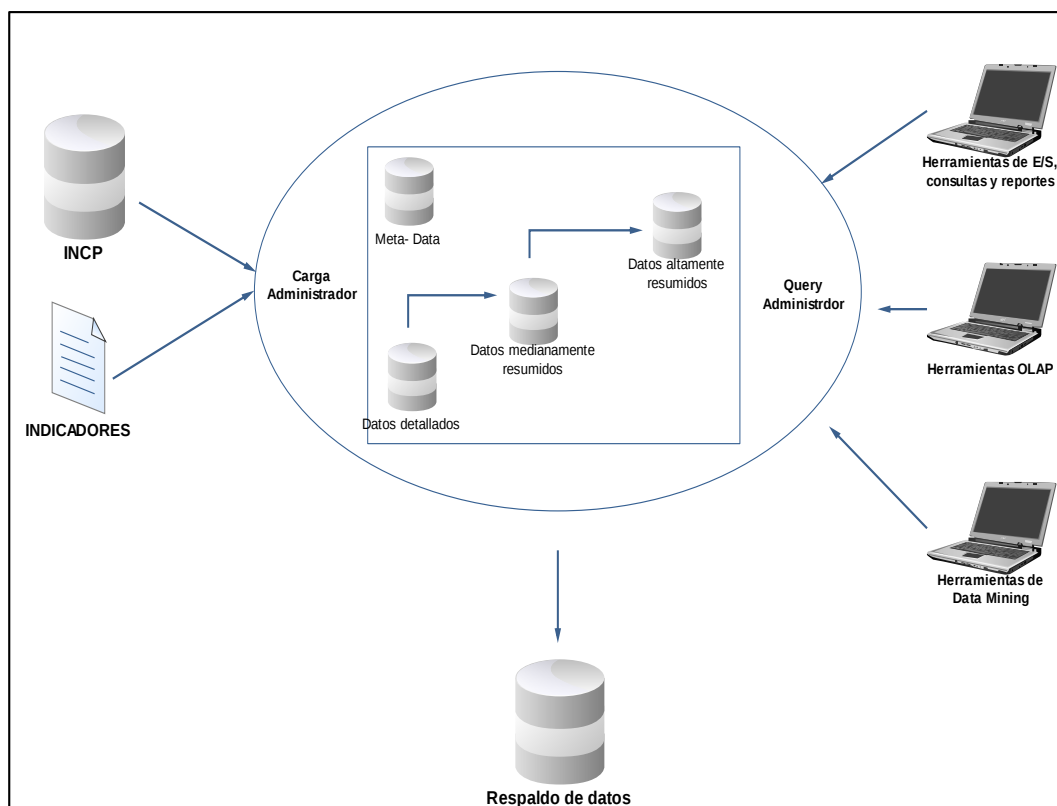
Fuente: elaboración propia

5.3.2. Análisis de los OLTP

5.3.2.1. Fuentes de datos

Representa los orígenes de los datos, una base de datos transaccional llamada inpc.

Figura 26. Fuentes de datos inpc

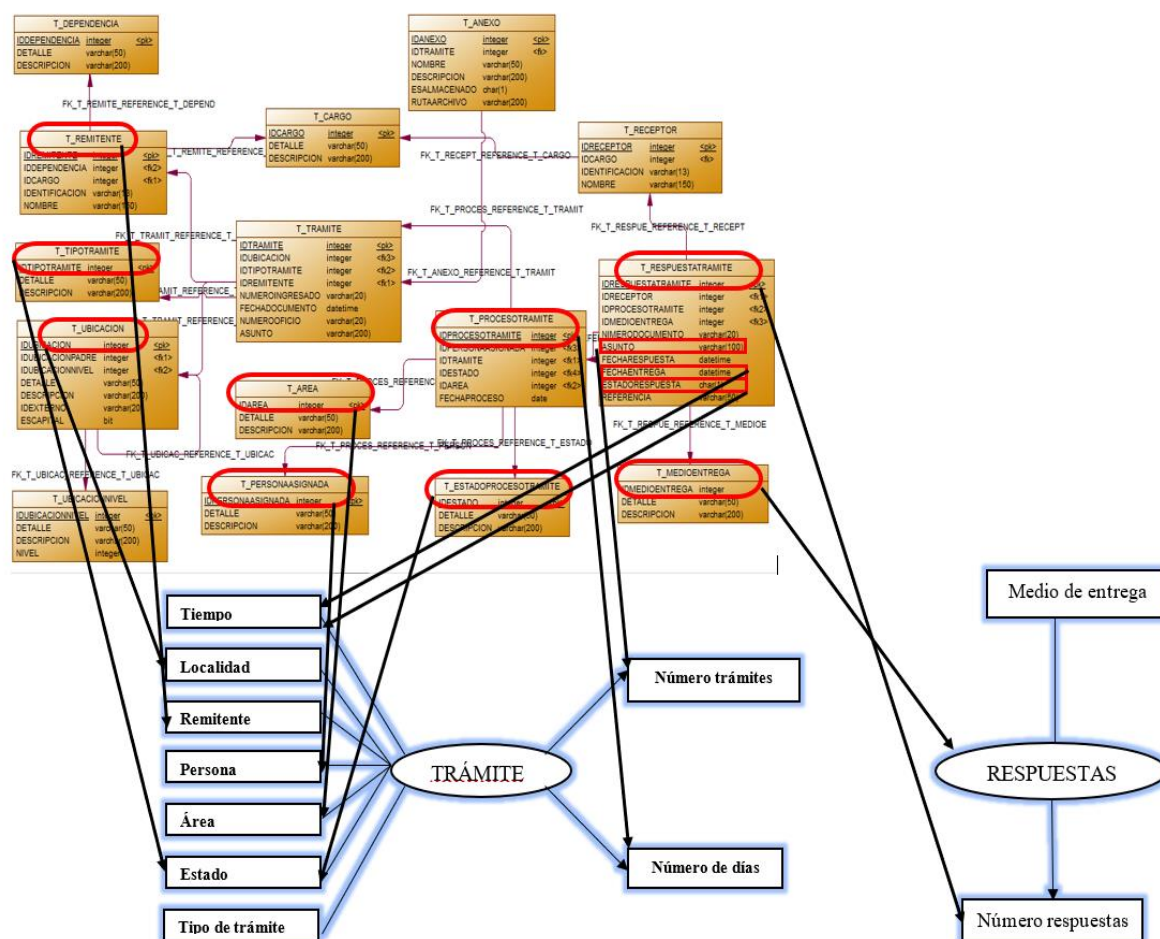


Fuente: elaboración propia

5.3.2.2. Modelo conceptual

El diseño de los modelos, parten de los indicadores y perspectivas obtenidas, los cuales permiten la observación de los alcances, este proceso sistemático se detalla en el apéndice B, en la sección del modelo conceptual, en la figura 27 se representa el modelo conceptual resultante para el INPC R3.

Figura 27. Modelo conceptual para el inpc



Fuente: elaboración propia

5.3.2.3. Modelo lógico del DW

Después de analizar los procesos que realiza el INPC R3 y las fuentes existentes, se procede con la construcción del modelo lógico de la estructura del *Data warehouse*, se apoyará en el modelo conceptual ampliado ubicado en el apéndice B en la sección del modelo conceptual ampliado, para el diseño de las tablas de dimensiones y hechos.

Figura 28. Modelo lógico de los procesos trámites



Fuente: elaboración propia

5.3.3. Integración de datos

El último paso que propone la metodología Hefestos es la integración de datos en el cual procedemos a ingresar los datos, para lo cual se usará métodos de limpieza y calidad de datos como ETL; para puntualizar reglas y políticas para actualización, este proceso se detalla en el apéndice B, en la sección de integración de datos.

5.3.4. Implementación de la data warehouse para el INPC R3

Una vez diseñado el modelo lógico y conceptual del DW, se procede a poblar el DW a través de la herramienta correspondiente.

5.3.4.1. Selección de la herramienta tecnológica

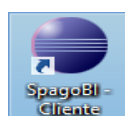
Tabla 52: Herramientas tecnológicas

Herramienta	Descripción
Base de datos relacional	MySQL; <i>Power Designer</i> para el diseño del <i>Data Mart</i> y definición de metadatos.
<i>Data Mart</i>	La construcción el <i>Data Mart</i> se lo realizará en MySQL; para la extracción de los datos y la generación de cubos multidimensionales se usará <i>SpagoBI</i>
Reportes	Para reportes y construcción de gráficos se utilizará <i>Power BI</i> .

Fuente: elaboración propia

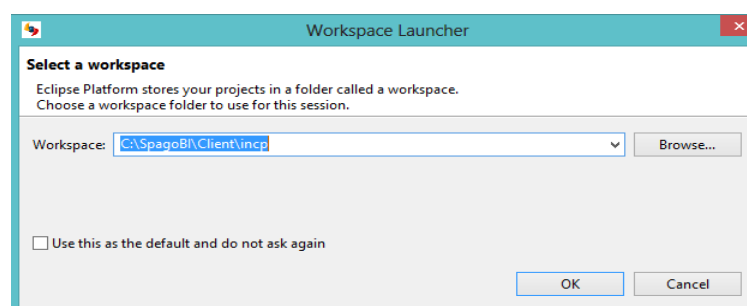
5.3.4.2. Implementación del componente de información - Instalación de *SpagoBI Studio*

1. Ingrese al instalador de *SpagoBI Studio*



2. Seleccionar la ubicación del *workspace* para el proyecto

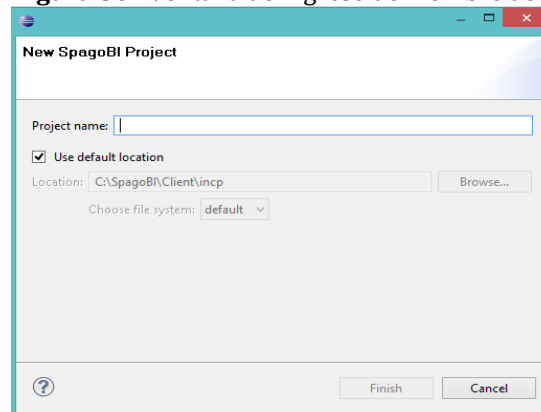
Figura 29. Ventana workspace launcher



Fuente: elaboración propia

3. Crear el proyecto y se digita el nombre del mismo

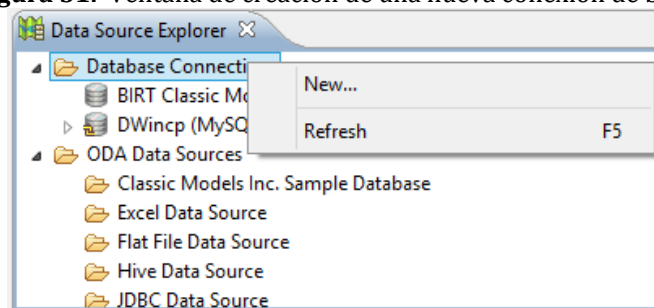
Figura 30. Ventana de ingreso del nombre del proyecto



Fuente: elaboración propia

4. Crear un *data source*, con click derecho sobre *Database Connection* -> *New*

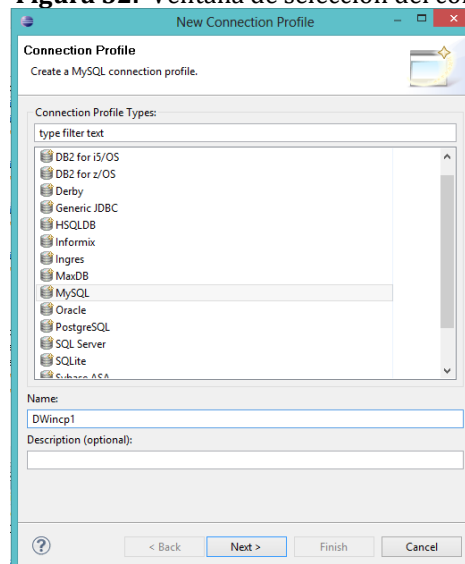
Figura 31. Ventana de creación de una nueva conexión de base de datos



Fuente: elaboración propia

5. Seleccionamos el controlador, en este caso es MySQL y se da clic en siguiente.

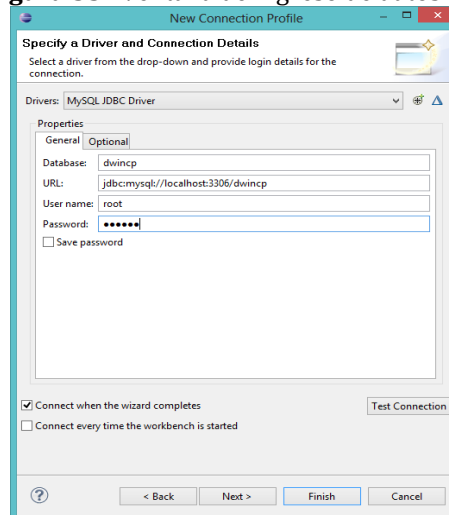
Figura 32. Ventana de selección del controlador



Fuente: elaboración propia

- Colocamos un nombre a la conexión y las credenciales para acceder a la base de datos

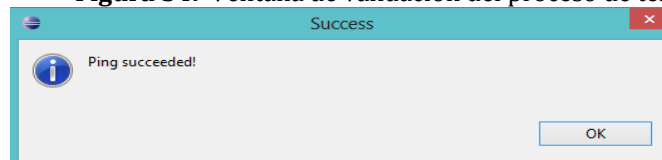
Figura 33. Ventana de ingreso de datos de conexión



Fuente: elaboración propia

- Testeamos la conexión

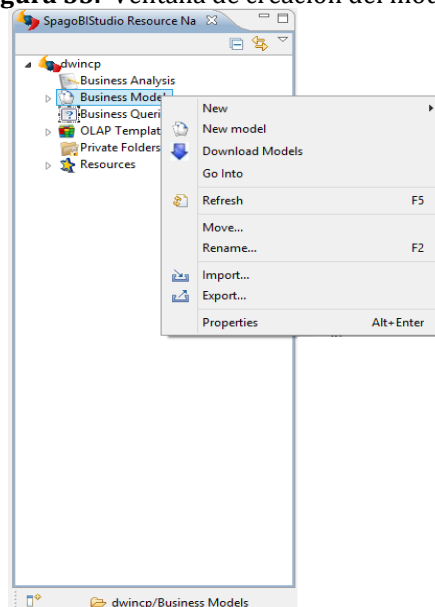
Figura 34. Ventana de validación del proceso de testeo



Fuente: elaboración propia

- Crear el *data model* para el Proyecto (*Business Model*)

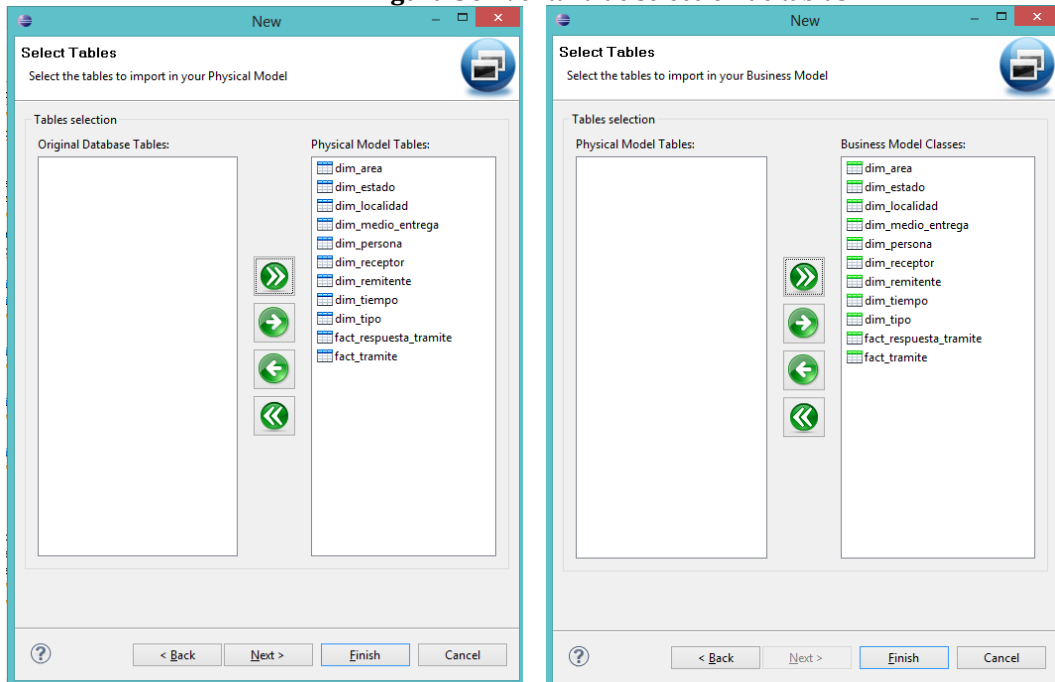
Figura 35. Ventana de creación del modelo de datos



Fuente: elaboración propia

9. Seleccionar las tablas para el modelo, este proceso de debe escoger las tablas que son útiles para la creación del modelo, este proceso se lo realiza dos veces.

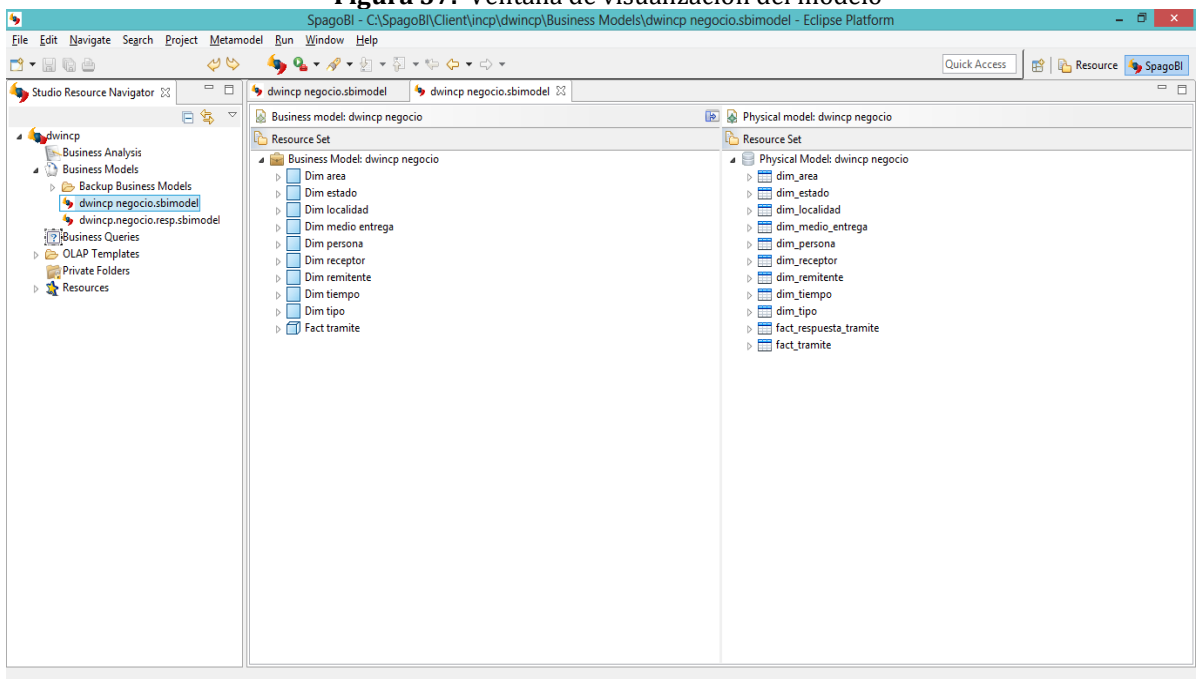
Figura 36. Ventana de selección de tablas



Fuente: elaboración propia

10. El modelo se conformará de la siguiente manera

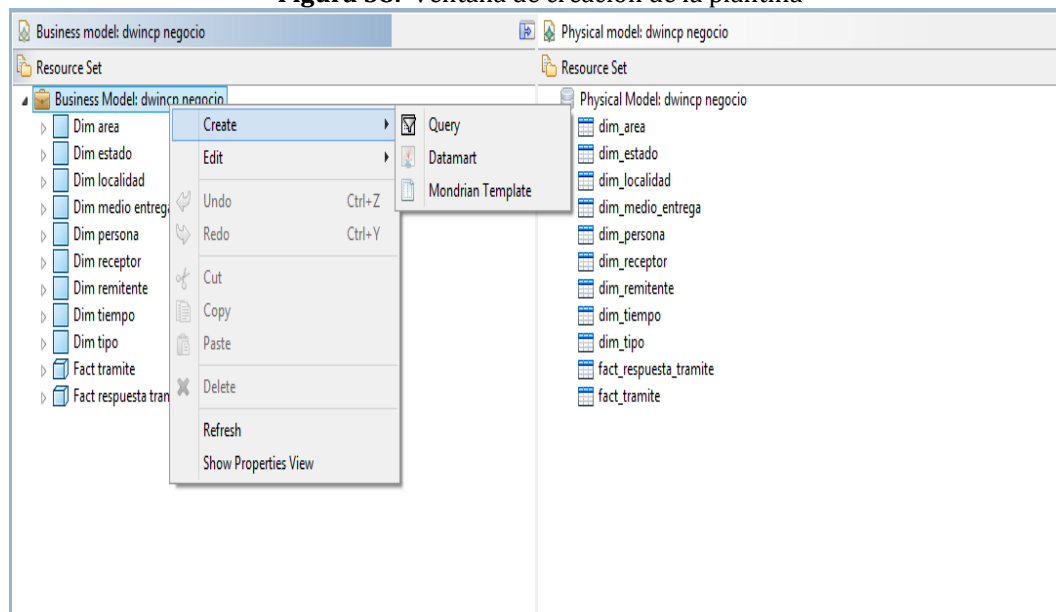
Figura 37. Ventana de visualización del modelo



Fuente: elaboración propia

11. Crear Mondrian Template

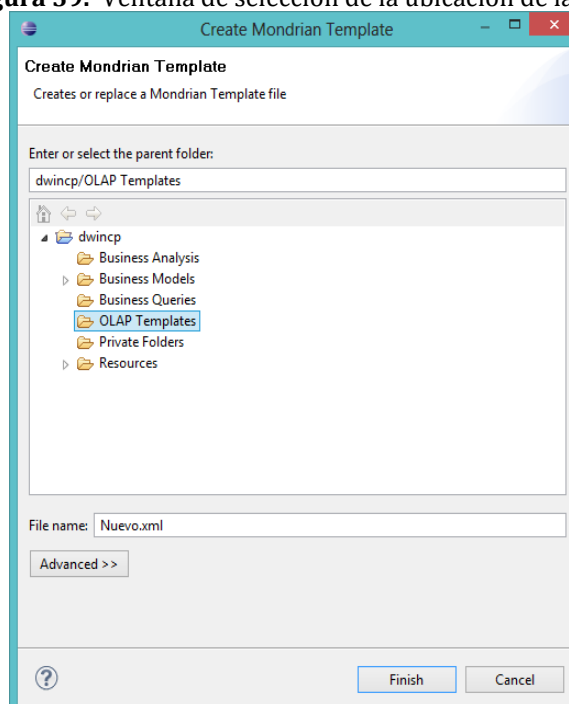
Figura 38. Ventana de creación de la plantilla



Fuente: elaboración propia

12. Seleccionar la ubicación del *template*, se lo ubica en los *templates* OLAP

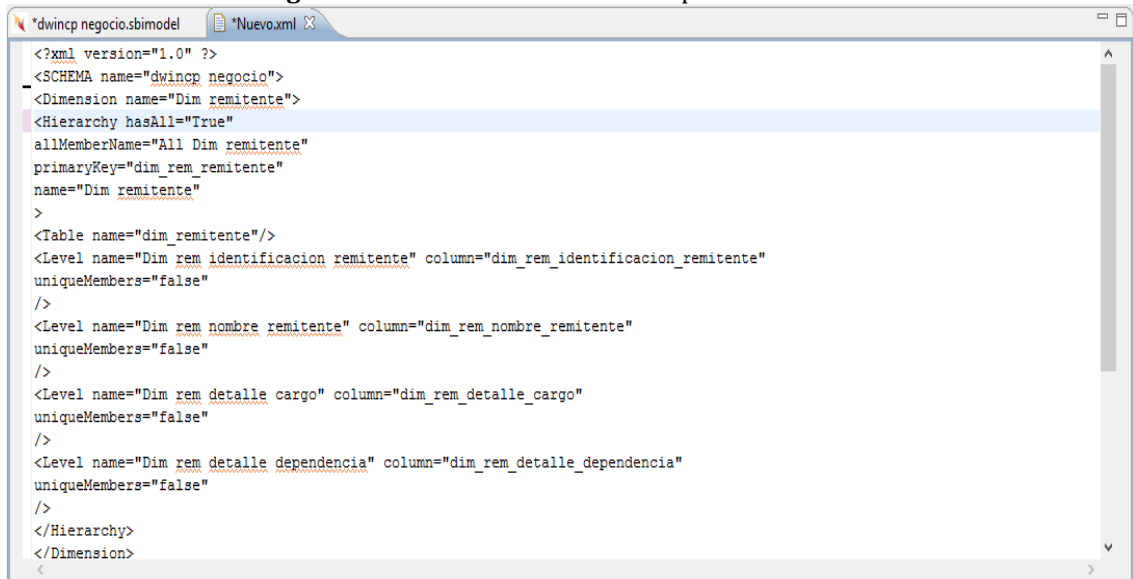
Figura 39. Ventana de selección de la ubicación de la plantilla



Fuente: elaboración propia

13. Editar el *template*, retiramos todas las dimensiones que no se van a usar en el reporte y completamos la palabra `<Hierarchy hasAll="True">` y guardamos

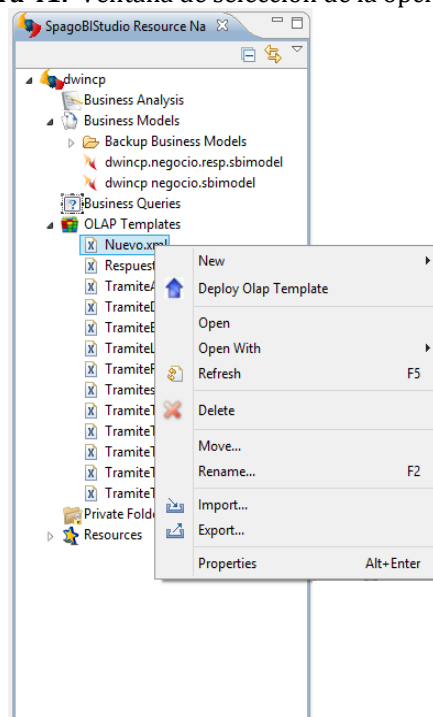
Figura 40. Ventana de edición de la plantilla creada



Fuente: elaboración propia

14. Para publicar, dar *click* derecho y seleccionar *Deploy Olap Template*

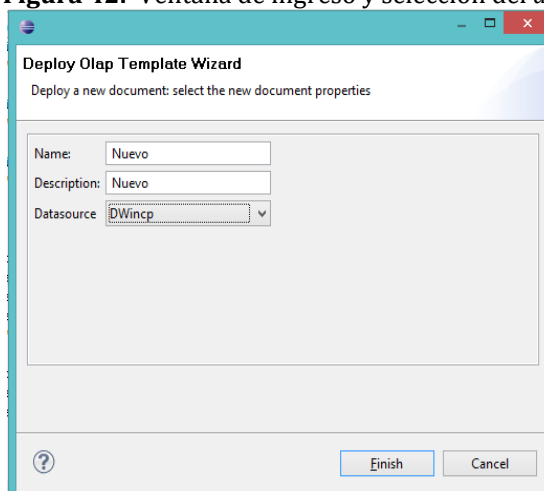
Figura 41. Ventana de selección de la opción publicar



Fuente: elaboración propia

15. Seleccionar el nombre la descripción y el *datasource*

Figura 42. Ventana de ingreso y selección del *datasource*

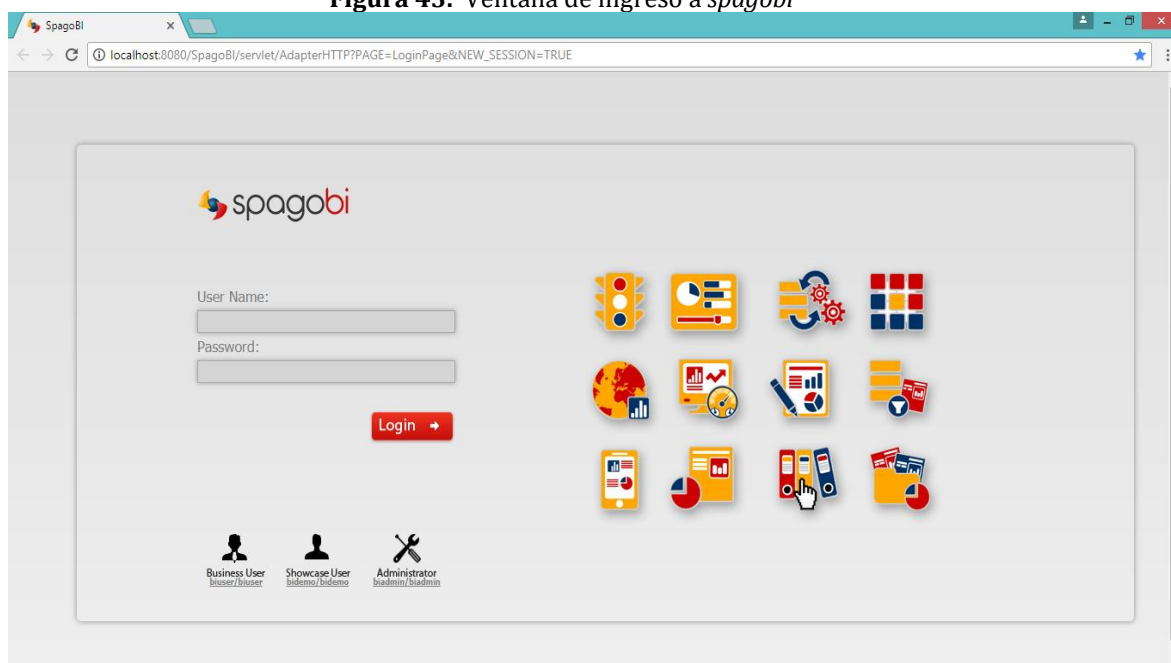


Fuente: elaboración propia

Una vez publicado el *datasource*, se mostrará un mensaje de éxito de la acción realizada.

16. La pantalla de la publicación de *spagobi* es la que se muestra a continuación, ingresamos las credenciales, en caso de no tenerla solicítela a su administrador

Figura 43. Ventana de ingreso a *spagobi*



Fuente: elaboración propia

17. Se despliega la pantalla principal de bienvenida a *spagobi*

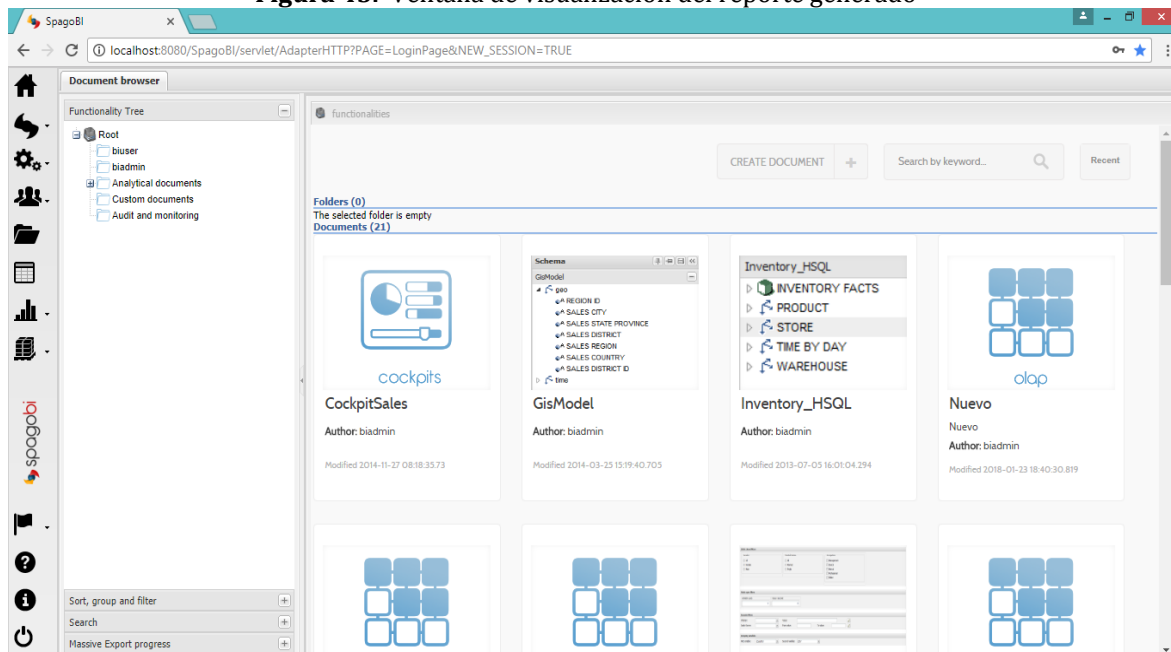
Figura 44. Ventana de bienvenida a *spagobi*



Fuente: elaboración propia

18. En el listado se ha añadido el reporte que publicamos desde el *spagobi Studio*

Figura 45. Ventana de visualización del reporte generado



Fuente: elaboración propia

19. Al abrir el reporte se muestra la siguiente pantalla

Figura 46. Ventana de visualización de un reporte



Dim remitente	Measures
-All Dim remitente	Fact tra numero 3,391
-0000000000	860
+AB. ALFREDO CAJAMARCA	1
+AB. CATHERINE GALLEGOS	1
+AB. MANUEL VILLA	2
+ABELARDO CHUQUITARCO CARATE	1
+ABG. ALEX AGUALONGO	1
+ABG. ANDRES M	1
+ABG. CARLOS ENRIQUE VARGAS PAZMIÑO	120
+ABG. CRISTIAN PAUL RUIZ COBA	22
+ABG. CRISTINA MARIBEL POVEDA ESPIN	65
+ABG. FRANKLIN BERRONES SANTANDER	1
+ABG. HERNÁN RAMIRO IDROVO SANCHEZ	48
+ABG. HÉCTOR SILVA G	1
+ABG. JOSÉ GUAMINGA	1
+ABG. MANUEL VARGAS	27
+ABG. MARCIA ARMIJO	
+ABG. MARÍA	
+ABG. MARÍA CECILIA MAYACELA SALAZAR	19
+ABG. MARÍA CRISTINA POVEDA ESPÍN	17
+ABG. MARÍA FERNANDA IBARRA VASCO	
+ABG. RITHA PAOLA CASTAÑEDA	1
+ABG. SEGUNDO AQUIETA	1
+AGR. DAVID GUITIERREZ	1
+ALEJANDRO PAZMIÑO ARIAS	
+ALEXA CALVOPÍÑA	1
+ALFONSO CUJILEMA	1
+ANA LUCÍA GUERRA	1

Fuente: elaboración propia

20. El diseño del reporte puede ser modificado, mediante el siguiente menú

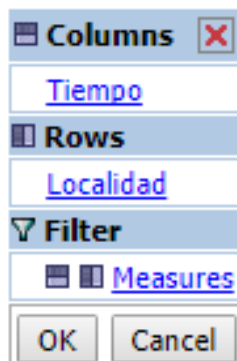
Figura 47. Menú de reporte



Fuente: elaboración propia

- El icono  permite la visualización de los filtros del reporte

Figura 48. Ventana de filtros de un reporte



Columns [X]

Tiempo

Rows

Localidad

Filter

Measures

OK Cancel

Fuente: elaboración propia

- Permite la organización del *grid* del reporte en filas y columnas

Figura 49. Visualización de la organización por filas y columnas de un *grid*

Tiempo	Localidad
☒ - Tiempo	☒ - Localidad
☒ - 2013	☒ - ZONA 3
☒ - #null ...	☒ - CHIMBORAZO
☒ + #null	☒ + ALAUSI
☒ + ABRIL	☒ + CHAMBO
☒ + AGOSTO	☒ + CHUNCHI
☒ + DICIEMBRE	☒ + COLTA
☒ + ENERO	☒ + CUMANDA
☒ + FEBRERO	☒ + GUAMOTE
☒ + JULIO	☒ + GUANO
☒ + JUNIO	☒ + PALLATANGA
☒ + MARZO	☒ + PENIPE
☒ + MAYO	☒ + RIOBAMBA
☒ + NOVIEMBRE	☒ - COTOPAXI
☒ + OCTUBRE	☒ + LA MANA
☒ - SEPTIEMBRE ...	☒ + LATACUNGA
☒ + SEPTIEMBRE	☒ + PANGUA
	☒ + PUJILI
	☒ + SALCEDO

Fuente: elaboración propia

- El icono  permite ordenar los datos del *grid* del reporte

Figura 50. Ventana para ordenar un *grid*

Sort Options
✕

Sort Mode

Keep Hierarchy Ascending ▼

Number of rows for Ranking

10

Show Memberproperties

☐

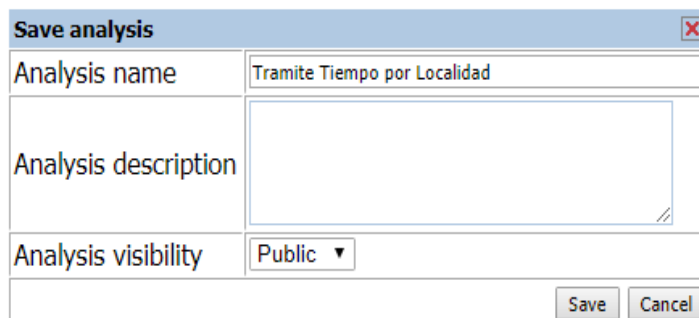
OK

Cancel

Fuente: elaboración propia

- El icono  permite guardar los cambios en una vista del reporte

Figura 51. Despliegue de la ventana guardar



Fuente: elaboración propia

Es necesario el ingreso de un nombre para la vista del reporte, si el nombre ya existe, el aplicativo reemplazará automáticamente el reporte existente. Para visualizar los reportes guardados, lo puede hacer mediante la pestaña **Shortcuts** , *show views*, selecciona la vista y genera, el reporte.

- El icono  permite visualizar el esquema de columnas del reporte

Figura 52. Visualización del esquema de las columnas del reporte

				Tiempo																	
				-Tiempo	Tiempo																+2014
					-2013	2013															
(All)	Zona	Provincia	Canton		#null	ABRIL	AGOSTO	DICIEMBRE	ENERO	FEBRERO	JULIO	JUNIO	MARZO	MAYO	NOVIEMBRE	OCTUBRE	SEPTIEMBRE				
-Localidad				3,391	892	2	66	58	55	101	51	78	82	97	87	67	64	84	1,057		
Localidad	-ZONA 3			3,391	892	2	66	58	55	101	51	78	82	97	87	67	64	84	1,057		
	ZONA 3 -CHIMBORAZO			2,833	712	1	49	49	51	87	40	65	60	78	70	50	48	64	916		
	CHIMBORAZO +ALAUSI			163	37		1	3		1	4		2	5	3	1	3	5	3	6	48
	+CHAMBO			35	12		1	1				1									
	+CHUNCHI			19	5		1							1	1		1	3	1	7	
	+COLTA																				6
	+CUMANDA			5	1										1						3
	+GUAMOTE			41	7				1	3								2		1	17
	+GUANO			67	14			1	1			1	3		3	2	2		1	25	
	+PALLATANGA			12																	5
	+PENIPE			14	8		3			3				2							2
	+RIOBAMBA			2,477	628	1	43	44	48	77	37	59	49	75	61	42	37	55	803		
	-COTOPAXI			303	111		12	7	1	8	6	10	10	13	14	7	7	16	75		
	COTOPAXI +LA MANA			9	1					1											3
	+LATACUNGA			180	89		12	5		7	6	7	9	11	13	4	2	13	48		
	+PANGUA																				
	+PUJILI			41	17			2	1			2	1	2	1	2	3	3	10		
	+SALCEDO																				
	+SAQUISILI																				
	+SIGCHOS			73	4							1				1	2			14	

Fuente: elaboración propia

- El icono  permite invertir la posición de filas y columnas

Figura 53. Visualización del esquema invertido filas y columnas

Tiempo	Localidad													
	Localidad	ZONA 3	CHIMBORAZO	ALAUSI	CHAMBO	CHUNCHI	COLTA	CUMANDA	GUAMOTE	GUANO	PALLATANGA	PENIPE	RIOBAMBA	COTOPAXI
-Tiempo	3,391	3,391	2,833	163	35	19		5	41	67	12	14	2,477	303
-2013	892	892	712	37	12	5		1	7	14		8	628	111
+ #null	2	2	1										1	
+ ABRIL	66	66	49	1	1	1						3	43	12
+ AGOSTO	58	58	49	3	1					1			44	7
+ DICIEMBRE	55	55	51	1					1	1			48	1
+ ENERO	101	101	87	4					3			3	77	8
+ FEBRERO	51	51	40	2	1								37	6
+ JULIO	78	78	65	5						1			59	10
+ JUNIO	82	82	60	3	2	1				3		2	49	10
+ MARZO	97	97	78	1	1			1					75	13
+ MAYO	87	87	70	3	1	2				3			61	14
+ NOVIEMBRE	67	67	50	5	1					2			42	7
+ OCTUBRE	64	64	48	3	3	1			2	2			37	7
+ SEPTIEMBRE	84	84	64	6	1				1	1			55	16
+2014	1,057	1,057	916	48	7	6		3	17	25	5	2	803	75
+2015	865	865	763	40	4	6		1	7	15	4	3	683	52
+2016	577	577	442	38	12	2			10	13	3	1	363	65
+2017														

Fuente: elaboración propia

- El ícono  permite visualizar el gráfico relacionado al reporte, mientras el ícono , permite la configuración del mismo

Figura 54. Visualización de las propiedades de un determinado gráfico

Chart Properties

Chart Type: Vertical Bar 3D

Enable Drill Through: ☐

Chart Title: Trámites por Localidad

Chart Title Font: SansSerif Bold 18

Horizontal axis label: Localidad

Vertical axis label: Trámites

Axes Label Font: SansSerif Plain 10

Axes Tick Label font: SansSerif Plain 10 90°

Show Legend: ☒ Bottom

Legend Font: SansSerif Plain 10

Show Slicer: ☒ Bottom Left

Slicer Font: SansSerif Plain 10

Chart Height: 400 Chart Width: 800

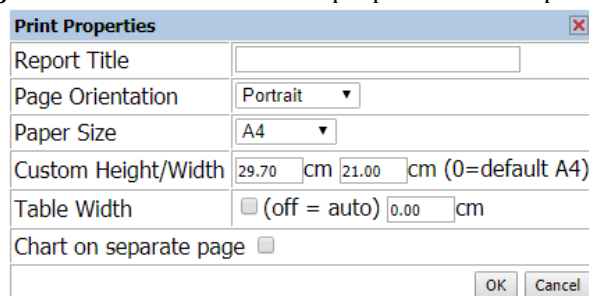
Background (R, G, B): 255 255 255

OK Cancel

Fuente: elaboración propia

- El icono  permite configurar el estilo de impresión para el reporte

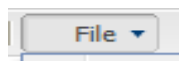
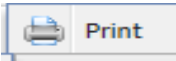
Figura 55. Visualización de las propiedades de impresión

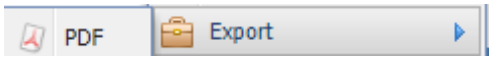


The 'Print Properties' dialog box contains the following fields and controls:

- Report Title:** A text input field.
- Page Orientation:** A dropdown menu set to 'Portrait'.
- Paper Size:** A dropdown menu set to 'A4'.
- Custom Height/Width:** Two input fields showing '29.70' cm and '21.00' cm, with a note '(0=default A4)'.
- Table Width:** A checkbox labeled '(off = auto)' and an input field showing '0.00' cm.
- Chart on separate page:** A checkbox.
- Buttons:** 'OK' and 'Cancel' buttons at the bottom right.

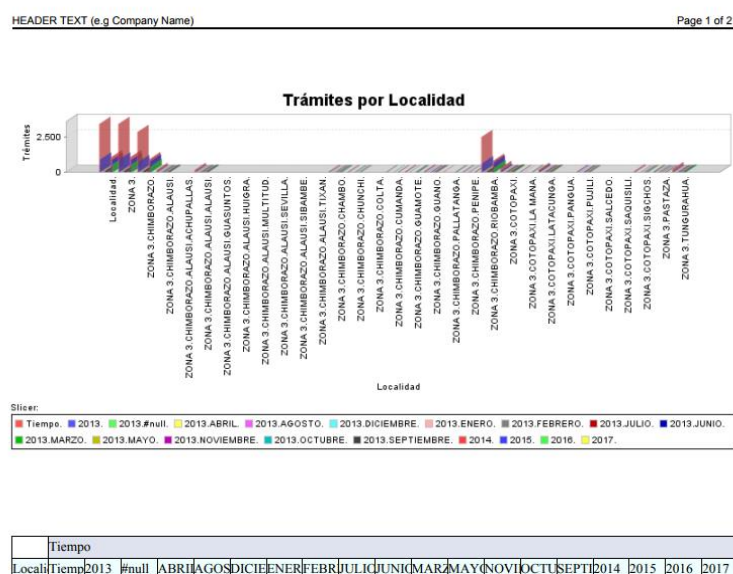
Fuente: elaboración propia

- Para imprimir un reporte se debe seleccionar la pestaña *File*  y escoger la opción *Print* 

- Para la exportación del PDF se lo realiza desde la pestaña *file*, donde seleccionamos *Export* y la opción deseada 

21. El reporte se visualiza de la siguiente manera

Figura 56. Visualización de un reporte seleccionado

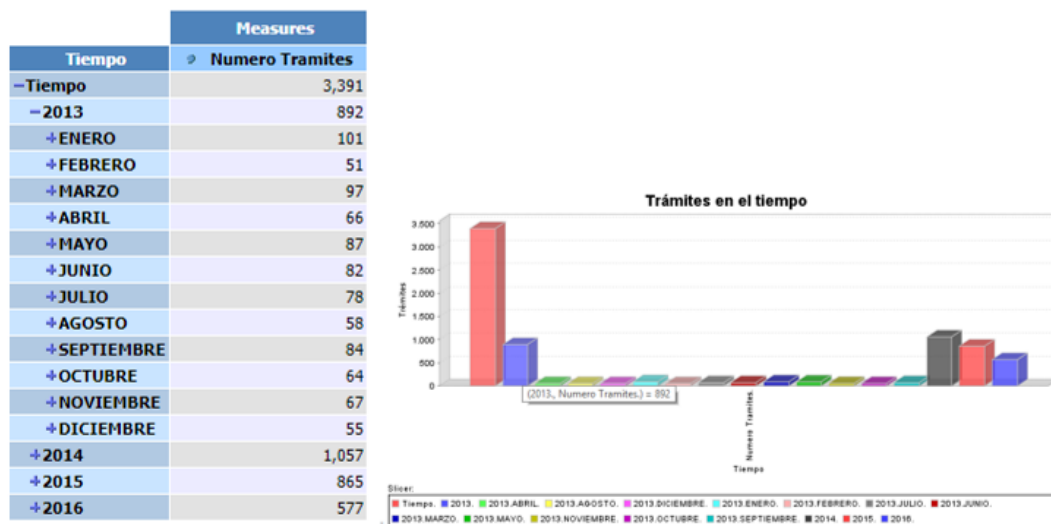


Fuente: elaboración propia

5.4. Reportes generados por el sistema

1. **Requerimiento:** Se requiere conocer el número de trámites receptados en un periodo de tiempo.

Gráfico 39. Título. Trámites en el tiempo



Fuente: elaboración propia

2. **Requerimiento:** Se requiere conocer el número de trámites por remitente y dependencia en un periodo de tiempo.

Gráfico 40. Título. Trámites por remitentes en el tiempo

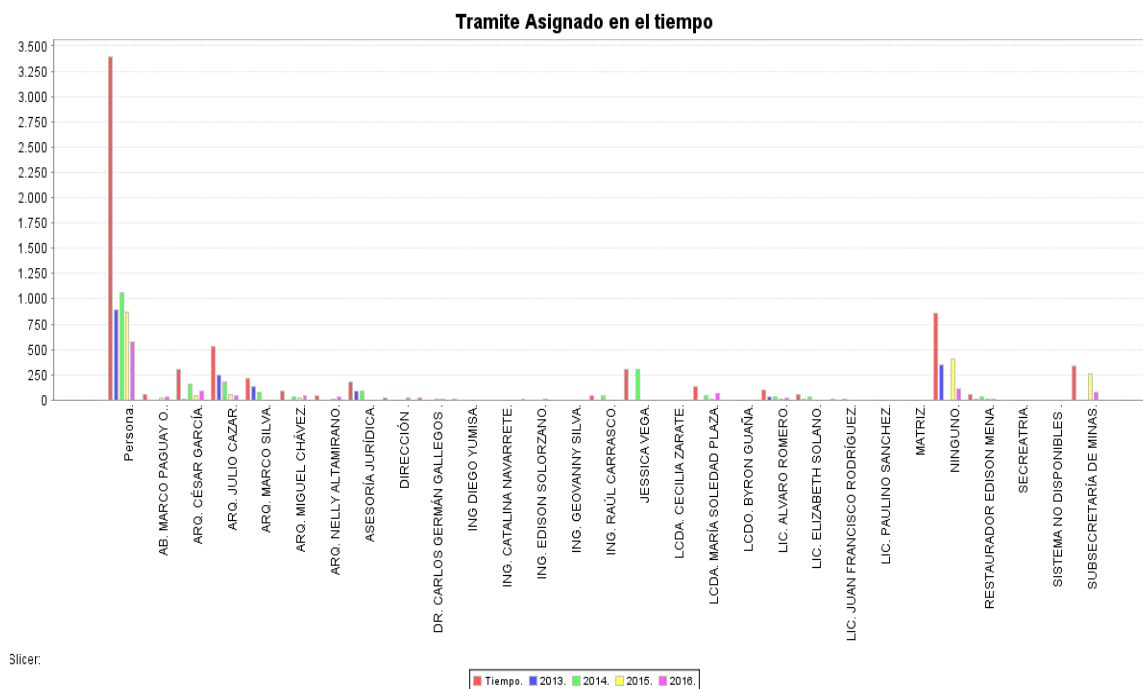
Remitente	Tiempo															
	Tiempo	2013	#null	ABRIL	AGOSTO	DICIEMBRE	ENERO	FEBRERO	JULIO	JUNIO	MARZO	MAYO	NOVIEMBRE	OCTUBRE	SEPTIEMBRE	84
-Remitente	3,391	892	2	66	58	55	101	51	78	82	97	87	67	64	84	
+ INGRID SAMANIEGO.																
+ SR. NELSON BENALCÁZAR																
+ WALTER MOYOTA																
+ AB. ALFREDO CAJAMARCA	1															
+ AB. CATHERINE GALLEGOS	1															
+ AB. MANUEL VILLA	2															
+ ABELARDO CHUQUITARCO CARATE	1															
+ ABG. ALEX AGUALONGO	1															
+ ABG. ANDRES H	1															
+ ABG. BOLÍVAR RAMOS ALTAMIRANO																
+ ABG. CARLOS ENRIQUE VARGAS PAZMIÑO	120															
+ ABG. CARLOS ENRIQUE VARGAS PAZMIÑO																
+ ABG. CECILIA MAYACELA SALAZAR																
+ ABG. CRISTIAN PAUL RUIZ COBA	22															
+ ABG. CRISTINA MARIBEL POVEDA ESPIN	65															
+ ABG. ELISA CESÉN ARTEAGA																
+ ABG. FRANKLIN BERRONES SANTANDER	1															
+ ABG. GRACE ARTAS NÚÑEZ																
+ ABG. HERNÁN RAMIRO IDROVO SANCHEZ	48															
+ ABG. HÉCTOR SILVA G	1															
+ ABG. JOSÉ GUAMINGA	1															
+ ABG. LEONOR LEMACHE																
+ ABG. MANUEL VARGAS	27															
+ ABG. MARCIA ARMIJO																
+ ABG. MARÍA																
+ ABG. MARÍA CECILIA MAYACELA SALAZAR	19															
+ ABG. MARÍA CRISTINA POVEDA ESPÍN	17															
+ ABG. MARÍA FERNANDA IBARRA VASCO																
+ ABG. MAURICIO MIRANDA BADILLO																

Fuente: elaboración propia

3. **Requerimiento:** Se requiere conocer el personal asignado durante el proceso de duración del trámite.

Gráfico 41. Título. Trámites asignados en el tiempo

Persona	Tiempo				
	-Tiempo	+2013	+2014	+2015	+2016
-Persona	3,391	892	1,057	865	577
AB. MARCO PAGUAY O.	57			21	36
ARQ. CÉSAR GARCÍA	302	15	160	39	88
ARQ. JULIO CAZAR	532	246	180	58	48
ARQ. MARCO SILVA	215	135	80		
ARQ. MIGUEL CHÁVEZ	91		32	20	39
ARQ. NELLY ALTAMIRANO	44			15	29
ASESORÍA JURÍDICA	178	86	92		
DIRECCIÓN	26				26
DR. CARLOS GERMÁN GALLEGOS	16			8	8
ING DIEGO YUMISA	5	2	2		1
ING. CATALINA NAVARRETE	2		2		
ING. EDISON SOLORZANO	12		3	3	6
ING. GEOVANNY SILVA	1		1		
ING. RAÚL CARRASCO	45		45		
JESSICA VEGA	305		305		
LCDA. CECILIA ZARATE	2		2		
LCDA. MARÍA SOLEDAD PLAZA	131		46	14	71
LCDO. BYRON GUAÑA	1				1
LIC. ALVARO ROMERO	101	36	34	10	21
LIC. ELIZABETH SOLANO	52	15	37		
LIC. JUAN FRANCISCO RODRÍGUEZ	11		7	3	1
LIC. PAULINO SANCHEZ	2				2
MATRIZ	1				1
NINGUNO	862	349	1	403	109
RESTAURADOR EDISON MENA	53	8	28	7	10
SECRETRIA	2				2
SISTEMA NO DISPONIBLES	1			1	
SUBSECRETARÍA DE MINAS	341			263	78



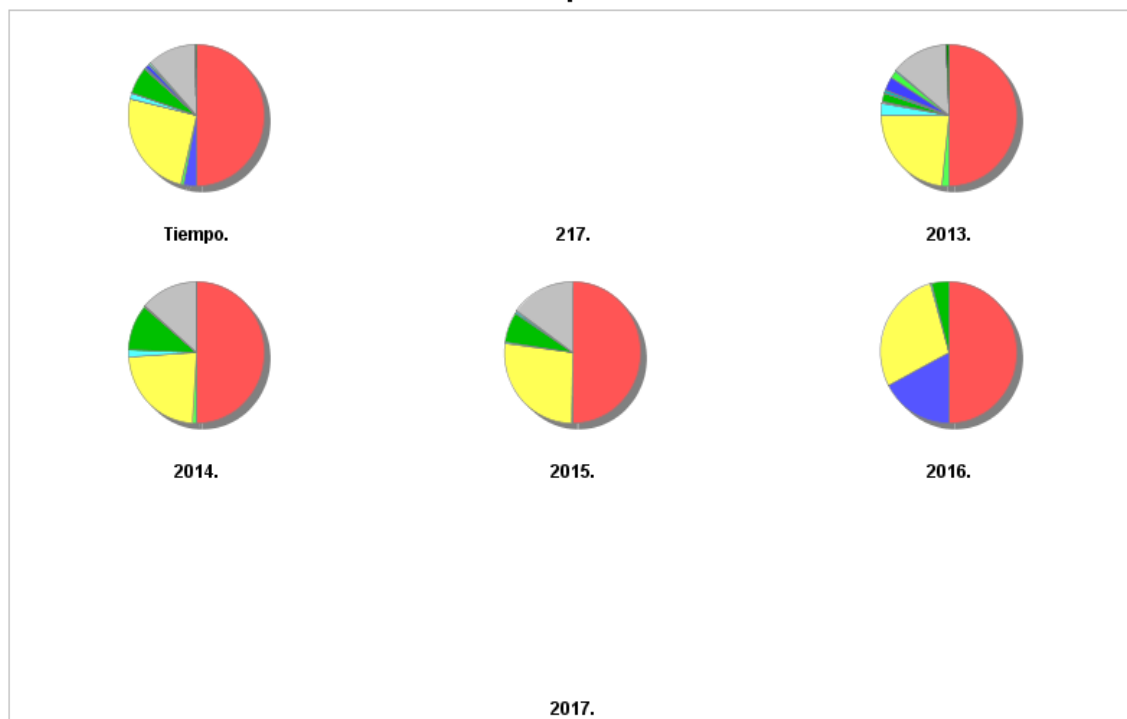
Fuente: elaboración propia

4. **Requerimiento:** Se requiere **conocer** el número de trámites por estado en un periodo de tiempo.

Gráfico 42. Título. Trámites por estados

Estado	Tiempo						
	-Tiempo	-217	-2013	-2014	-2015	-2016	-2017
-Estado	3,391		892	1,057	865	577	
ARCHIVADO	203			1	2	200	
ASISTIÓ	50		31	17	2		
ATENDIDO	1,692		413	488	465	326	
DESISTIÓ DE TRÁMITE	2		2				
EN CONOCIMIENTO	90		49	33	4	4	
EN REVISIÓN	5		5				
EN SEGUIMIENTO							
FALTA INSPECCIÓN E INFORME	2		2				
FALTA OFICIO	1		1				
NINGUNO	423		32	224	120	47	
NO SE ASISTIÓ	2		1	1			
NO SE REALIZÓ	2		2				
PENDIENTE	16		11		5		
PENDIENTE ACTA	1		1				
PENDIENTE CERTIFICACIÓN	1		1				
PENDIENTE INFORME	56		56				
PENDIENTE INSPECCIÓN - INFORME	28		27	1			
PROBLEMAS SISTEMA QUIPUS	3			3			
REALIZADO	2			2			
REASIGNADO A QUITO	12		7		5		
REASIGNADO A SUBSECRETARIA MINERÍA	785		237	286	262		
RECIBIDA LA DOCUMENTACIÓN	1			1			
RESPUESTA	2		2				
SIN RESPUESTA	12		12				

Trámites por estado



Slicer:

● Estado. ● ARCHIVADO. ● ASISTIÓ. ● ATENDIDO. ● DESISTIÓ DE TRÁMITE. ● EN CONOCIMIENTO. ● EN REVISIÓN. ● EN SEGUIMIENTO. ● FALTA INSPECCIÓN E INFORME. ● FALTA OFICIO. ● NINGUNO. ● NO SE ASISTIÓ. ● NO SE REALIZÓ. ● PENDIENTE. ● PENDIENTE ACTA. ● PENDIENTE CERTIFICACIÓN. ● PENDIENTE INFORME. ● PENDIENTE INSPECCIÓN - INFORME. ● PROBLEMAS SISTEMA QUIPUS. ● REALIZADO. ● REASIGNADO A QUITO. ● REASIGNADO A SUBSECRETARIA MINERÍA. ● RECIBIDA LA DOCUMENTACIÓN. ● RESPUESTA. ● SIN RESPUESTA.

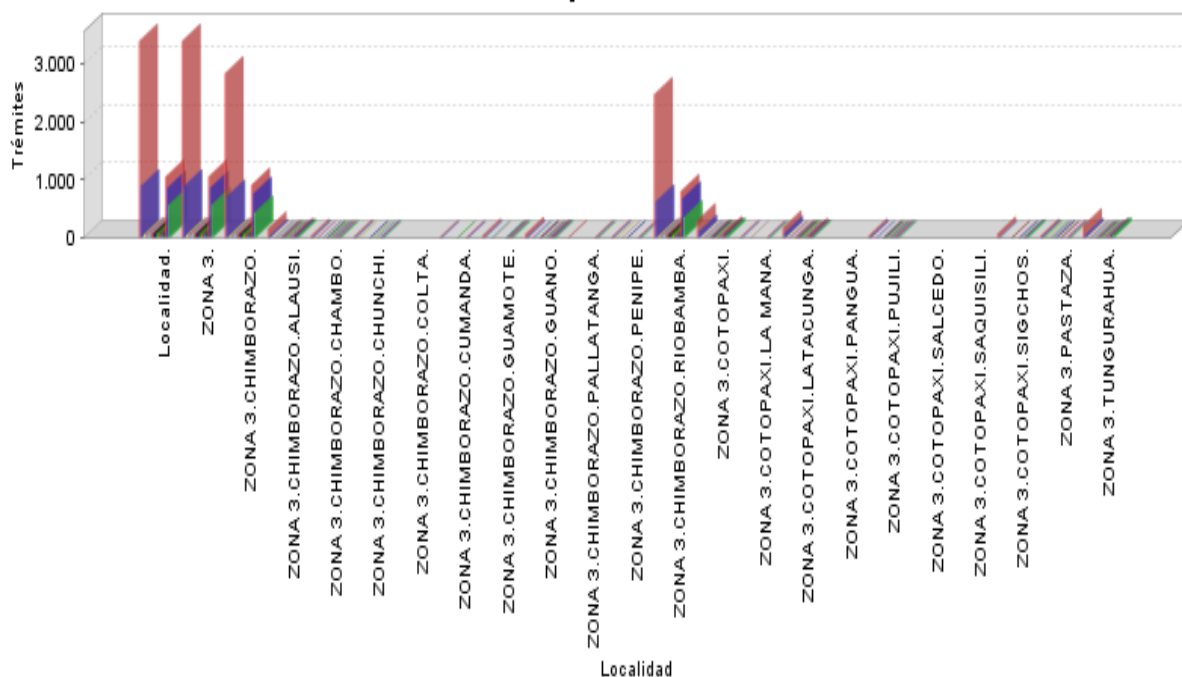
Fuente: elaboración propia

5. **Requerimiento:** Se requiere conocer el número de trámites por procedencia geográfica a ser procesados.

Gráfico 43. Título. Trámites por localidad

	Tiempo														
Localidad	→Tiempo	→2013	→#null	→ABRIL	→AGOSTO	→DICIEMBRE	→ENERO	→FEBRERO	→JULIO	→JUNIO	→MARZO	→MAYO	→NOVIEMBRE	→OCTUBRE	→SEPTIEMBRE
→Localidad	3,391	892	2	66	58	55	101	51	78	82	97	87	67	64	84
→ZONA 3	3,391	892	2	66	58	55	101	51	78	82	97	87	67	64	84
→CHIMBORAZO	2,833	712	1	49	49	51	87	40	65	60	78	70	50	48	64
→ALAUSI	163	37		1	3	1	4	2	5	3	1	3	5	3	6
→CHAMBO	35	12		1	1			1		2	1	1	1	3	1
→CHUNCHI	19	5		1						1		2		1	
→COLTA															
→CUMANDA	5	1									1				
→GUAMOTE	41	7				1	3							2	1
→GUANO	67	14			1	1			1	3		3	2	2	1
→PALLATANGA	12														
→PENIPE	14	8		3			3			2					
→RIOBAMBA	2,477	628	1	43	44	48	77	37	59	49	75	61	42	37	55
→COTOPAXI	303	111		12	7	1	8	6	10	10	13	14	7	7	16
→LA MANA	9	1					1								
→LATACUNGA	180	89		12	5		7	6	7	9	11	13	4	2	13
→PANGUA															
→PUJILI	41	17			2	1			2	1	2	1	2	3	3
→SALCEDO															
→SAQUISILI															
→SIGCHOS	73	4							1				1	2	
→PASTAZA	37	21	1	1		1	2	1		11		1	3		
→TUNGURAHUA	218	48		4	2	2	4	4	3	1	6	2	7	9	

Trámites por Localidad



Slicer:

■ Tiempo. ■ 2013. ■ 2013.#null. ■ 2013.ABRIL. ■ 2013.AGOSTO. ■ 2013.DICIEMBRE. ■ 2013.ENERO. ■ 2013.FEBRERO. ■ 2013.JULIO. ■ 2013.JUNIO. ■ 2013.MARZO. ■ 2013.MAYO. ■ 2013.NOVIEMBRE. ■ 2013.OCTUBRE. ■ 2013.SEPTIEMBRE. ■ 2014. ■ 2015. ■ 2016. ■ 2017.

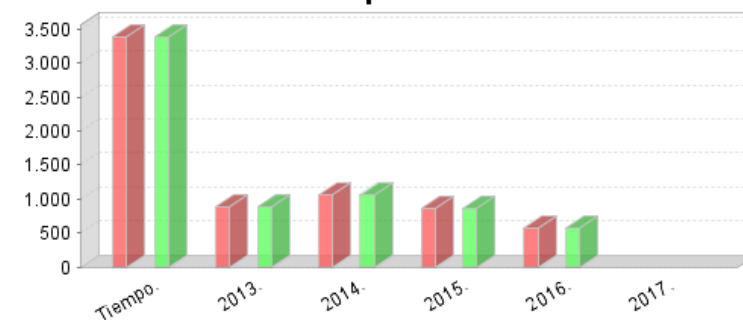
Fuente: elaboración propia

6. **Requerimiento:** Se requiere conocer el número de trámites por prioridad.

Gráfico 44. Título. Trámites por prioridad

	Tipo			
Tiempo	Tipo	BAJO	MEDIO	PRIORITARIO
-Tiempo	3,391		3,391	
+2013	892		892	
+2014	1,057		1,057	
+2015	865		865	
+2016	577		577	
+2017				

Trámites por Prioridad



Slicer:

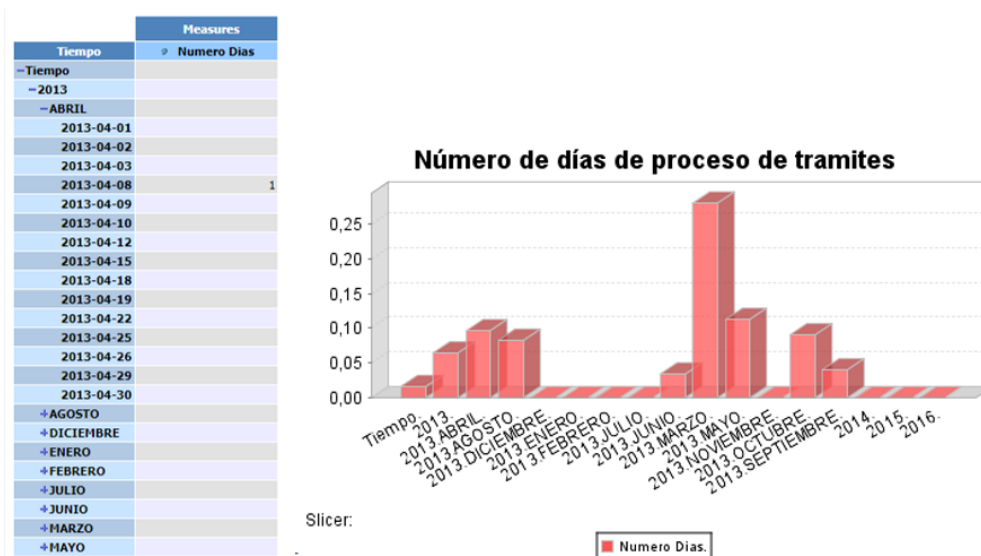
Numero Tramites.Tipo. Numero Tramites.BAJO. Numero Tramites.MEDIO.
Numero Tramites.PRIORITARIO.

Slicer:

Fuente: elaboración propia

7. **Requerimiento:** Se requiere conocer el número de días transcurridos hasta la finalización del trámite.

Gráfico 45. Título. Número de días de proceso de trámite



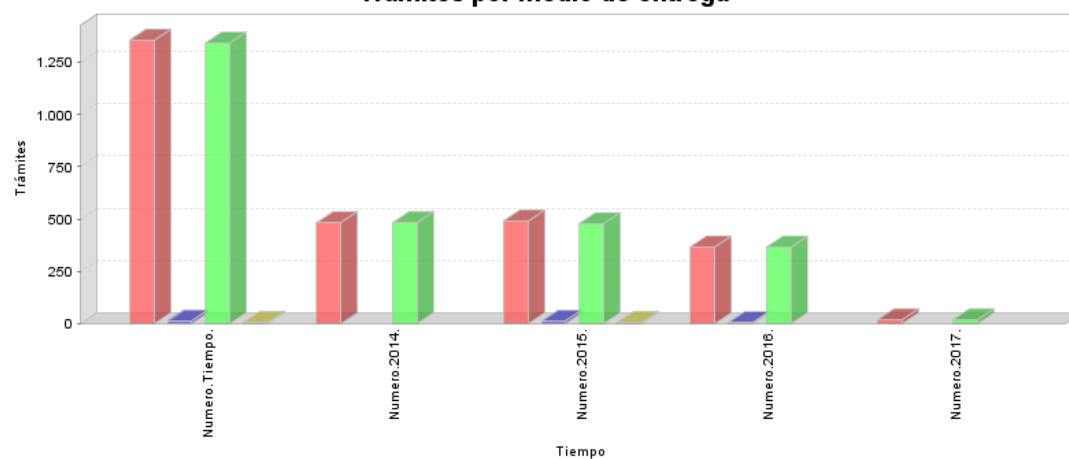
Fuente: elaboración propia

8. **Requerimiento:** Se requiere conocer el número de respuestas por medio de entrega en un periodo de tiempo

Gráfico 46. Título. Trámites por medio de entrega

Measures	Tiempo	Medio entrega			
		–Medio entrega	CORREOS DEL ECUADOR	NINGUNO	VENTANILLA
Numero	–Tiempo	1,359	14	1,342	3
	+ 2014	481		481	
	+ 2015	493	11	479	3
	+ 2016	368	3	365	
	+ 2017	17		17	

Trámites por medio de entrega



Slicer:

Slicer:

Medio entrega. CORREOS DEL ECUADOR. NINGUNO. VENTANILLA.

Fuente: elaboración propia

Capítulo 6

Conclusiones y Recomendaciones

6.1. Conclusiones

La investigación de los principios teóricos de las metodologías Kimball, Hefestos y SAS, se han convertido en el sustento para el desarrollo de soluciones *Data Warehouse*, gracias a la rapidez y sencillez en la obtención de reportes.

La selección de los parámetros de comparación de las metodologías a emplearse en un *Data Warehouse*, se desarrollaron en base a la investigación del número de fases requeridas para la elaboración de un *Data Warehouse*, así como en las características que brindan las metodologías Kimball, Hefestos y SAS a un proyecto pequeño, permitió conocer que la metodología Hefestos, proporciona una identificación ágil de los objetivos y resultados que persigue el INPC R3 para garantizar la toma de decisiones.

Los datos históricos de las hojas electrónicas del INPC R3 ayudaron al diseño e implementación de un *Data Warehouse* y el apoyo de una metodología madura, completa y orientada a la calidad fue la combinación más asertiva para la construcción de la solución con información integra.

La implementación del *Data Warehouse* para el INPC R3, permitió obtener reportes detallados en tiempo real; lo que ayuda a los funcionarios a proporcionar un control continuo, sobre los trámites y los tiempos de respuesta de las consultas en cada ejecución, garantizando acciones de mejora para la toma de decisiones en esta entidad.

6.2. Recomendaciones

Se recomienda el análisis de las necesidades de la solución a construirse para aplicar la metodología que mejor se adapte al desarrollo del proyecto de forma ágil y eficaz, permitiendo a las empresas la toma de decisiones en base a información íntegra.

Se recomienda el uso de la metodología Hefestos para la construcción de *Data Warehouse* en instituciones cuya información es relativamente pequeña y requiera la identificación ágil de objetivos y resultados; así como la interacción entre altos mandos y operativos.

Se recomienda la integración previa de la información de los archivos contenidos en hojas electrónicas de datos históricos, mediante los cuales se podrá obtener datos comparativos de años anteriores, pudiendo medir de manera eficaz el avance o retardo de las operaciones que el INPC R3 ha generado durante el ejercicio de sus actividades.

Es necesaria la evaluación de los requerimientos que la institución considera idóneos para un mejor desenvolvimiento, así como el procesamiento de los datos que permita obtener como resultado información específica que sirva de apoyo a los altos mandos en la toma de decisiones.

APÉNDICES

Apéndice A

METODOLOGÍA KIMBALL

Introducción.....	129
1. Planificación del proyecto.....	129
1.1. Definición del Proyecto	129
1.2. Justificación del proyecto en el negocio	129
1.3. Alcance del proyecto	129
1.4. Objetivos	130
1.4.1. Objetivo General	130
1.4.2. Objetivo General	130
1.5. Etapas del proyecto.....	130
1.6. Resumen del cronograma de actividades	130
2. Definición de requerimientos.....	131
2.1. Técnicas de recolección de información.....	131
2.2. Técnicas de recolección de información.....	131
2.3. Análisis de documentos y sistemas fuente.....	131
2.4. Preparación y publicación de requerimientos.....	132
2.4.1. Levantamiento de requerimientos	132
2.4.2. Documentación de requerimientos.....	132
2.4.2.1. Ver número de trámites en el tiempo.....	133
2.4.2.2. Ver trámites por asignación en el tiempo	133
2.4.2.3. Ver número de trámites por remitente y su dependencia en el tiempo.....	133
2.4.2.4. Ver número de trámites por prioridad (tipos) y su dependencia en el tiempo	133
2.4.2.5. Ver número de trámites por estado y su dependencia en el tiempo	133
2.4.2.6. Ver número de trámites por procedencia y su dependencia en el tiempo	133

2.4.2.7.	Ver número de trámites por estado y su dependencia en el tiempo	133
2.4.2.8.	Ver el número de días transcurridos por estado del trámite	134
3.	Diseño técnico de la arquitectura	134
3.1.	Arquitectura del Data Mart.....	134
3.1.1.	Back Room.....	134
3.1.1.1.	Servicios	134
3.1.1.2.	Repositorios	134
3.1.2.	Front Room	135
3.1.2.1.	Servicios	135
3.1.2.2.	Repositorios	135
3.2.	Metadatos.....	135
3.2.1.	Metadatos del sistema fuente	135
3.2.2.	Metadatos del área de ETL.....	136
3.2.3.	Metadatos del sistema fuente	136
3.3.	Selección de productos	137
3.3.1.	Productos para el Back Room	137
3.3.2.	Productos para el Front Room.....	137
4.	Modelado dimensional	138
4.1.	Modelo dimensional	138
4.1.1.	Conformación de dimensiones	138
4.1.2.	Conformación de las tablas de hechos.....	139
4.1.2.1.	Selección del Data Mart.....	139
4.1.2.2.	Selección de dimensiones	139
4.1.2.3.	Definición de la granularidad	140
4.1.2.4.	Selección de medidas	140
4.1.3.	Diagrama de la tabla de hechos	140
4.1.4.	Detalle de la tabla de hechos	141
4.1.5.	Detalle de las tablas de dimensiones.....	142
4.1.6.	Identificación de la fuente de datos para tablas de hechos y dimensiones	147
4.1.7.	Mapeo detallado de la fuente de datos	148
5.	Diseño físico, construcción y despliegue del Data Mart	151
5.1.	Diseño Físico.....	151

5.1.1.	Elaboración de estándares	151
5.1.1.1.	Estándares para objetos de base	151
5.1.1.2.	Estándares para objetos de ETL	152
5.1.2.	Diseño del modelo físico	153
5.1.2.1.	Modelo físico de la tabla de hechos de trámites	154
5.1.2.2.	Modelo físico de la tabla de respuestas de trámites	155
5.1.3.	Cubos OLAP	155
6.	Proceso de extracción, transformación y carga	160
6.1.	Elaboración de plan	161
6.1.1.	Carga inicial de datos	162
6.1.2.	Carga incremental de datos	163
6.1.3.	Actualización de datos	163
6.2.	Definición de transformaciones y cálculos	163
6.3.	ETL de las tablas de dimensión.....	166
6.3.1.	ETL de la dimensión tiempo	166
6.3.2.	ETL de la dimensión localidad	167
6.3.3.	ETL de la dimensión remitente.....	167
6.3.4.	ETL de la dimensión persona.....	167
6.3.5.	ETL de la dimensión área.....	167
6.3.6.	ETL de la dimensión receptor	167
6.3.7.	ETL de la dimensión medio entrega	168
6.3.8.	ETL de la dimensión tipo trámite	168
6.3.9.	ETL de la dimensión estado	168
6.3.10.	ETL de las tablas de hechos.....	168
7.	ANEXOS.....	169
7.1.	Entrevista al Director del INPC	169
7.2.	Entrevista Jefe de sistemas.....	169

Introducción

1. Planificación del proyecto

1.1. Definición del Proyecto

Para el proyecto desarrollado se ha identificado la necesidad de la institución de identificar de mejor manera los trámites, su proceso y respuesta emitida; lo cual permitirá una mejor toma de decisiones, en lo que respecta a la distribución de los trámites, para su rápida gestión.

Esta demanda puede ser solventada con la implementación de un *Data Warehouse*, el cual se encargará de transformar los datos del sistema relacional en información útil sobre los trámites que se realizan; esta solución cuenta con un soporte de datos históricos, lo cual permitirá un análisis a profundidad de las necesidades de la institución de forma eficiente y rápida, puesto que el comparativo no tendrá que esperar a datos futuros.

1.2. Justificación del proyecto en el negocio

El proyecto a realizarse es de índole académico, y busca incrementar la productividad en el proceso de trámites justificándose su implementación mediante los beneficios que se plantean a continuación:

- Disminuir el tiempo en el proceso de un trámite
- Al determinar los responsables de cada proceso en un trámite, es posible establecer el responsable con menor carga de trabajo, para la asignación de un nuevo trámite
- Determinar la procedencia de los trámites

1.3. Alcance del proyecto

El proyecto será diseñado para interactuar con fuentes de datos procedentes de libros de Excel; y la información proporcionada por una base de datos relacional implementada en MySQL. La realización del proceso de extracción, transformación y carga se realizará a un modelo dimensional en MySQL enlazada a la herramienta de consulta de datos Power BI.

Este *Data Warehouse* se encargará de transformar los datos proporcionados por las fuentes externas y transformarlos en información concerniente a los trámites que se procesan en la

institución. Se desarrollarán e implementarán los requerimientos determinados luego del levantamiento y análisis de requerimientos que se acuerden con la institución.

1.4. Objetivos

1.4.1. Objetivo General

Construir un *Data Warehouse* que apoye la toma de decisiones en el proceso de trámites del INPC R3, utilizando la metodología Kimball.

1.4.2. Objetivo General

- Integrar un *Data Warehouse* con el sistema transaccional del INPC R3.
- Integrar los datos de archivos en Excel de datos históricos del INPC R3.
- Ofrecer opciones de consultas directas e interactivas de la información de los trámites que procesa la institución.

1.5. Etapas del proyecto

La metodología plateada por Ralf Kimball es correspondiente a una metodología *Bottom-up*; este ciclo de vida consta de tareas secuenciales y en paralelo, proporcionando una alta flexibilidad.

1.6. Resumen del cronograma de actividades

Se plantea un cronograma de las actividades que se realizarán durante el desarrollo e implementación del *Data Warehouse*

Actividad	Fecha inicio	Fecha fin	Tiempo (semanas)
Planificación del proyecto	Febrero 02	Febrero 16	2
Definición de requerimientos	Febrero 17	Marzo 02	2
Diseño técnico de la arquitectura	Marzo 03	Marzo 23	3
Modelo dimensional	Marzo 24	Abril 13	3
Diseño físico, construcción y despliegue	Abril 14	Mayo 04	3
Proceso ETL	Mayo 05	Junio 25	3

2. Definición de requerimientos

En esta etapa es necesaria la definición de los requerimientos necesarios para el cumplimiento de las etapas subsiguientes, se determinan requisitos de arquitectura y selección de herramientas; además de las especificaciones que el usuario final requiere se implementen dentro del proyecto.

2.1. Técnicas de recolección de información

La recolección de la información se realiza mediante las técnicas:

- Entrevistas personales con los directores del INPC R3
- Encuestas a los usuarios finales
- Análisis de documentos proporcionados por el sistema transaccional y archivos Excel

2.2. Técnicas de recolección de información

Se efectuó la entrevista al director de la INPC R3, con el propósito de conocer las necesidades de la institución con respecto al proceso de gestión; lo cual nos permite entender el seguimiento que se da a los trámites antes mencionados y conocer cuáles son los reportes y peticiones de información solicitados constantemente.

Se realizó una reunión con el jefe de sistemas de la institución, para verificar las fuentes de datos, los reportes que necesitan de gestión directa con el motor de base de datos, las herramientas mediante las cuales los datos son transformados en información

2.3. Análisis de documentos y sistemas fuente

La verificación de varias consultas SQL que son usadas para la obtención de la información nos permite la identificación de un indicador: el número de trámites por cada uno de las tablas con las cuales se relaciona, y el intervalo de tiempo que un trámite está en proceso.

Se identificó una categoría de la información, la ubicación, la cual agrupa parroquias, ciudades, provincias y zonas.

Proceso: Registro de trámites	
Indicador	Número de trámites
	Intervalo de tiempo desde el inicio hasta el fin del proceso

Categorías	Remitente, tipo trámite, ubicación (Zona, Provincia, Cantón, Parroquia), persona asignada, estado del trámite, área
Proceso: Registro de respuesta de trámite	
Indicador	Número de trámites Intervalo de tiempo desde inicio del trámite hasta el momento que se genera una respuesta
	Estado trámite, receptor, medio de entrega

2.4. Preparación y publicación de requerimientos

La recopilación de información permite el levantamiento de los requerimientos para el *Data Warehouse*

2.4.1. Levantamiento de requerimientos

Los requerimientos fueron establecidos mediante el empleo de entrevistas a los usuarios de la solución. Los requerimientos del negocio se describen a continuación.

- Ver número de trámites en el tiempo.
- Ver número de trámites por remitente y su dependencia en el tiempo.
- Ver trámites por asignación en el tiempo.
- Ver número de trámites por prioridad (tipos) en el tiempo.
- Ver número de trámites por estado en el tiempo.
- Ver trámites por respuestas en el tiempo.
- Ver número de trámites por procedencia geográfica en el tiempo.
- Ver el número de días transcurridos por estado del trámite.

Estos requerimientos están acordes a las necesidades de la empresa, así como a los datos que se proveerán al sistema.

2.4.2. Documentación de requerimientos

La requerimentación que se presenta a continuación se enfoca en conocer las necesidades de la institución, las fuentes de datos y la información que será proporcionada por cada requerimiento.

2.4.2.1. Ver número de trámites en el tiempo.

Descripción: Esta consulta, permitirá explorar los trámites por fechas receptados por la institución; al hacer drill down se limita el criterio del reporte, permitiendo analizar los trámites por año, semestre, trimestre y día.

Fuente de datos: base de datos MySQL

2.4.2.2. Ver trámites por asignación en el tiempo

Descripción: Esta consulta, permitirá verificar la asignación de los trámites en un periodo de tiempo; el drill down limitará el criterio del reporte, permitiendo analizar los trámites por año, semestre, trimestre, mes y semana

Fuente de datos: base de datos MySQL

2.4.2.3. Ver número de trámites por remitente y su dependencia en el tiempo

Descripción: Esta consulta, permitirá identificar los trámites en un periodo de tiempo; el drill down facilitará la exploración de los trámites por remitente.

Fuente de datos: base de datos MySQL

2.4.2.4. Ver número de trámites por prioridad (tipos) y su dependencia en el tiempo

Descripción: Esta consulta, permitirá identificar los trámites en un periodo de tiempo; el drill down facilitará la exploración de los trámites por prioridad.

Fuente de datos: base de datos MySQL

2.4.2.5. Ver número de trámites por procedencia y su dependencia en el tiempo

Descripción: Esta consulta, permitirá identificar los trámites en un periodo de tiempo; el drill down facilitará la exploración de los trámites por procedencia geográfica.

Fuente de datos: base de datos MySQL

2.4.2.6. Ver número de trámites por estado y su dependencia en el tiempo

Descripción: Esta consulta, permitirá identificar las respuestas de un trámite en un periodo de tiempo.

Fuente de datos: base de datos MySQL

2.4.2.7. Ver el número de días transcurridos por estado del trámite

Descripción: Esta consulta, permitirá identificar el número de días transcurridos desde que se inicia el trámite hasta su finalización.

Fuente de datos: base de datos MySQL

2.4.2.8. Ver el número respuestas por medio de entrega y su dependencia en el tiempo

Descripción: Esta consulta, permitirá identificar el número respuestas y su dependencia en el tiempo.

Fuente de datos: base de datos MySQL

3. Diseño técnico de la arquitectura

3.1. Arquitectura del Data Mart

3.1.1. Back Room

3.1.1.1. Servicios

Recoge todos los procesos y recursos necesarios para el procesamiento de los datos para el ETL.

- **Servicio de extracción de datos:** Permite que los datos sean extraídos desde el origen
- **Servicio de transformación de datos:** Algunos datos deben ser transformados durante esta etapa; los procesos más comunes son: generación de las claves subrogadas, desnormalización de tablas, cambios o conversiones en tipos de datos, generación de campos calculados.
- **Servicio de carga de datos:** Se encarga del almacenamiento de los datos.
- **Servicio de control de tareas:** El control de errores y manejo de excepciones se realiza en esta etapa, al igual que el monitoreo y tareas programadas

3.1.1.2. Repositorios

- **Sistemas fuente:** origen de datos desde donde el *Data Mart* es alimentado, es sobre este repositorio donde se realiza la extracción de datos.

DBMS: MySQL

BDD: inpc

Número de tablas: 15 tablas

- **Area ETL:** Repositorio temporal donde se realizan los procesos de transformación y la calendarización de tareas, así como un monitoreo de las mismas
- **DMBS** Repositorio final de los datos que procesados a través del ETL.

3.1.2. Front Room

3.1.2.1. Servicios

Son los procesos que deben ser útiles para la presentación de la información

- **Servicio de navegación sobre el *Data Mart*:** Es el proceso mediante el cual se realiza la búsqueda de un dato.
- **Servicio de acceso y seguridad:** La autenticación y conexión se encuentra dentro de este servicio, este servicio puede ser controlado por otro servicio.
- **Servicio de administración de consultas:** Permite interacción del usuario con la base de datos a través de consultas directas.
- **Servicio de reportes estándar:** Servicio que permite la generación de reportes en base a plantillas predefinidas

3.1.2.2. Repositorios

- **Motor OLAP:** Generador de cubos y dimensiones con la información que es almacenada, procesador de consultas solicitadas por servicio de consultas
- **Herramienta de acceso a datos:** Herramienta que permite la interacción con los datos, mediante consultas MDX y SQL

3.2. Metadatos

La información proporcionada en esta etapa sirve como base para los siguientes procesos de la construcción del *Data Mart*, cada uno de estos datos debe ser desarrollado según las necesidades y el proceso.

Para los metadatos *Back Room*, se establece la información necesaria acerca de los sistemas fuente y el ETL.

3.2.1. Metadatos del sistema fuente

Especificaciones de la fuente de datos: Información de la base de datos relacional de la aplicación del INPC R3

Plataforma		MySQL
Nombre		Inpc
Owner	Usuario	Root
	Password	11032013
	Privilegios	Administrador

Diccionario de datos: Contiene la especificación de tablas y columnas de la base de datos.

3.2.2. Metadatos del área de ETL

Políticas de cambio de dimensiones: manera como se administran los cambios para cada una de las dimensiones

Tipo cambio	Descripción
Tipo 0	Sin Cambio
Tipo 1	Sobre escritura
Tipo 2	Creación de nuevo registro en la dimensión

Conexiones de base de datos: Se detalle el nombre de la conexión y cada uno de sus componentes (servidor, DBMS, BD, usuario, password).

Transformación de datos: Identificación de las columnas sobre las cuales se realizan los cambios y cálculos.

Flujo de datos de la fuente al destino. Diagramas de flujo de datos sobre el cada una de las etapas que el proceso debe recorrer (nombre del paso, objeto provisto por la herramienta ETL, descripción)

Para el Front Room,

3.2.3. Metadatos del sistema fuente

Especificación de la Base de Datos del *Data Mart*: Información de la base de datos del *Data Mart* construido para el INPC R3

Plataforma		MySQL
Nombre		DWHInpc
Número de tablas		10
Owner	Usuario	Root

	Password	11032013
	Privilegios	Administrador

Dimensiones y tablas de hechos: Información detallada de las tablas de hechos: claves dimensionales, medidas y agregaciones; al igual que tablas de dimensiones con sus: nombres, descripciones, políticas de cambio y ejemplos)

Índices: Índices por tabla con su nombre, tipo, y columnas que lo conforman

Diccionario de datos: Descripción de las tablas y columnas de la Base de datos para el *Data Mart*: nombre de columnas, claves subrogadas, claves operacionales, tipos de datos, descripciones.

Capacitación a usuarios: Manual de despliegue, manual de usuario

3.3. Selección de productos

3.3.1. Productos para el *Back Room*

Producto seleccionado para la DBMS:

Nombre	MySQL
Versión	5.7.18

Producto seleccionado para el área ETL

Nombre	SpagoBI studio
Versión	5.1

3.3.2. Productos para el *Front Room*

Producto seleccionado para el motor OLAP

Nombre	SpagoBI studio
Versión	5.1

Producto seleccionado como herramienta de acceso a datos

Nombre	SpagoBI studio
Versión	5.1

4. Modelado dimensional

4.1. Modelo dimensional

Se puede establecer qué modelo dimensional y un modelo entidad relación contienen la misma información, cuya información es empaquetada en un formato simétrico, y entre los objetivos que persigue tenemos: el usuario debe comprender la información, las consultas deben generarse con un alto rendimiento y debe ser flexible al cambio.

4.1.1. Conformación de dimensiones

- **Dimensión Tiempo**

En esta dimensión comprende las fechas: Fecha Documento (trámite), Fecha Recepción (trámite), Fecha Proceso (proceso trámite), Fecha Respuesta (respuesta trámite). Estos datos están agrupados por: años, semestres, trimestres, meses, semanas, días.

- **Dimensión Localidad**

Información descriptiva de la ubicación geográfica, posee datos como las Provincias, Ciudades y Parroquias.

- **Dimensión Remitente – Cargo – Dependencia (Remitente)**

Información descriptiva concerniente a los cargos que un remitente ocupa, datos como: Identificación, Nombres, Cargo, Detalle de la dependencia

- **Dimensión Persona Asignada**

Información descriptiva de la persona a quien se le asigna un trámite, datos como Identificación, Nombre

- **Dimensión Área**

Información descriptiva de la dependencia y el área hacia donde es emitida la respuesta de un trámite, datos como: Detalle del área

- **Dimensión Receptor**

Información descriptiva del receptor de la respuesta que genera un trámite, datos como: Identificación, Nombres.

- **Dimensión Medio Entrega**

Información descriptiva de los medios por los cuales se entrega la respuesta de un trámite, datos como: Detalle del medio de entrega

- **Dimensión Tipo Trámite (Tipo)**

Información del tipo de trámite, datos como: Detalle del tipo de trámite.

- **Dimensión Estado Proceso Trámite (Estado)**

Información descriptiva sobre el estado en el cual se encuentra el trámite, datos como: Detalle del estado del proceso de trámite

4.1.2. Conformación de las tablas de hechos

Según el análisis de los requerimientos realizados en la fase de análisis, se determina la Tabla de hechos de Trámite. Para realizar el diseño se ha dividido en 4 pasos:

4.1.2.1. Selección del Data Mart

Este proceso es usado para soluciones *Data Warehouse* con múltiples *Data Mart*, en este caso solo se establece un único *Data Mart* sobre el cual se conformarán las tablas de hechos

4.1.2.2. Selección de dimensiones

Las dimensiones fueron establecidas para la tabla de hechos "Trámites";

Tabla de hechos	Proceso	Respuesta
Dimensiones		
Dimensión Tiempo	X	
Dimensión Localidad	X	
Dimensión Remitente	X	
Dimensión Persona Asignada	X	
Dimensión Tipo Trámite	X	
Dimensión Estado	X	
Dimensión Receptor		X
Dimensión Medio Entrega		X
Dimensión Área	X	

4.1.2.3. Definición de la granularidad

- Se identificó los niveles mínimos de acceso a los cuales se puede apreciar la información
- Tabla de hechos de proceso de trámites (trámite, anexo, proceso)
- Cada registro del proceso en los que se encuentre un trámite es un registro en esta tabla
- Tabla de hechos de respuestas de trámites
- Cada registro de respuesta generado por un proceso de trámite se registra en esta tabla.

4.1.2.4. Selección de medidas

Esta selección se realiza mediante los indicadores y la granularidad determinada para cada tabla de hechos:

Tabla de hechos de proceso de trámites

Captura el evento de registro de trámites y su proceso; las medidas para esta tabla son: el número de trámites que se registre en la granularidad de las dimensiones con las que se encuentra relacionado (el número de trámites por remitente, el número de trámites por estado, el número de trámites por tipo, el número de trámites por dependencia geográfica, el número de trámites por persona asignada), los intervalos de tiempo en que se registran los trámites y su fin de proceso.

Tabla de hechos de respuestas de trámites

Captura el evento de registro de respuestas para cada uno de los procesos del trámite que se registre en la granularidad de las dimensiones con las que se encuentra relacionado (el número de respuestas por trámite, el número de respuestas por receptor, el número de respuestas por dependencia).

4.1.3. Diagrama de la tabla de hechos

Diagrama de la tabla de hechos de trámites

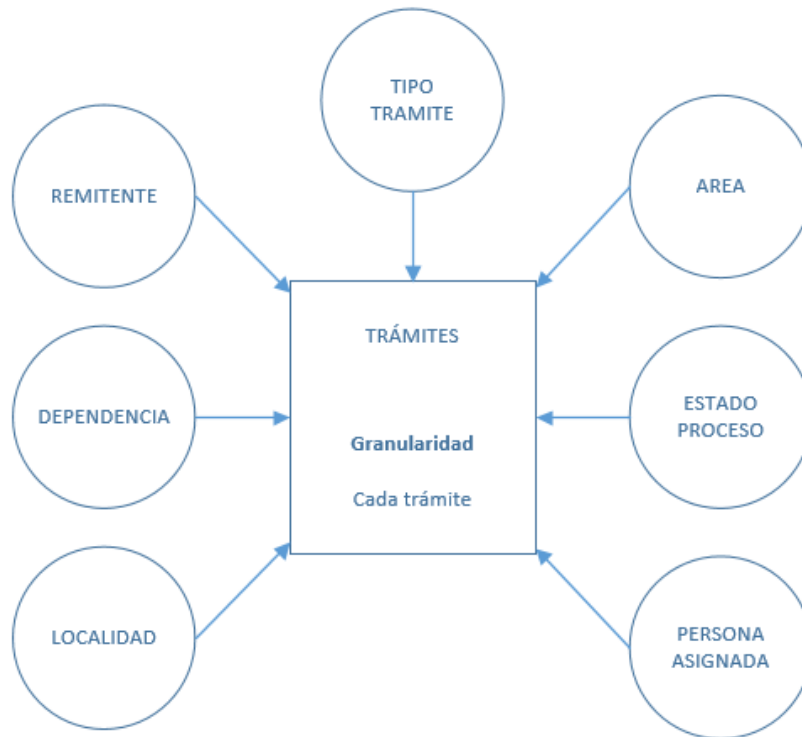
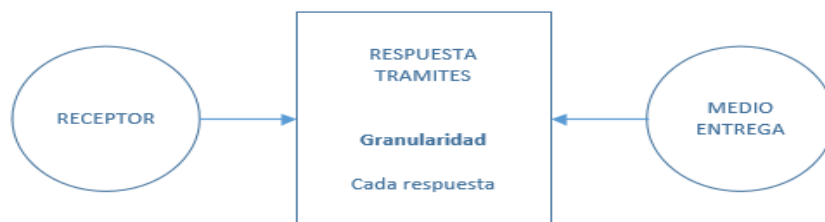


Diagrama de la tabla de hechos de respuesta de trámites



4.1.4. Detalle de la tabla de hechos

Las tablas de hechos presentan información sobre las claves dimensionales o subrogadas de las dimensiones relacionadas a esta tabla, información sobre las medidas, la regla de agregación y la fórmula mediante la cual se obtienen.

Tabla de hechos de proceso de trámites

Claves	Descripción
Key: remitente	Clave de la dimensión remitente
Key: estado	Clave de la dimensión estado
Key: tipo	Clave de la dimensión tipo

Key: localidad	Clave de la dimensión localidad
Key: tiempo- fecha de proceso	Clave de la dimensión tiempo con respecto a la fecha de registro de cada proceso
Key: tiempo- fecha de documento	Clave de la dimensión tiempo con respecto a la fecha de emisión del documento
Key: persona-asignada	Clave de la dimensión persona asignada
Key: area	Clave de la dimensión área
Medidas	Descripción
Fact: Trámites	Número de trámites registrados en base a la granularidad de las dimensiones
	Tipo Transformación
	Agregación Suma
Fact: Proceso (Días)	Número de días transcurridos entre la fecha de inicio de proceso (FIP) y la fecha de finalización del proceso (FFP)
	Tipo Transformación
	Formula $TI = \text{day}(\text{FFP}) - \text{day}(\text{FIP})$
	Agregación Promedio

Tabla de hechos de respuesta de trámites

Claves	Descripción
Key: receptor	Clave de la dimensión remitente
Key: medioEntrega	Clave de la dimensión persona-asignada
Pk: procesoTrámite	Clave de la tabla de hechos de proceso de trámite
Medidas	Descripción
Fact: Respuestas	Número de trámites registrados en base a la granularidad de las dimensiones
	Tipo Transformación
	Agregación Suma

4.1.5. Detalle de las tablas de dimensiones

Las tablas de dimensiones presentan información sobre sus atributos, y la descripción de cada uno de ellos; para ello se consideran: Identificadores de fuentes de datos de origen (Id); Claves

Dimensión tiempo

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: tiempo	Clave subrogada de la dimensión	Tipo 0	1
Fecha	Fecha completa: año, mes, día	Tipo 0	2015-01-10
Dia	Día en la fecha	Tipo 0	10
Mes	Mes en la fecha	Tipo 0	01
nombreMes	Nombre del mes	Tipo 0	Enero
Anio	Nombre del cargo que desempeña	Tipo 0	2017
Categorías			
Categorías	Jerarquía	Ejemplo	
Tiempo	Año	-	2015
		-	Enero
		-	01
	Mes	-	02
		-	03
		-	Febrero
	Día	-	2016

Dimensión localidad

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: localidad	Clave subrogada de la dimensión	Tipo 0	1
Pk: idUbicacion	Clave operacional de la ubicación geográfica	Tipo 1	4
Id: idParroquia	Identificador del atributo parroquia	Tipo 1	4
Id: idCanton	Identificador del atributo cantón	Tipo 1	3
Id: idProvincia	Identificador del atributo provincia	Tipo 1	2
Id: idZona	Identificador del atributo zona	Tipo 1	1
Parroquia	Nombre de la parroquia	Tipo 1	Maldonado
Canton	Nombre del cantón	Tipo 1	Riobamba
Provincia	Nombre de la provincia	Tipo 1	Chimborazo
Zona	Nombre de la zona	Tipo 1	Zona 3
codigoParroquia	Código externo de parroquia	Tipo 1	0234
codigoCanton	Código externo de cantón	Tipo 1	0112
codigoProvincia	Código externo de provincia	Tipo 1	0012
codigoRegion	Código externo de región	Tipo 1	0001
Categorías			

Categorías	Jerarquía	Ejemplo
Localidad	zona región provincia cantón parroquia	- Zona 3 - Cotopaxi - Chimborazo - Riobamba - Maldonado - Velasco - Veloz - Lizarzaburu - Chambo - Tungurahua

Dimensión Remitente

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: remitente	Clave subrogada de la dimensión	Tipo 0	1
Pk: idRemitente	Clave operacional del remitente	Tipo 1	1
Pk: idCargo	Clave operacional del cargo	Tipo 1	1
Pk: idDependencia	Clave operacional de la dependencia	Tipo 1	1
nombreRemitente	Nombres y apellidos del remitente	Tipo 1	Ing. Christian Aguirre Merino
identificacionRemitente	Identificación del remitente	Tipo 1	0603785163
detalleDependencia	Detalle de la dependencia	Tipo 1	ESPOCH
detalleCargo	Detalle del cargo	Tipo 1	Director de la Escuela de Ecoturismo
estadoRegistro	Estado del registro del remitente	Tipo 1	A
Categorías			
Remitente - Estado	Estado del registro	-Insertado -Actualizado -Eliminado	

Dimensión Persona

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: persona	Clave subrogada de la dimensión	Tipo 0	1
Pk: idPersonaAsignada	Clave operacional del remitente	Tipo 1	1
nombrePersona	Nombres y apellidos del remitente	Tipo 1	Ab. Marco Paguay
identificacionPersona	Identificación del remitente	Tipo 1	1002299793
estadoRegistro	Estado del registro del remitente	Tipo 1	A
Categorías			
Remitente – Estado	Estado del registro	-Insertado -Actualizado -Eliminado	

Dimensión Área

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: area	Clave subrogada del área	Tipo 0	1
Pk: idArea	Clave operacional del área	Tipo 1	1
detalleArea	Detalle del área	Tipo 1	Patrimonio Material
estadoRegistro	Estado del registro del área	Tipo 1	A
Categorías			
Remitente – Estado	Estado del registro	-Insertado -Actualizado -Eliminado	

Dimensión Receptor

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: receptor	Clave subrogada de la dimensión	Tipo 0	1
Pk: idReceptor	Clave operacional del remitente	Tipo 1	1
Pk: idCargo	Clave operacional del cargo	Tipo 1	1
nombreReceptor	Nombres y apellidos del remitente	Tipo 1	Germán Gavilanes
identificacionReceptor	Identificación del remitente	Tipo 1	0603785163
detalleCargo	Detalle del cargo	Tipo 1	Particular
estadoRegistro	Estado del registro del receptor	Tipo 1	A

Categorías		
Remitente – Estado	Estado del registro	-Insertado -Actualizado -Eliminado

Dimensión Medio de Entrega

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: medioEntrega	Clave subrogada del medio de entrega	Tipo 0	1
Pk: IdMedioEntrega	Clave operacional del medio de entrega	Tipo 1	1
detalleMedioEntrega	Detalle del medio de entrega	Tipo 1	Correos del ecuador
estadoRegistro	Estado del registro del medio de entrega	Tipo 1	A
Categorías			
Remitente – Estado	Estado del registro	-Insertado -Actualizado -Eliminado	

Dimensión Tipo de trámite

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: tipo	Clave subrogada del tipo de trámite	Tipo 0	1
Pk: IdTipoTramite	Clave operacional del tipo de trámite	Tipo 1	1
detalleTipoTramite	Detalle del tipo de trámite	Tipo 1	Prioritario
estadoRegistro	Estado del registro del tipo de trámite	Tipo 1	A
Categorías			
Remitente – Estado	Estado del registro	-Insertado -Actualizado -Eliminado	

Dimensión Estado del proceso del trámite

Atributos			
Atributo	Descripción	Cambio	Ejemplo
Key: estado	Clave subrogada del estado de proceso de trámite	Tipo 0	1
Pk: IdEstadoTramite	Clave operacional del estado de proceso de trámite	Tipo 1	1
detalleEstadoTramite	Detalle del estado del proceso de trámite	Tipo 1	Inicio Proceso Fin
estadoRegistro	Estado del registro del tipo de trámite	Tipo 1	A
Categorías			
Remitente – Estado	Estado del registro	-Insertado -Actualizado -Eliminado	

4.1.6. Identificación de la fuente de datos para tablas de hechos y dimensiones

La fuente de datos es la base de datos del sistema de control de trámites; la correspondencia entre las tablas se detalla en la siguiente tabla

<i>Data Warehouse</i>	Base relacional
Tabla de hechos de trámites	• T_TRAMITE • T_PROCESOTRAMITE
Tabla de hechos respuestas	• T_RESPUESTATRAMITE
Dimensión Tiempo	• T_PROCESOTRAMITE • T_RESPUESTATRAMITE
Dimensión Localidad	• T_UBICACION
Dimensión Remitente	• T_REMITENTE • T_CARGO • T_DEPENDENCIA
Dimensión Persona	• T_PERSONAASIGNADA
Dimensión Área	• T_AREA
Dimensión Receptor	• T_RECEPTOR

Dimensión Medio Entrega	T_MEDIOENTREGA
Dimensión Tipo	T_TIPOTRAMITE
Dimensión Estado Trámite	T_ESTADOPROCESOTRAMITE

4.1.7. Mapeo detallado de la fuente de datos

Fuente		Destino	
Dimensión Tiempo			
Columna	Tipo Dato	Tabla	Columna
Key: tiempo*	Integer	---	---
fecha*	Date	- T_TRAMITE - T_PROCESOTRAMITE - T_RESPUESTATRAMITE	FECHADOCUMENTO
dia*	Integer		FECHAPROCESO
mes*	Integer		FECHARESPUESTA
anio*	Integer		
nombreMes*	Varchar(30)		

Fuente		Destino	
Dimensión Localidad			
Columna	Tipo Dato	Tabla	Columna
Key: localidad*		---	---
Pk: idUbicacion	Integer	- T_UBICACION - T_UBICACIONNIVEL	
id: idParroquia*	Integer		IDUBICACIONNIVEL
id: idCanton*	Integer		IDUBICACION
id: idProvincia*	Integer		IDUBICACIONPADRE
id: idZona*	Integer		
Parroquia	Varchar(50)		DETALLE
Canton	Varchar(50)		
Provincia	Varchar(50)		
Zona	Varchar(50)		
CodigoParroquia*	Varchar(20)		IDEXTERNO
CodigoCanton*	Varchar(20)		
CodigoProvincia*	Varchar(20)		
CodigoZona*	Varchar(20)		
Fuente		Destino	
Dimensión Remitente			
Columna	Tipo Dato	Tabla	Columna

Key: remitente*	Integer	---	---
Pk: idRemitente	Integer	• T_REMITENTE	• IDREMITENTE
nombreRemitente	Varchar(150)		• NOMBRE
identificacionRemitente	Varchar(13)		• IDENTIFICACION
Pk: idDependencia	Integer	• DEPENDENCIA	• IDDEPENDENCIA
detalleDependencia	Varchar(50)		• DETALLE
Pk: idCargo	Integer	• CARGO	• IDCARGO
detalleCargo	Varchar(50)		• NOMBRE
estadoRegistro	Char(1)	---	---

Fuente		Destino	
Dimensión Persona			
Columna	Tipo Dato	Tabla	Columna
Key: persona*	Integer	---	---
Pk: idPersonaAsignada	Integer	• T_PERSONAASIGNADA	• IDPERSONAASIGNADA
nombrePersona	Varchar(150)		• NOMBRE
identificacionPersona	Varchar(13)		• IDENTIFICACION
estadoRegistro	Char(1)	---	---

Fuente		Destino	
Dimensión Área			
Columna	Tipo Dato	Tabla	Columna
Key: area*	Integer	---	---
Pk: idArea	Integer	● T_AREA	● IDAREA
detalleArea	Varchar(150)		● DETALLE
estadoRegistro	Char(1)	---	---

Fuente		Destino	
Dimensión Receptor			
Columna	Tipo Dato	Tabla	Columna
Key: receptor*	Integer	---	---
Pk: idReceptor	Integer	• T_RECEPTOR	• IDRECEPTOR
nombreReceptor	Varchar(150)		• NOMBRE
identificacionReceptor	Varchar(13)		• IDENTIFICACION
Pk: idCargo	Integer	• CARGO	• IDCARGO

detalleCargo	Varchar(50)		• NOMBRE
estadoRegistro	Char(1)	---	---

Fuente		Destino	
Dimensión Medio entrega			
Columna	Tipo Dato	Tabla	Columna
Key: medioEntrega	Integer	---	---
Pk: IdMedioEntrega	Integer	● T_MEDIOENTREGA	● IDMEDIOENTREGA
detalleMedioEntrega	Varchar(150)		● DETALLE
estadoRegistro	Char(1)	---	---

Fuente		Destino	
Dimensión Tipo			
Columna	Tipo Dato	Tabla	Columna
Key: tipo	Integer	---	---
Pk: IdTipoTramite	Integer	● T_TIPOTRAMITE	● IDTIPOTRAMITE
detalleTipoTramite	Varchar(150)		● DETALLE
estadoRegistro	Char(1)	---	---

Fuente		Destino	
Dimensión Estado			
Columna	Tipo Dato	Tabla	Columna
Key: tipo	Integer	---	---
Pk: IdEstadoTramite	Integer	● T_ESTADOPROCESO TRAMITE	● IDESTADO
detalleEstadoTramite	Varchar(150)		● DETALLE
estadoRegistro	Char(1)	---	---

(*) Están sujetas a generaciones, transformaciones; detallados en el proceso ETL

5. Diseño físico, construcción y despliegue del *Data Mart*

5.1. Diseño Físico

La conversión del diseño lógico en una base de datos física se realiza en esta etapa; las dimensiones, tablas de hechos se definen como tablas de base de datos respetando los tipos de datos asignados a los atributos conforme se plantee en el mapeo; para ello se realizan los siguientes procesos:

- Elaboración de estándares.
- Diseño del modelo físico
- Plan de índices
- Código DDL para base de datos
- Cubos OLAP

5.1.1. Elaboración de estándares

Para la elaboración un estándar para los nombres de los objetos de la base de datos y ETL es necesario tener en cuenta los componentes: nombre <nombre>, clase <clase>, calificador <calificador>

5.1.1.1. Estándares para objetos de base

Para el estándar de los nombres de tablas, columnas e índices de la base de datos.

Formato	<clase>_<nombre>	
Clases	dim: representa las tablas de dimensiones	
	fact: representa las tablas de hechos	
Nombre	Corresponde al nombre determinado en el modelo dimensional	
Modelo lógico	Modelo físico	Abreviatura
Dimensión tiempo	dim_tiempo	dim_tim
Dimensión localidad	dim_localidad	dim_loc
Dimensión remitente	dim_remitente	dim_rem
Dimensión persona asignada	dim_persona_asignada	dim_per
Dimensión área	dim_area	dim_are
Dimensión receptor	dim_receptor	dim_rec
Dimensión medio entrega	dim_medio_entrega	dim_med
Dimensión tipo trámite	dim_tipo_tramite	dim_tip
Dimensión estado trámite	dim_estado_tramite	dim_est
Tabla de hechos trámite	fact_tramite	fac_tra
Tabla de hechos de respuesta	fact_respuesta	fac_res

Formato	<clase>_<nombre>_<calificador>	
Clases	Abreviatura de la tabla a la cual pertenece la columna	
Nombre	Corresponde al nombre determinado en el modelo dimensional	
Calificador	Opcional	
	Key: Identifica a la columna clave subrogada de la tabla	
	Id: Identifica la columna primaria heredada de la base de datos fuente	
Ejemplo		
Modelo lógico		Modelo físico
Key: receptor		dim_rec_receptor
Pk: idReceptor		dim_rec_idReceptor
NombreReceptor		dim_rec_nombreReceptor

Formato	<clase>_<nombre>
Clases	ik: representa un índice de clave primaria de la tabla is: representa un índice compuesto por una columna, que no es la clave primaria de la tabla im: representa un índice compuesta por dos o más columnas

5.1.1.2. Estándares para objetos de ETL

Tabla: estándar para nombres de pasos

Formato	Pasos de carga y extracción: <nombre>_<calificador2>(<calificador1>) Pasos de transformación: <nombre>(<calificador1>)
Nombre	Pasos de carga y extracción: abreviatura de la tabla de dimensión o tabla de hechos a la cual representa el paso. Pasos de transformación: Descripción corta sobre la operación que realiza el paso
Calificador1	Opcional para los pasos de transformación DM: La operación se realiza sobre las tablas del <i>Data Mart</i> DS: La operación se realiza sobre la base de datos OR: Registros antiguos de una tabla de hechos NR: Registros nuevos de una tabla de hechos N: Inserción de datos en tabla U: Actualización de datos en tabla D: Eliminación de datos en tabla Adicionalmente se puede utilizar la abreviatura de las tablas sobre las que ocurre la operación para evitar ambigüedades
Calificador2	Opcional Información adicional que describe al paso de carga o extracción, para evitar ambigüedades

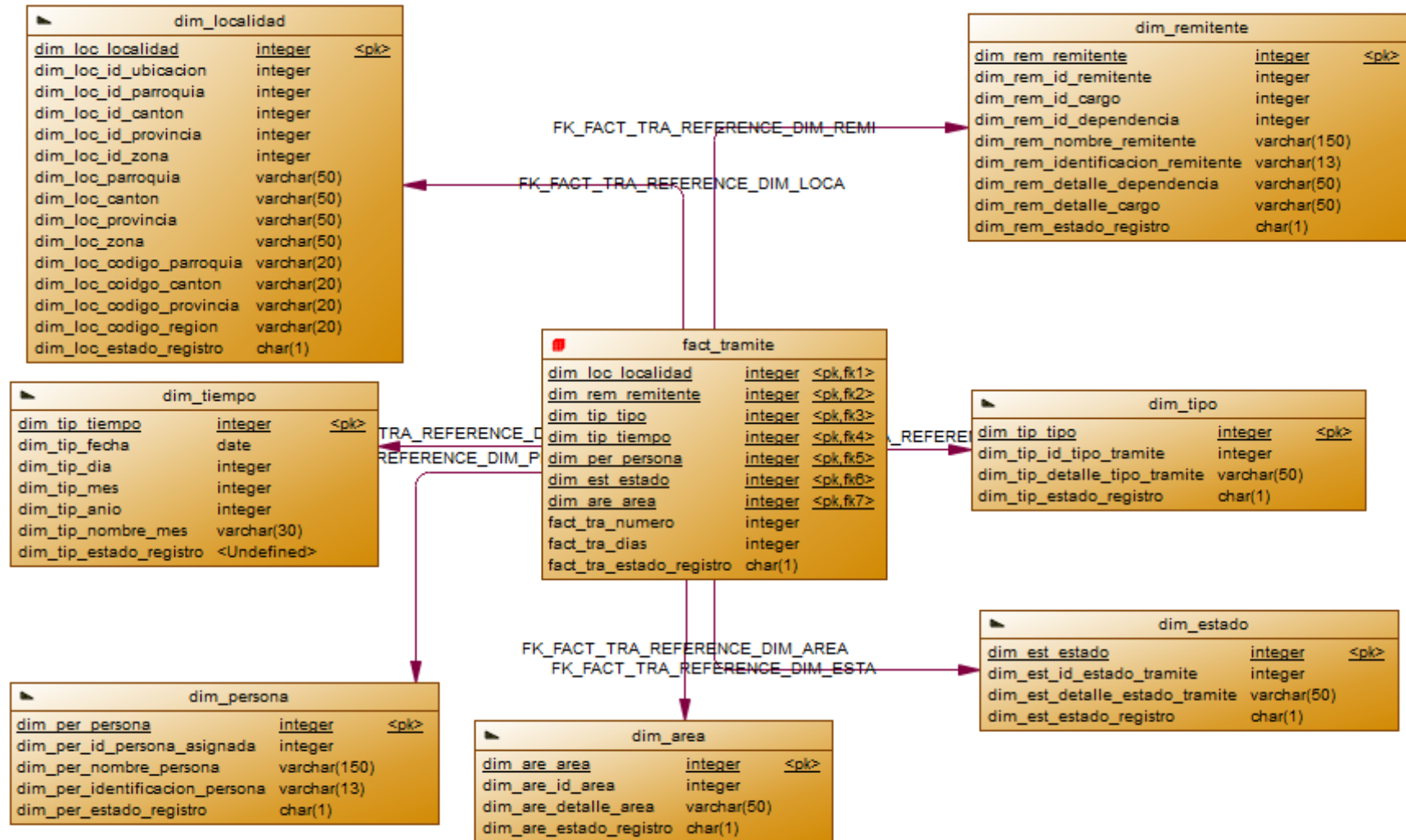
Tabla: estándar para nombres de archivos

Formato	<clase> _<nombre>.<extensión>
Clase	job: Archivos que representa flujos tipo Job. elt: Archivos que representan flujos tipo Transformación val: Archivos que representan flujos tipo Job para realizar validaciones. cal: Archivos que representan flujos tipo Transformación para realizar la definición de parámetros.
Nombre	Abreviatura de la tabla de dimensión o tabla de hechos sobre la cual se ejecuta el ETL.
Extensión	Extensión de los archivos: kjb: para archivos de tipo Job ktr: para archivos de tipo Transformación

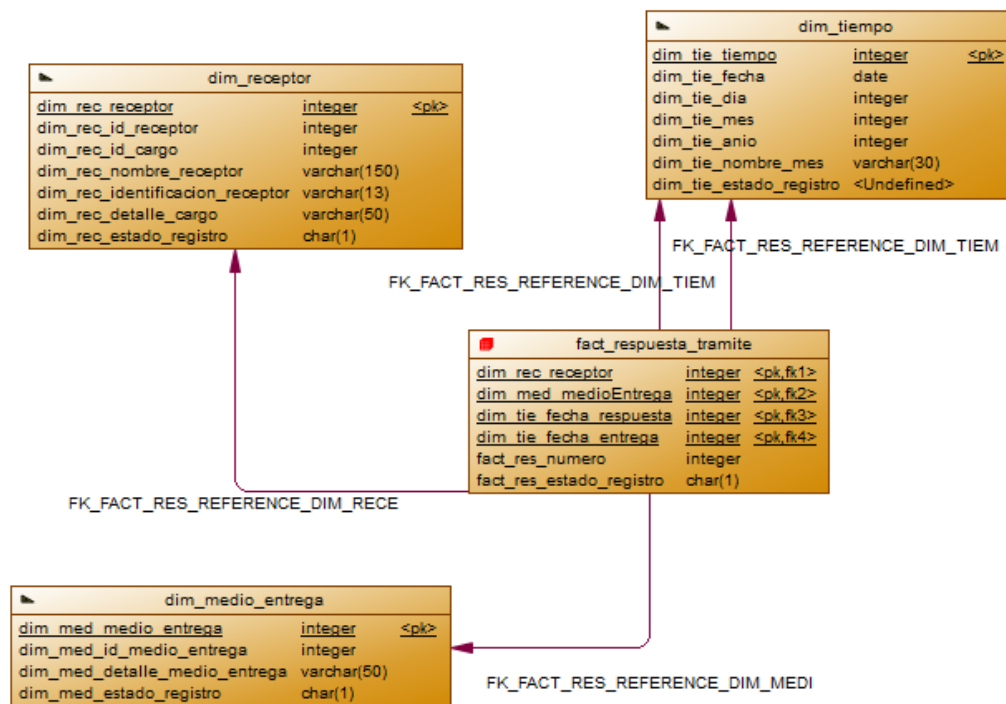
5.1.2. Diseño del modelo físico

La definición del modelo físico se lo ha realizado mediante una herramienta de modelado de datos, introduciendo en el modelo los estándares antes mencionados para la definición de los objetos de la base de datos.

5.1.2.1. Modelo físico de la tabla de hechos de trámites



5.1.2.2. Modelo físico de la tabla de respuestas de trámites



5.1.3. Cubos OLAP

El análisis OLAP de la solución es realizado con la herramienta *SpagoBI*. Por lo tanto, el cubo refleja el diseño del modelo dimensional del *Data Mart*.

6.2.1.1. Estructura del cubo

Para la elaboración del *Data Mart* se establecen los siguientes cubos:

- Trámites Días Tiempo
- Trámite Estado
- Trámite Localidad
- Trámite Remitente
- Trámite Tiempo
- Trámite Tipo

6.2.1.1.1. Estructura Trámites Días Tiempo

- Nombre cubo: Estructura Trámites Días Tiempo
- Nombre de dimensiones: Dim tiempo
- Tabla de hechos: Fact tramite

```
<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
```

```

allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Numero Dias" column="fact_tra_dias" aggregator="avg"
formatString="#,###"/>
</Cube>
</SCHEMA>

```

6.2.1.1.2. Estructura Trámites Estado

- Nombre cubo: Trámites Estado
- Nombre de dimensiones: Dim estado, Dim tiempo
- Tabla de hechos: Fact tramite

```

<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim estado">
<Hierarchy hasAll="true"
allMemberName="Estado"
primaryKey="dim_est_estado"
name="Estado"
>
<Table name="dim_estado"/>
<Level name="Estado tramite" column="dim_est_detalle_estado_tramite"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>

```

```

<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim estado" source="Dim estado"
foreignKey="dim_est_estado"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>

```

6.2.1.1.3. Estructura Trámites Localidad

- Nombre cubo: Trámites Localidad
- Nombre de dimensiones: Dim localidad, Dim tiempo
- Tabla de hechos: Fact tramite

```

<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim localidad">
<Hierarchy hasAll="true"
allMemberName="Localidad"
primaryKey="dim_loc_localidad"
name="Localidad"
>
<Table name="dim_localidad"/>
<Level name="Zona" column="dim_loc_zona"
uniqueMembers="false"
/>
<Level name="Provincia" column="dim_loc_provincia"
uniqueMembers="false"
/>
<Level name="Canton" column="dim_loc_canton"
uniqueMembers="false"
/>
<Level name="Parroquia" column="dim_loc_parroquia"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
<Level name="Fecha" column="dim_tip_fecha"

```

```

uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim localidad" source="Dim localidad"
foreignKey="dim_loc_localidad"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>

```

6.2.1.1.4. Estructura Trámites Remitente

- Nombre cubo: Trámites Remitente
- Nombre de dimensiones: Dim remitente, Dim Tiempo
- Tabla de hechos: Fact tramite

```

<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim remitente">
<Hierarchy hasAll="true"
allMemberName="Remitente"
primaryKey="dim_rem_remitente"
name="Remitente"
>
<Table name="dim_remitente"/>
<Level name="Nombre Remitente" column="dim_rem_nombre_remitente"
uniqueMembers="false"
/>
<Level name="Identificacion" column="dim_rem_identificacion_remitente"
uniqueMembers="false"
/>
<Level name="Cargo" column="dim_rem_detalle_cargo"
uniqueMembers="false"
/>
<Level name="Dependencia" column="dim_rem_detalle_dependencia"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>

```

```

<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim remitente" source="Dim remitente"
foreignKey="dim_rem_remitente"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>

```

6.2.1.1.5. Estructura Trámites Tiempo

- Nombre cubo: Trámites Tiempo
- Nombre de dimensiones: Dim tiempo
- Tabla de hechos: Fact tramite

```

<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>

```

6.2.1.1.6. Estructura Trámites Tipo

- Nombre cubo: Trámites Tipo
- Nombre de dimensiones: Dim Tipo, Dim Tiempo
- Tabla de hechos: Fact tramite

```

<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim tiempo">
<Hierarchy hasAll="True"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Dimension name="Dim tipo">
<Hierarchy hasAll="true"
allMemberName="Tipo"
primaryKey="dim_tip_tipo"
name="Tipo"
>
<Table name="dim_tipo"/>
<Level name="Tipo Tramite" column="dim_tip_detalle_tipo_tramite"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<DimensionUsage name="Dim tipo" source="Dim tipo"
foreignKey="dim_tip_tipo"/>
<Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>

```

6. Proceso de extracción, transformación y carga

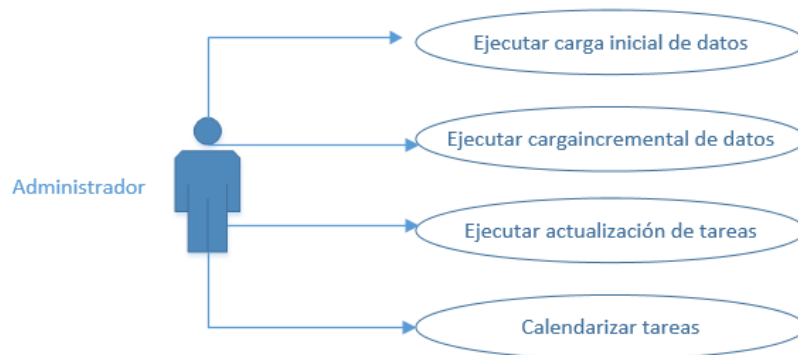
El proceso ETL consiste en la extracción, limpieza, transformación y carga de datos, desde la base de datos fuente hasta el *Data Mart*, está definido mediante las siguientes tareas:

- Elaboración de plan
- Definición de transformaciones y cálculos
- ETL de las dimensiones
- ETL de los hechos

6.1. Elaboración de plan

El proceso de carga de datos desde sistemas fuentes hacia el *Data Warehouse* se realiza en tres pasos

1. Ejecutar carga inicial de datos
2. Ejecutar carga incremental de datos
3. Ejecutar actualización de tareas



Escenario de proceso área ETL

Cada uno de los procesos consta de tres partes:

- Validación

Verifica la existencia de las tablas y archivos para la realización del proceso ETL, en caso de que algún objeto no existe, el proceso es abortado.

- ETL

Extracción, de datos desde sistemas fuentes, transformación de datos, generación y cálculo de medidas; carga de datos al *Data Mart*

- Actualización

Actualización de índices y estadísticas.

Diagrama de flujo ETL para dimensión

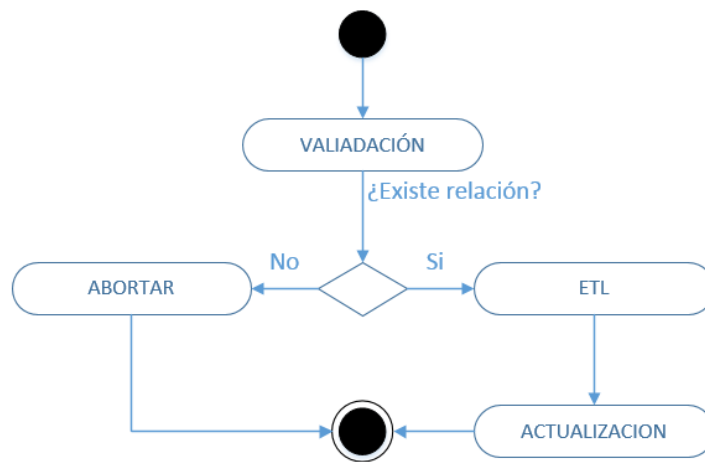
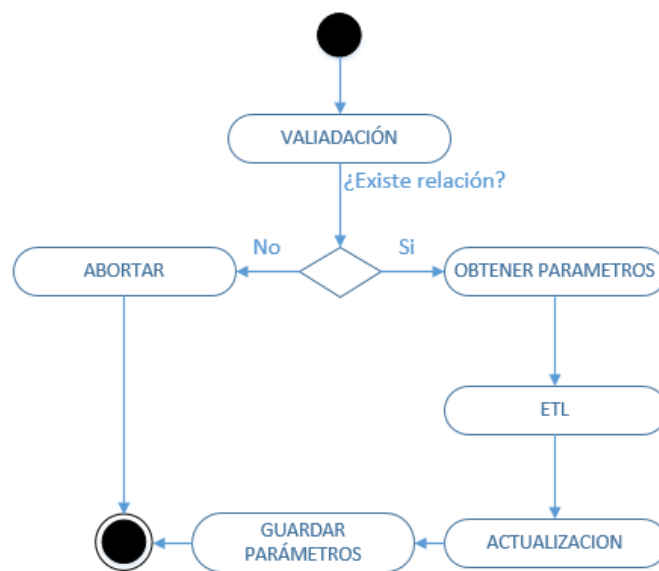


Diagrama de flujo ETL para tablas de hechos



6.1.1. Carga inicial de datos

La primera carga de datos al *Data Mart* consiste en:

1. Creación de la base de datos
2. Cargar la dimensión tiempo
3. Cargar las dimensiones adicionales
4. Cargar tablas de hechos

En esta primera etapa se cargan las tablas de dimensiones y hechos históricos; desde el inicio hasta los ingresados hasta la fecha de corte para la carga inicial.

6.1.2. Carga incremental de datos

En esta etapa se realizan las actualizaciones de dimensiones de acuerdo a las políticas de cambio para cada atributo, en caso de un nuevo registro se insertan; para las tablas de hechos, se cargan los nuevos registros comprendidos entre la última ejecución de la carga incremental o inicial hasta la fecha de corte establecida.

6.1.3. Actualización de datos

Es importante mantener actualizadas las dimensiones, a través de la carga o actualización de registros. La actualización en las tablas de hechos es poco probable, pero permite mantener actualizados los registros correspondientes a un rango de fechas.

6.2. Definición de transformaciones y cálculos

Dimensión Tiempo

Id	Dimensión	Columna	Transformación/cálculo
1	dim_tiempo	dim_tie_dia	Se calcula a partir de la columna dim_tie_fecha, extrayendo el componente "día" Tal que dim_tie_fecha es x, dim_tie_dia es y: $y = \text{día}(x)$
2	dim_tiempo	dim_tie_mes	Se calcula a partir de la columna dim_tie_fecha, extrayendo el componente "mes" Tal que dim_tie_fecha es x, dim_tie_mes es y: $y = \text{mes}(x)$
3	dim_tiempo	dim_tie_dia	Se calcula a partir de la columna dim_tie_fecha, extrayendo el componente "año" Tal que dim_tie_fecha es x, dim_tie_anio es y: $y = \text{año}(x)$

Dimensión Localidad

Id	Dimensión	Columna	Transformación/cálculo
4	dim_localidad	dim_loc_id_zona dim_loc_codigo_zona dim_loc_zona	Datos asignados en base al valor de la columna nombre, de acuerdo al nivel de ubicación. Tal que dim_loc_id_zona es y1, dim_loc_zona es y2, dim_loc_codigo_zona es y3, dim_loc_id_nivel_ubicación es x <pre> switch (x) { case 2: y1=IDUBICACION; y2=DETALLE; y3=IDEXTERNO } </pre>
5	dim_localidad	dim_loc_id_provincia dim_loc_codigo_provincia dim_loc_provincia	Datos asignados en base al valor de la columna nombre, de acuerdo al nivel de ubicación. Tal que dim_loc_id_provincia es y1, dim_loc_provincia es y2, dim_loc_codigo_provincia es y3, dim_loc_id_nivel_ubicacion es x <pre> switch (x) { case 3: y1=IDUBICACION; y2=DETALLE; y3=IDEXTERNO } </pre>
6	dim_localidad	dim_loc_id_canton dim_loc_codigo_canton dim_loc_canton	Datos asignados en base al valor de la columna nombre, de acuerdo al nivel de ubicación. Tal que dim_loc_id_canton es y1, dim_loc_canton es y2, dim_loc_codigo_canton es y3, dim_loc_id_nivel_ubicacion es x

			<pre> switch (x) { case 4: y1=IDUBICACION; y2=DETALLE; y3=IDEXTERNO } </pre>
6	dim_localidad	dim_loc_id_parroquia dim_loc_codigo_parroquia dim_loc_parroquia	Datos asignados en base al valor de la columna nombre, de acuerdo al nivel de ubicación. Tal que dim_loc_id_parroquia es y1, dim_loc_parroquia es y2, dim_loc_codigo_parroquia es y3, dim_loc_id_nivel_parroquia es x <pre> switch (x) { case 4: y1=IDUBICACION; y2=DETALLE; y3=IDEXTERNO } </pre>

Tabla de hechos trámites

Id	Dimensión	Columna	Transformación/cálculo
7	fact_tramite	fact_tra_dias	Se calcula en base a las fechas de recepción de trámite y la fecha de finalización del trámite, se restan y se transforma el resultado a días Tal que, fact_tra_dias es y, dim_tie_fecha_proceso x, dim_est_estado z <pre> if (z=1) { fecha1=FECHAPROCESO; </pre>

			<pre> } if(z=2) { Fecha2=FECHAPROCESO; } y=cast_to_int(convertir_a_dias(fecha2- fecha1)) </pre>
8	fact_tramite	fact_tra_numero	Se calcula en base al número de registros de tramite Tal que fact_tra_numero es y, fact_tra_id_tramite es x y=count(x);

6.3. ETL de las tablas de dimensión

Para el proceso ETL de las tablas de dimensiones se consideran los siguientes pasos

- Extracción de datos origen
Mediante código SQL se extraen los datos de las columnas de cada una de las dimensiones. En este paso se lleva a cabo la primera transformación
- Transformación y cálculo

Se realiza las transformaciones y cálculos antes definidos, de ser necesario se añaden columnas a los registros

- Asignación de clave subrogada

En base a las claves operacionales se asigna un número secuencial único al registro, esta es la clave subrogada

- Carga y utilización de datos

Los registros del paso anterior son cargados a la tabla de dimensiones, los registros nuevos son insertados mientras los alterados son actualizados

6.3.1. ETL de la dimensión tiempo

1. Se generan 45000 registros con la (1900-01-01) por defecto
2. Para cada registro se asigna un número secuencial comenzando desde 1
3. Se lleva a cabo las transformaciones 1,2,3

4. Se filtran las fechas para ingresar solo valores hasta un año determinado se ordena las fechas en base a la columna dim_tie_fecha
5. Se cargan los registros en la dimensión

6.3.2. ETL de la dimensión localidad

1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión localidad, la extracción se realizó en dos etapas
 - Asignación de códigos y detalles de zonas, regiones, provincias, parroquias
 - Generación de datos mediante script sql
2. Se combinan los registros de la extracción
3. Se llevan a cabo las transformaciones 4,5,6
4. Se asignan la clave subrogada para cada registro en base a la clave operacional: dim_loc_id_ubicacion
5. Se cargan y actualizan registros

6.3.3. ETL de la dimensión remitente

1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional dim_rem_id_remitente, dim_rem_id_cargo, dim_rem_id_dependencia
3. Se cargan y actualizan los registros en la dimensión

6.3.4. ETL de la dimensión persona

1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional dim_rem_id_persona_asignada
3. Se cargan y actualizan los registros en la dimensión

6.3.5. ETL de la dimensión área

1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional dim_are_id_area
3. Se cargan y actualizan los registros en la dimensión

6.3.6. ETL de la dimensión receptor

1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional dim_rec_id_receptor, dim_rec_id_cargo
3. Se cargan y actualizan los registros en la dimensión

6.3.7. ETL de la dimensión medio entrega

1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional
dim_med_id_medio_entrega
3. Se cargan y actualizan los registros en la dimensión

6.3.8. ETL de la dimensión tipo trámite

1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional
dim_med_id_tipotramite
3. Se cargan y actualizan los registros en la dimensión

6.3.9. ETL de la dimensión estado

1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional dim_med_id_estadoe
3. Se cargan y actualizan los registros en la dimensión

6.3.10. ETL de las tablas de hechos

- Extracción de datos de las tablas origen
- Transformaciones y cálculos
- Obtención de claves dimensionales
- Validación de claves
- Extracción de datos de la tabla de hechos
- Comparación y Marcado de filas
- Actualización de registros

Anexos

Entrevista al Director del INPC R3

- Cuáles son los objetivos que la institución persigue
- Cuáles son las áreas o departamentos de la organización
- Cuáles son las etapas por las atraviesa un proceso de trámite
- Cuáles son las dependencias que participan en cada uno de los procesos
- Cuáles son los reportes que se solicitan con mayor frecuencia
- Como son obtenidos estos reportes en la actualidad
- Cuál es el objetivo que se persigue al obtener estos datos
- ¿Se ha solicitado información no prevista en los reportes anteriores?, si es así, ¿cuánto tiempo se tomó recabar la información?
- Que otros reportes serían de apoyo para la gestión de trámites
- La información histórica, es necesaria para la gestión, ¿cuál es el tiempo a considerar para que la información sea útil para la gestión?

Entrevista Jefe de sistemas

- ¿Cuáles son los orígenes de datos?
- ¿Cuáles son las herramientas mediante las cuales se gestiona la información?
- ¿Cuáles son los reportes más frecuentes?
- ¿Cómo se gestionan los reportes no previstos?
- ¿Existen factores que disminuyan la calidad de los reportes?
- ¿Ha considerado las ventajas de implementar un *Data Warehouse* para la obtención de información que requiere?

Apéndice B

METODOLOGÍA HEFESTO

Introducción	172
1. Análisis de la situación actual	172
2. Identificación de la institución.....	172
1.1. Visión.....	172
1.2. Misión.....	172
1.3. Objetivos.....	172
1.4. Objetivos Estratégicos.....	173
1.5. Organigrama.....	175
3. Relación de las metas de la organización con el <i>Data Warehouse</i>	175
4. Responsabilidades	176
5. Ambiente tecnológico.....	176
5.1. Producto seleccionado para la DBMS: Justificación de MySQL	176
5.2. Producto seleccionado para el área ETL	176
5.3. Producto seleccionado para el motor OLAP	176
5.4. Producto seleccionado como herramienta de acceso a datos	176
5.5. Producto seleccionado para la generación de cubos multidimensionales.....	176
6. Pasos y aplicación metodológica	176
7. Características:	177
7.1. PASO 1) ANÁLISIS DE REQUERIMIENTOS.....	177
7.1.1. Identificar preguntas.....	177
7.1.2. Identificar indicadores y perspectivas	178
7.1.3. Modelo Conceptual.....	180
7.2. PASO 2) ANÁLISIS DE LOS OLTP	180
7.2.1. Conformar indicadores	180
7.2.2. Establecer correspondencias	181
7.2.3. Nivel de granularidad	183
7.2.4. Modelo Conceptual ampliado.....	186
7.3. PASO 3) MODELO LÓGICO DEL DW	190
7.3.1. Tipo de Modelo Lógico del DW.....	190
7.3.2. Tablas de dimensiones	190
7.3.3. Tablas de hechos	194

7.3.4.	Uniones	196
7.4.	PASO 4) INTEGRACIÓN DE DATOS.....	197
7.4.1.	Carga Inicial	197
7.4.2.	Actualización.....	202
8.	Creación de Cubos Multidimensionales	203
8.1.	Estructura del cubo	203
8.1.1.	Estructura Trámites Días Tiempo.....	203
8.1.2.	Estructura Trámites Estado	204
8.1.3.	Estructura Trámites Localidad.....	204
8.1.4.	Estructura Trámites Remitente	206
8.1.5.	Estructura Trámites Tiempo	207
8.1.6.	Estructura Trámites Tipo.....	207

Introducción

Hefesto se ha convertido con el paso de los años en una metodología robusta para la creación de *Data Warehouse* (DW) desde cero, su pragmatismo facilita la adaptación con cualquier ciclo de vida al enfocarse en el análisis de requerimientos, análisis de fuentes de datos para la elaboración e implementación de DW.

1. Análisis de la situación actual

Hesfesto no precisa de fases extensas de desarrollo monolítico que involucre tiempo en exceso y fases de despliegue largas, busca priorizar en la primera implementación satisfacer las necesidades casi por completo demostrando así las ventajas de un DW y motivando a los usuarios.

2. Identificación de la institución

1.1. Visión

En 2017 el Instituto Nacional de Patrimonio Cultural se consolidará como un centro generador de conocimiento sobre los elementos que constituyen el sustento de nuestra identidad nacional multiétnica y pluricultural. Se fundamentará en la investigación formal, en la conservación y en la difusión multifacética del patrimonio cultural y de la memoria histórica del Ecuador. Este caudal será difundido para concienciar a los actores involucrados y a la ciudadanía sobre la importancia de las raíces y de la identidad. Así, mediante el uso y desarrollo de teorías, métodos y técnicas tanto inter como transdisciplinarias, el INPC R3 facilitará el acceso, el uso social y la preservación de todo el acervo cultural de la Nación.

1.2. Misión

El Instituto Nacional de Patrimonio Cultural es una entidad del sector público con ámbito nacional, encargada de aplicar las políticas sectoriales, así como de investigar, normar, regular y asesorar para la preservación, conservación, apropiación, promoción y uso adecuado del patrimonio cultural material e inmaterial.

1.3. Objetivos

- Consolidar al Instituto como un Centro de Investigación-Desarrollo y Sistematización del Conocimiento Científico-Técnico del patrimonio cultural, mediante el estudio, análisis y aplicación de teorías, metodologías y técnicas para: catalogar, documentar, conservar, proteger y potenciar el patrimonio material o tangible e inmaterial o intangible, con la

finalidad de difundir hacia los actores involucrados las mejores opciones de conservación del patrimonio.

- Desarrollar, sugerir y vigilar la implementación y el cumplimiento de normas, políticas y procedimientos, orientados para la conservación preventiva del patrimonio cultural, a fin de regular y controlar el adecuado manejo del patrimonio por parte de los actores sociales, esto incluye: prevención de riesgo (identificación, conservación y alerta), uso (vulnerabilidad patrimonial) y seguridad (ámbito legal- jurídico).
- Diseñar e implementar sistemas de información para la gestión de los bienes patrimoniales y culturales: El INPC R3 tiene como fin la recopilación de la documentación producida en la institución a nivel nacional, tanto histórica como actual. Toda la información que genere esta base de datos, servirá para la GESTIÓN DEL PATRIMONIO sobre todo para los gobiernos autónomos descentralizados GADs, cuyo mandato constitucional ahora es total. Por otro lado, esta herramienta permitirá conocer nuestros bienes culturales y patrimoniales a través de la WEB y estará al servicio de los estudiosos y de la sociedad en general.
- Sistematizar y comunicar el conocimiento patrimonial material o tangible e inmaterial o intangible mediante la elaboración de publicaciones científicas, las cuales deberán difundirse a través de medios y canales adecuados, así como en el Ministerio de Cultura, mismos que pueden ser conferencias, ponencias y foros públicos, documentos impresos y/o virtuales, así como la realización de cursos propios de sus áreas de especialidad que propendan a la formación de técnicos y personal especializado en materia de patrimonio y seguridad patrimonial.

1.4. Objetivos Estratégicos

Incrementar la investigación y la generación de conocimiento sobre el patrimonio cultural y la memoria histórica (INPC, 2015).

Estrategias:

Formular y ejecutar proyectos de investigación.

Suscribir acuerdos y convenios con la academia y entidades públicas y privadas a nivel nacional e internacional.

Desarrollar nuevas líneas de investigación de patrimonio.

Producir publicaciones científicas de acuerdo a los estándares internacionales de Investigación.

Incrementar la conservación y salvaguardia del patrimonio cultural del Ecuador

Estrategias:

Fortalecer la prevención de los riesgos antrópicos y naturales que afectan al patrimonio cultural del Ecuador.

Controlar el cumplimiento de la normativa para la preservación y salvaguardia del patrimonio cultural.

Actualizar la normativa para la preservación y salvaguardia del patrimonio cultural.

Formular normativa para el combate del tráfico ilícito de bienes culturales patrimoniales.

Fortalecer las acciones y el control del tráfico ilícito de los bienes culturales.

Ampliar el registro e inventario de los bienes culturales del Estado.

Desarrollar los planes de salvaguardia y conservación del patrimonio cultural.

Identificar los patrimonios culturales emergentes y resignificados con la participación de las comunidades.

Incrementar la transferencia de conocimiento del patrimonio cultural a la ciudadanía

Estrategias:

Desarrollar el sistema de información de bienes culturales y patrimoniales.

Fortalecer la producción editorial de la institución.

Implementar un plan de promoción y capacitación sobre el patrimonio cultural

Implementar nuevas tecnologías de información y comunicación.

Diseñar programas de comunicación para difundir los patrimonios culturales emergentes y resignificados.

Incrementar el fortalecimiento institucional del INPC R3, a través de la eficiencia operacional, con énfasis en el desarrollo del talento humano y la desconcentración administrativa - financiera, contribuyendo así a mejorar la calidad del servicio público en el ámbito del patrimonio cultural

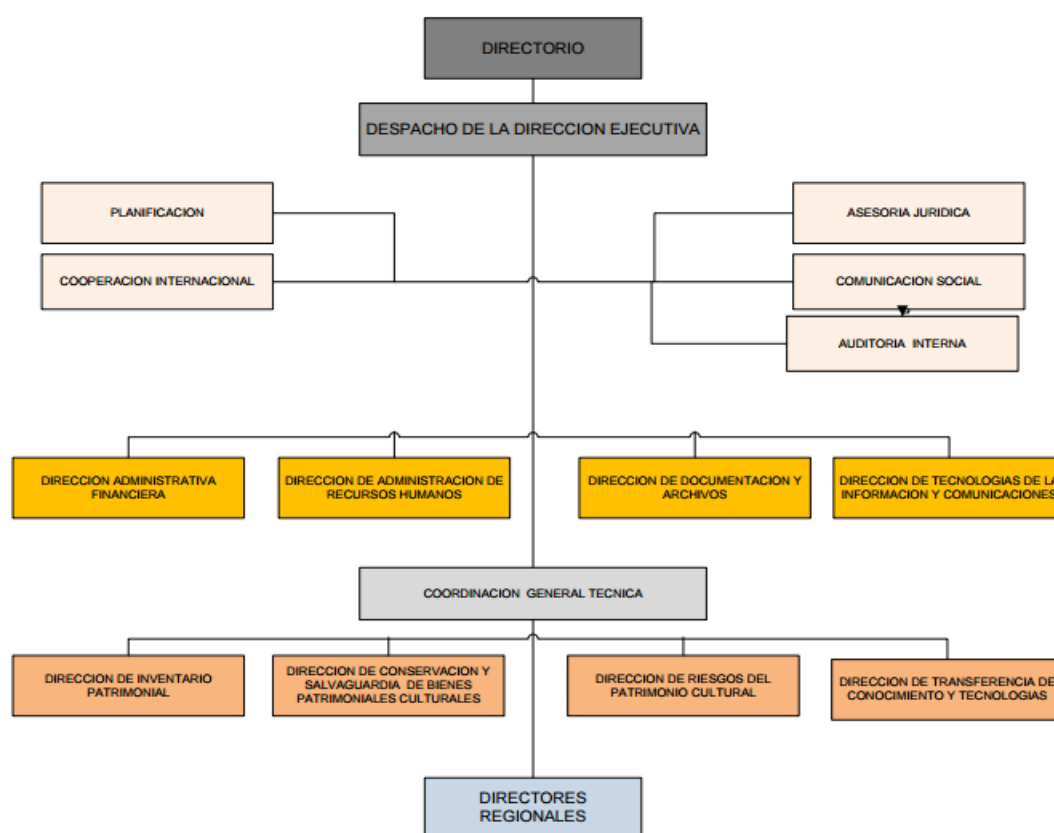
Estrategias:

Fortalecer la aplicación de mecanismos de selección, profesionalización, promoción, seguimiento y evaluación del Talento Humano del INPC R3, a fin de garantizar la eficiencia y la calidad, a través de una gestión integral.

Administrar el presupuesto donde se ejecute los recursos asignados al INPC R3 mediante la eficiencia y transparencia en el manejo de los mismos, procurando proveer las necesidades institucionales de forma equitativa y oportuna.

Diseñar e implementar un sistema de gestión de la calidad que estandarice los procesos del INPC R3 con criterios de calidad y excelencia, con la aplicación de buenas prácticas y con la adopción de estándares internacionales.

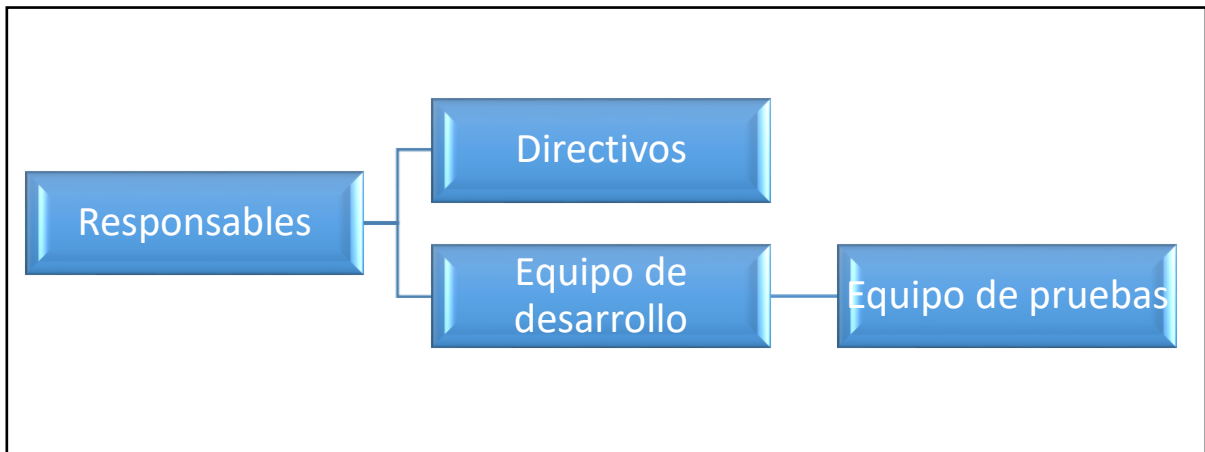
1.5. Organigrama



3. Relación de las metas de la organización con el *Data Warehouse*

Data Warehouse coincide con las metas que desea alcanzar el INPC R3, ya que por medio del uso de almacenes de datos facilitaría la toma de decisiones y el acceso de los datos de cada uno de los proyectos de investigación a los que se les da seguimiento.

4. Responsabilidades



5. Ambiente tecnológico

5.1. Producto seleccionado para la DBMS: Justificación de MySQL

Nombre	MySQL
Versión	5.7.18

5.2. Producto seleccionado para el área ETL

Nombre	SpagoBI studio
Versión	5.1

5.3. Producto seleccionado para el motor OLAP

Nombre	SpagoBI studio
Versión	5.1

5.4. Producto seleccionado como herramienta de acceso a datos

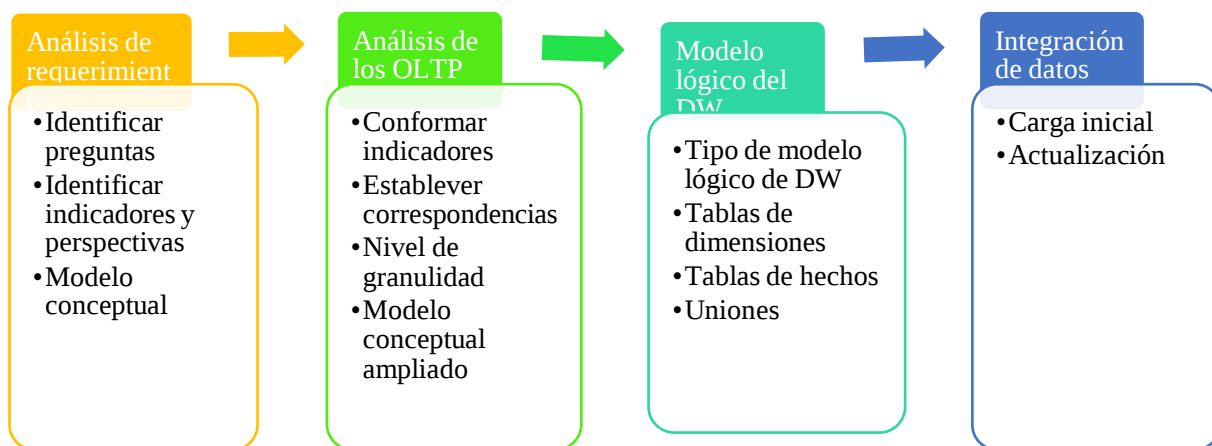
Nombre	SpagoBI studio
Versión	5.1

5.5. Producto seleccionado para la generación de cubos multidimensionales

Nombre	SpagoBI studio
Versión	5.1

6. Pasos y aplicación metodológica

Para facilitar la construcción del modelo conceptual tenemos los siguientes pasos a seguir:



7. Características:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tome decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el siguiente paso.
- Se aplica tanto para *Data Warehouse* como para *Data Mart*.

7.1. PASO 1) ANÁLISIS DE REQUERIMIENTOS

Esta metodología se utilizará para la construcción de un *Data Warehouse* que alcance las metas a las que se enfoca INPC R3, como el área de recepción y tramitación de documentos.

7.1.1. Identificar preguntas

Mediante las entrevistas realizadas a los usuarios se ha determinado la creación de un OLTP, que permita el tratamiento del proceso de control de trámites.

Al identificar las variables o perspectivas a utilizar para la toma de decisiones mediante las respuestas de los usuarios se determinando cuales indicadores representan mejor este proceso de trámites.

Las preguntas obtenidas fueron las siguientes:

- Se requiere conocer el número de trámites receptados en un periodo de tiempo.
- Se requiere conocer el número de trámites por remitente y dependencia en un periodo de tiempo.
- Se requiere conocer el personal asignado durante el proceso de duración del trámite.
- Se requiere conocer el número de trámites por estado en un periodo de tiempo.
- Se requiere conocer el número de trámites por procedencia geográfica a ser procesados.
- Se requiere conocer el número de trámites por prioridad.
- Se requiere conocer el número de días transcurridos hasta la finalización del trámite.
- Se requiere conocer el número de respuestas por medio de entrega en un periodo de tiempo

7.1.2. Identificar indicadores y perspectivas

Número de trámites receptados en un período de tiempo.



Número de trámites por remitente y dependencia dentro de un periodo de tiempo.



Conocer las personas asignadas durante el proceso del trámite dentro de un período de tiempo.



Listar los trámites por prioridad (tipos) en un periodo de tiempo.



Identificar número de trámites por estados dentro de un periodo de tiempo.



Número de trámites por procedencia geográfica en un periodo determinado de tiempo.



Conocer el número de días transcurridos hasta la finalización del trámite



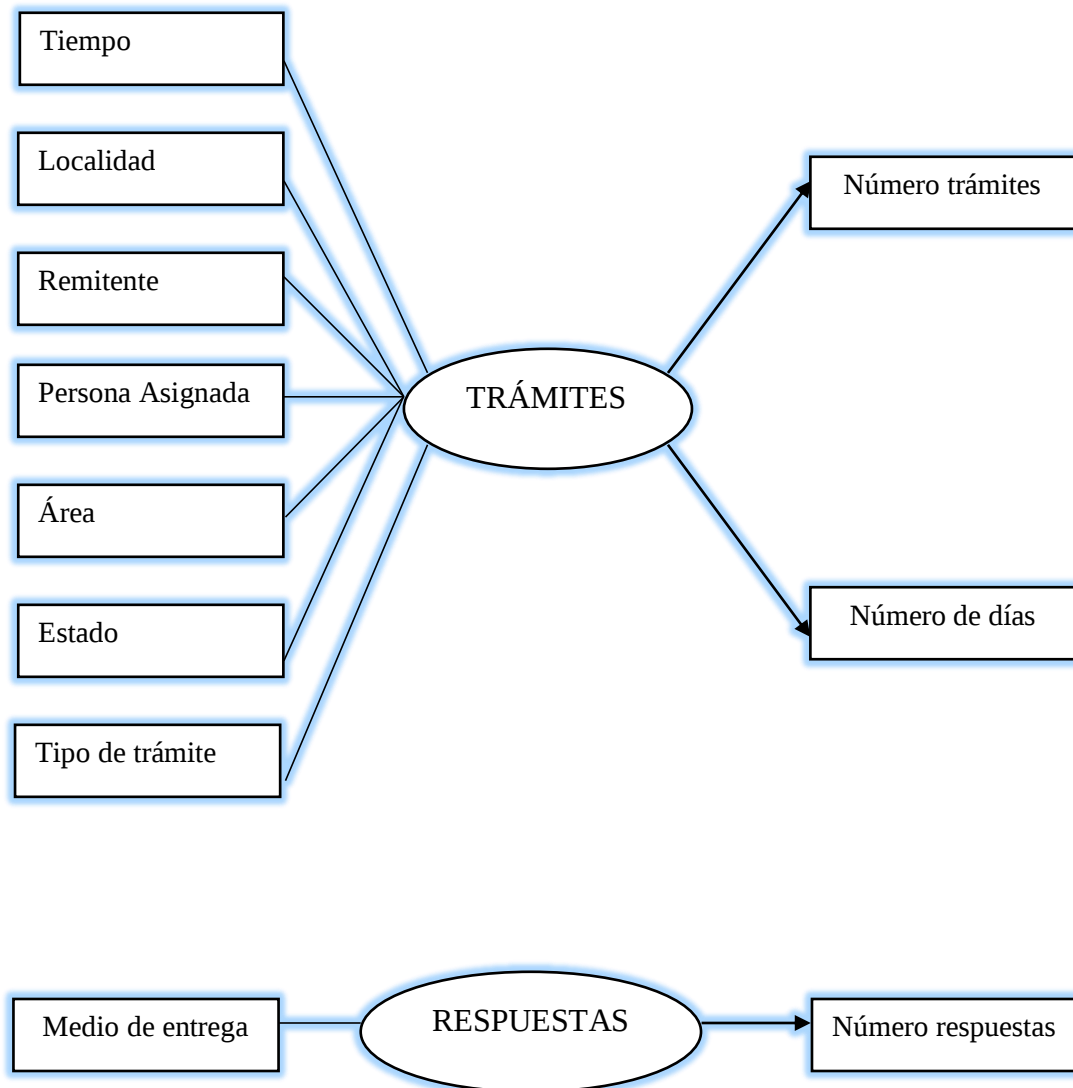
- Respuestas
- Número de días

Las perspectivas del análisis son:

- Tiempo
- Localidad
- Remitente
- Persona asignada
- Área
- Medio de entrega
- Estado
- Tipo de trámite

7.1.3. Modelo Conceptual

El diseño de los modelos parte de los indicadores y perspectivas obtenidas con anterioridad, los cuales permiten la observación de los alcances.



7.2. PASO 2) ANÁLISIS DE LOS OLTP

Al analizar los OLTP se determina la manera de realizar los cálculos de los indicadores y el establecimiento de las correspondencias del modelo conceptual y las fuentes de datos.

7.2.1. Conformar indicadores

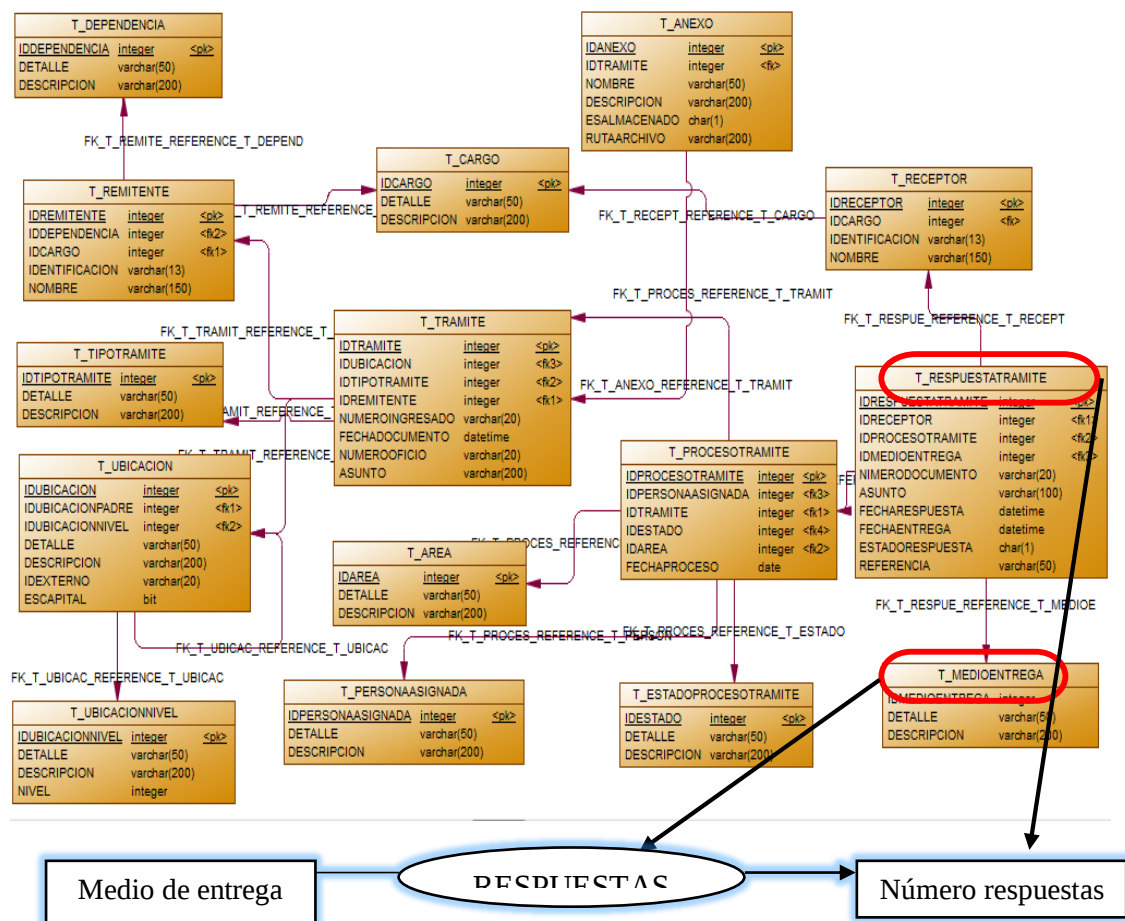
- Trámites
 - Hechos: Números de trámites
 - Función de diferencia entre fechas: DAYDIFF

- ### 7.2.2. Establecer correspondencias



Relaciones identificadas:

- La tabla “dim_tiempo” se relaciona con la perspectiva “Tiempo”.
- La tabla “dim_localidad” se relaciona con la perspectiva “Localidad”.
- La tabla “dim_remitente” se relaciona con la perspectiva “Remitente”.
- La tabla “dim_persona” se relaciona con la perspectiva “Persona”.
- La tabla “dim_area” se relaciona con la perspectiva “Área”.
- La tabla “dim_estado” se relaciona con la perspectiva “Estado”.
- La tabla “dim_tipo” se relaciona con la perspectiva “Tipo de trámite”.
- El campo de “fact_tra_numero” de la tabla fact_tramite almacena el conteo del indicador “Número de trámites”.
- El campo “dim_tie_fecha_proceso” con el campo “dim_est_estado” en la comparación del tipo de estado en recepción restado el campo “dim_tie_fecha_proceso” con el campo “dim_est_estado” en la comparación del tipo de estado en finalizado se almacena en el campo “fact_numero_dias”.



Relaciones identificadas:

- La tabla “dim_medio_entrega” se relaciona con la perspectiva “Medio de entrega”.
- El campo fact_res_numero se realiza el conteo con el indicador Número de respuestas.

7.2.3. Nivel de granularidad

Siendo puntualizadas las relaciones con los OLTP, se identificó los campos que contienen cada perspectiva que contribuyen con la elaboración de consultas.

Perspectiva Tiempo:

ATRIBUTO	DESCRIPCIÓN
Fecha	Fecha completa: año, mes, día
Día	Día en la fecha
Mes	Mes en la fecha
Año	Año en la fecha
nombre_mes	Nombre del mes
estado_registro	Estado del registro

Perspectiva Localidad

ATRIBUTO	DESCRIPCIÓN
id_ubicacion	Clave operacional de la ubicación geográfica
id_Parroquia	Identificador del atributo parroquia
id_Canton	Identificador del atributo cantón
id_provincia	Identificador del atributo provincia
id_zona	Identificador del atributo zona
nombre_parroquia	Nombre de la parroquia
nombre_canton	Nombre del cantón
nombre_provincia	Nombre de la provincia
codigo_zona	Nombre de la zona
codigo_parroquia	Código externo de parroquia
codigo_canton	Código externo de cantón
codigo_provincia	Código externo de provincia
codigo_region	Código externo de región
codigo_registro	Estado del registro

Perspectiva Remitente:

ATRIBUTO	DESCRIPCIÓN
id_remitente	Clave operacional del remitente
id_cargo	Clave operacional del cargo
id_dependencia	Clave operacional de la dependencia
nombre_remitente	Nombres y apellidos del remitente
identificacion_remitente	Identificación del remitente
detalle_dependencia	Detalle de la dependencia
detalle_cargo	Detalle del cargo
estado_registro	Estado del registro del remitente

Dimensión Persona

ATRIBUTO	DESCRIPCIÓN
id_persona_asignada	Clave operacional del remitente
nombre_persona	Nombres y apellidos del remitente
identificacion_persona	Identificación del remitente
estado_registro	Estado del registro del remitente

Dimensión Área

ATRIBUTO	DESCRIPCIÓN
id_area	Clave operacional del área
detalle_area	Detalle del área
estado_registro	Estado del registro del área

Perspectiva Medio de Entrega

ATRIBUTO	DESCRIPCIÓN
id_medio_entrega	Clave operacional del medio de entrega
detalle_medio_entrega	Detalle del medio de entrega
estado_registro	Estado del registro del medio de entrega

Perspectiva Estado del proceso del trámite

ATRIBUTO	DESCRIPCIÓN
id_estado_tramite	Clave operacional del estado de proceso de trámite
detalle_estado_tramite	Detalle del estado del proceso de trámite
estado_registro	Estado del registro del tipo de trámite

Perspectiva Tipo de trámite

ATRIBUTO	DESCRIPCIÓN
id_tipo_tramite	Clave operacional del tipo de trámite
detalle_tipo_tramite	Detalle del tipo de trámite
estado_registro	Estado del registro del tipo de trámite

Los datos importantes que permiten analizar los indicadores son:

Tiempo:

- id_tiempo
- fecha
- dia
- mes
- año

Localidad:

- id_localidad
- id_ubicacion
- id_paroquia
- id_canton
- id_provincia
- id_zona

Remitente:

- id_remitente
- id_cargo

- id_dependencia

Persona:

- id_persona
- id_persona_asignada

Área:

- id_area
- detalle_area

Medio de entrega:

- id_entrega
- detalle_medio_entrega

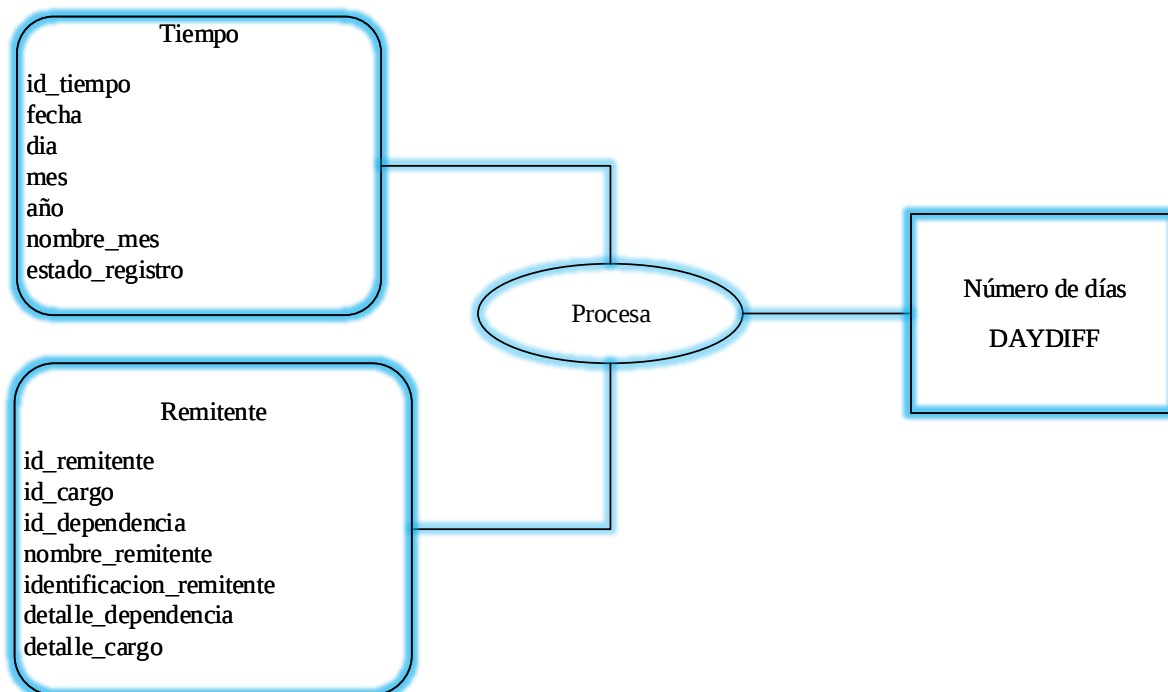
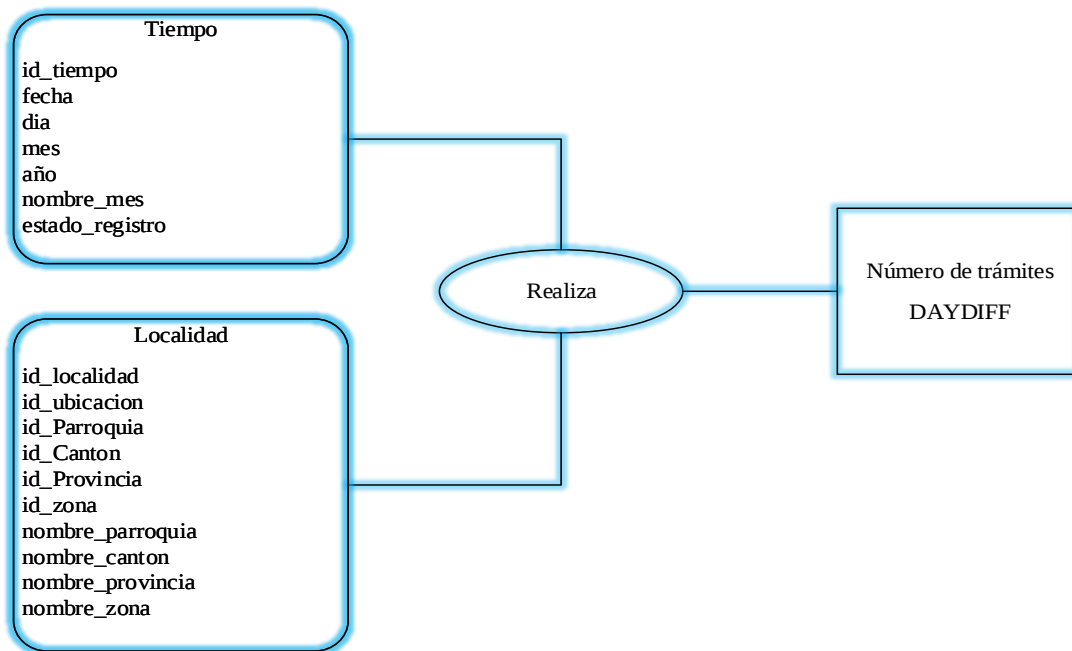
Estado

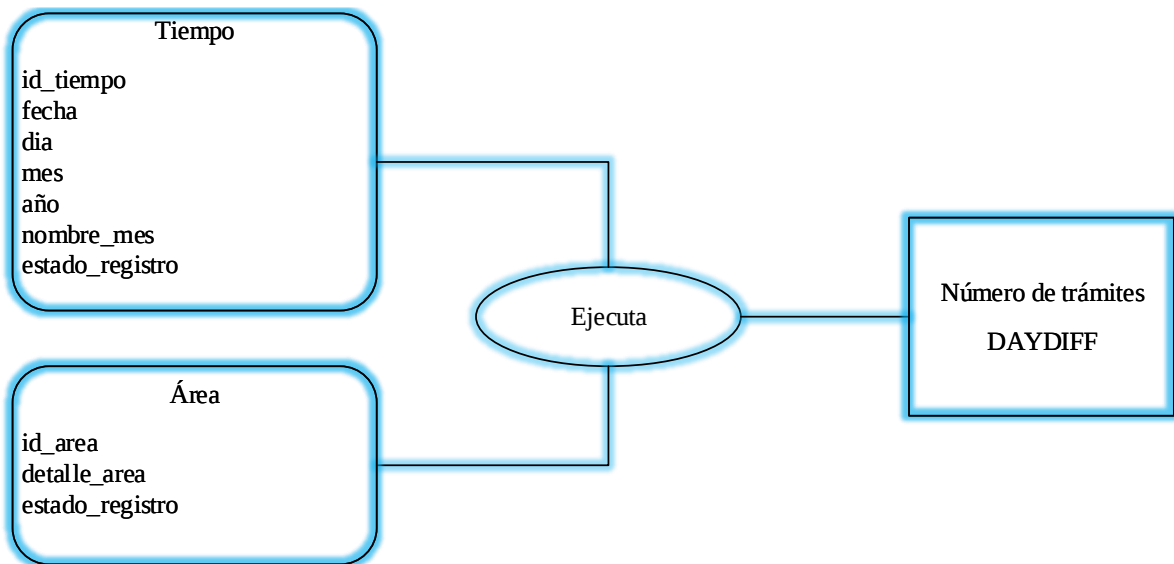
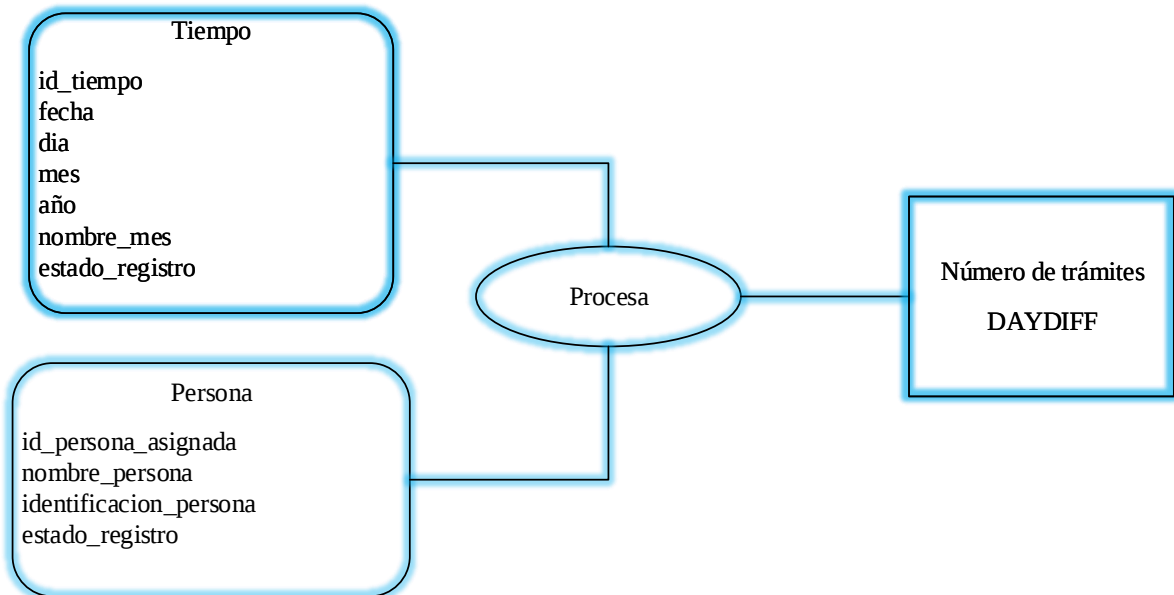
- id_estado
- detalle_estado_tramite

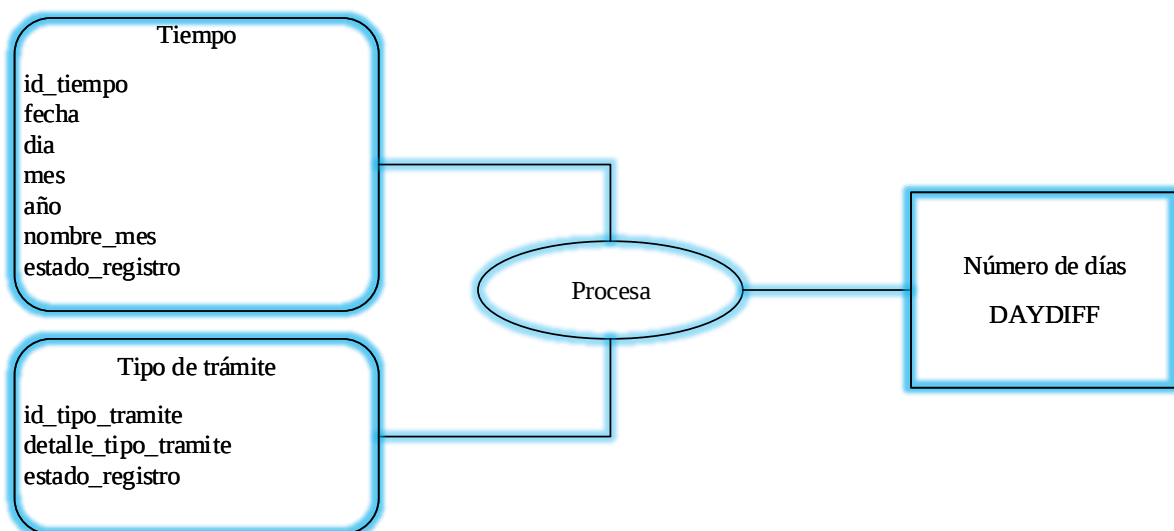
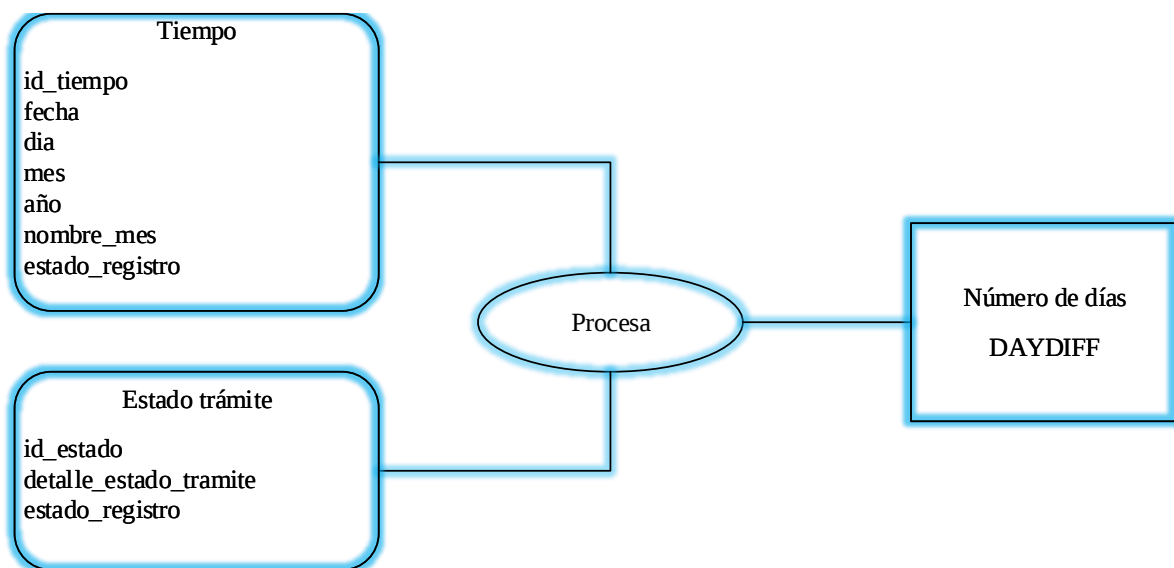
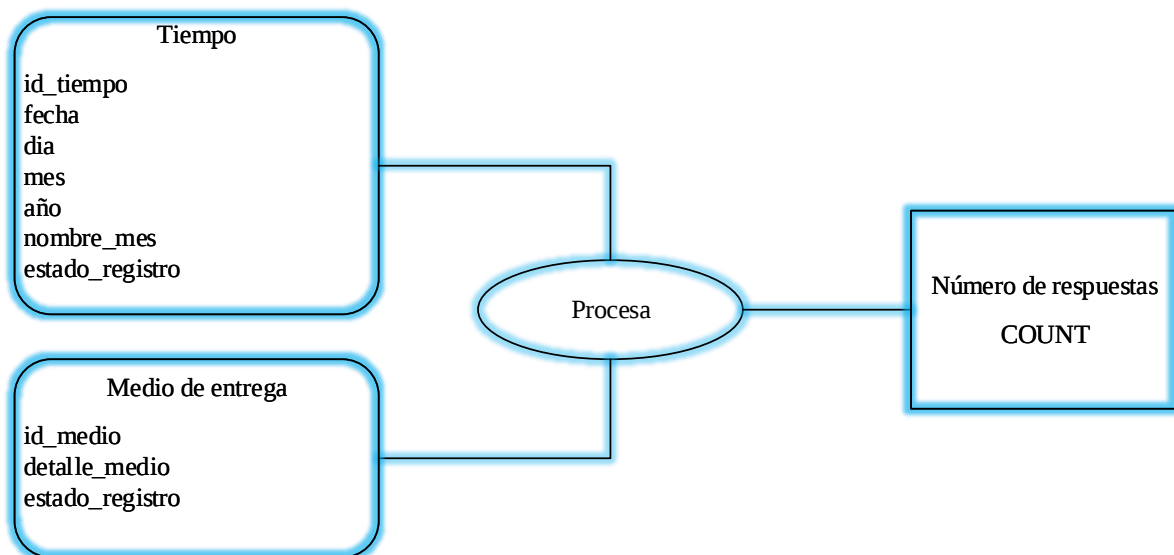
Tipo de trámite:

- id_tipo_tramite
- detalle_tipo_tramite

7.2.4. Modelo Conceptual ampliado







7.3. PASO 3) MODELO LÓGICO DEL DW

Para la construcción del modelo lógico de la estructura del *Data warehouse*, se apoyará en el modelo conceptual ampliado, para el diseño de las tablas de dimensiones y hechos.

7.3.1. Tipo de Modelo Lógico del DW

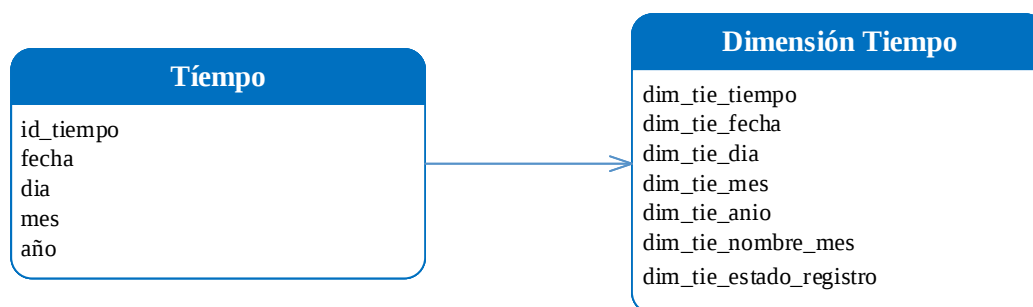
El esquema a utilizar será el de estrella ya se adecua a las necesidades y requerimientos de los usuarios.

7.3.2. Tablas de dimensiones

Las tablas de dimensiones que forman parte del *Data Warehouse* son:

- Dimensión Tiempo

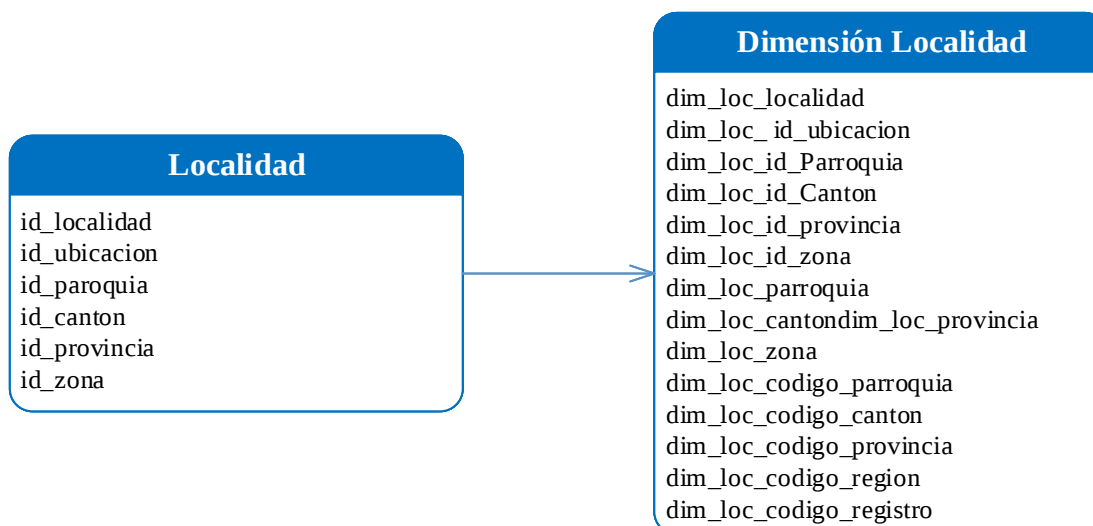
En esta dimensión comprende las fechas: Fecha Documento (trámite), Fecha Recepción (trámite), Fecha Proceso (proceso trámite), Fecha Respuesta (respuesta trámite). Estos datos están agrupados por: años, semestres, trimestres, meses, semanas, días.



ATRIBUTO	DESCRIPCIÓN	TIPO DE DATO
Key: dim_tie_tiempo	Clave subrogada de la dimensión	Integer
dim_tie_fecha	Fecha completa: año, mes, día	Date
dim_tie_dia	Día en la fecha	Integer
dim_tie_mes	Mes en la fecha	Integer
dim_tie_anio	Año en la fecha	Integer
dim_tie_nombre_mes	Nombre del mes	Varchar(30)
dim_tie_estado_registro	Estado del registro	Char(1)

- Dimensión Localidad

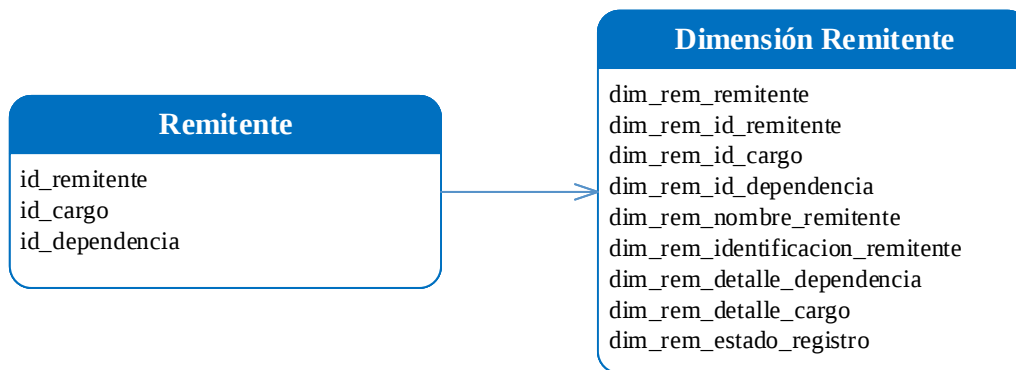
Información descriptiva de la ubicación geográfica, posee datos como las Provincias, Ciudades y Parroquias.



ATRIBUTO	DESCRIPCIÓN	TIPO DE DATO
Key: dim_loc_localidad	Clave subrogada de la dimensión	Integer
Pk: dim_loc_id_ubicacion	Clave operacional de la ubicación geográfica	Integer
Id: dim_loc_id_Parroquia	Identificador del atributo parroquia	Integer
Id: dim_loc_id_Canton	Identificador del atributo cantón	Integer
Id: dim_loc_id_provincia	Identificador del atributo provincia	Integer
Id: dim_loc_id_zona	Identificador del atributo zona	Integer
dim_loc_paroquia	Nombre de la parroquia	Varchar(50)
dim_loc_canton	Nombre del cantón	Varchar(50)
dim_loc_provincia	Nombre de la provincia	Varchar(50)
dim_loc_zona	Nombre de la zona	Varchar(50)
dim_loc_codigo_paroquia	Código externo de parroquia	Varchar(20)
dim_loc_codigo_canton	Código externo de cantón	Varchar(20)
dim_loc_codigo_provincia	Código externo de provincia	Varchar(20)
dim_loc_codigo_region	Código externo de región	Varchar(20)
dim_loc_codigo_registro	Estado del registro	Varchar(50)

- Dimensión Remitente

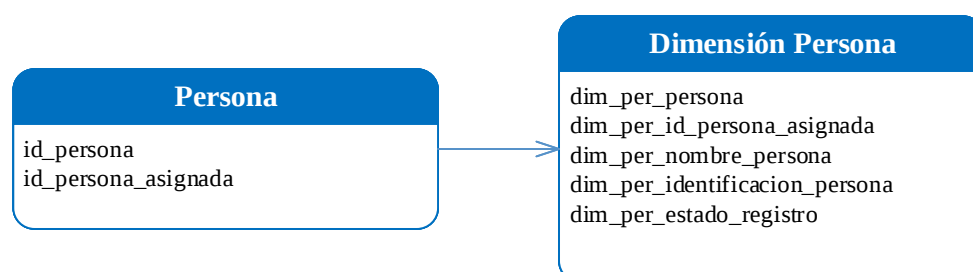
Información descriptiva concerniente a los cargos que un remitente ocupa, datos como: Identificación, Nombres, Cargo, Detalle de la dependencia.



ATRIBUTO	DESCRIPCIÓN	TIPO DE DATO
Key: dim_rem_remitente	Clave subrogada de la dimensión	Integer
Pk: dim_rem_id_remitente	Clave operacional del remitente	Integer
Pk: dim_rem_id_cargo	Clave operacional del cargo	Varchar(150)
Pk: dim_rem_id_dependencia	Clave operacional de la dependencia	Varchar(13)
dim_rem_nombre_remitente	Nombres y apellidos del remitente	Integer
dim_rem_identificacion_remitente	Identificación del remitente	Varchar(50)
dim_rem_detalle_dependencia	Detalle de la dependencia	Integer
dim_rem_detalle_cargo	Detalle del cargo	Varchar(50)
dim_rem_estado_registro	Estado del registro del remitente	Char(1)

- Dimensión Persona Asignada

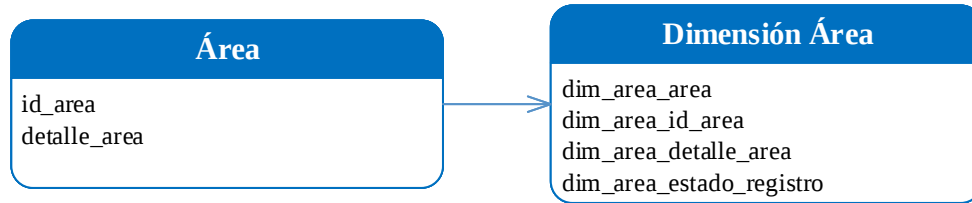
Información descriptiva de la persona a quien se le asigna un trámite, datos como Identificación, Nombre.



ATRIBUTO	DESCRIPCIÓN	TIPO DE DATO
Key: dim_per_persona	Clave subrogada de la dimensión	Integer
Pk: dim_per_id_persona_asignada	Clave operacional del remitente	Integer
dim_per_nombre_persona	Nombres y apellidos del remitente	Varchar(150)
dim_per_identificacion_persona	Identificación del remitente	Varchar(13)
dim_per_estado_registro	Estado del registro del remitente	Char(1)

- Dimensión Área

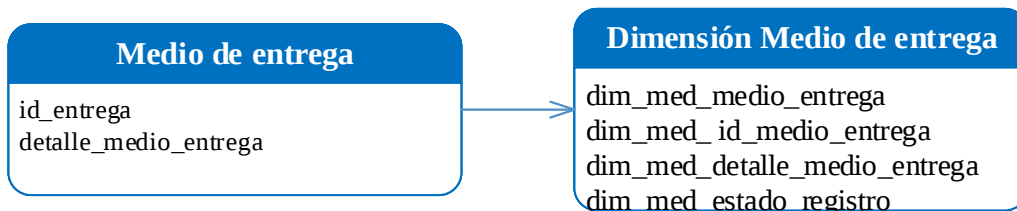
Información descriptiva de la dependencia y el área hacia donde es emitida la respuesta de un trámite, datos como: Detalle del área.



ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Key: dim_area_area	Clave subrogada del área	Integer
Pk: dim_area_id_area	Clave operacional del área	Integer
dim_area_detalle_area	Detalle del área	Varchar(150)
dim_area_estado_registro	Estado del registro del área	Char(1)

- Dimensión Medio Entrega

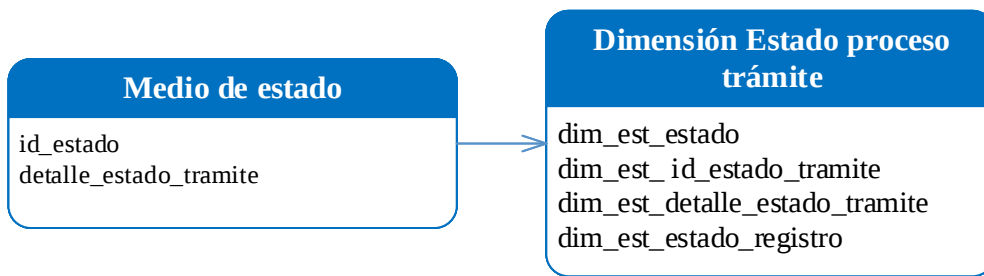
Información descriptiva de los medios por los cuales se entrega la respuesta de un trámite, datos como: Detalle del medio de entrega.



ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Key: dim_med_medio_entrega	Clave subrogada del medio de entrega	Integer
Pk: dim_med_id_medio_entrega	Clave operacional del medio de entrega	Integer
dim_med_detalle_medio_entrega	Detalle del medio de entrega	Varchar(150)
dim_med_estado_registro	Estado del registro del medio de entrega	Varchar(13)

- Dimensión Estado Proceso Trámite (Estado)

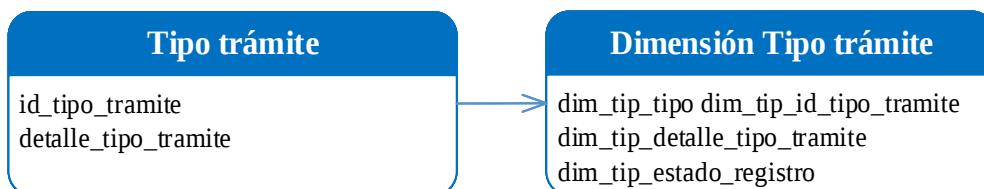
Información descriptiva sobre el estado en el cual se encuentra el trámite, datos como: Detalle del estado del proceso de trámite.



ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Key: dim_est_estado	Clave subrogada del estado de proceso de trámite	Integer
Pk: dim_est_id_estado_tramite	Clave operacional del estado de proceso de trámite	Integer
dim_est_detalle_estado_tramite	Detalle del estado del proceso de trámite	Varchar(150)
dim_est_estado_registro	Estado del registro del tipo de trámite	Char(1)

- Dimensión Tipo Trámite (Tipo)

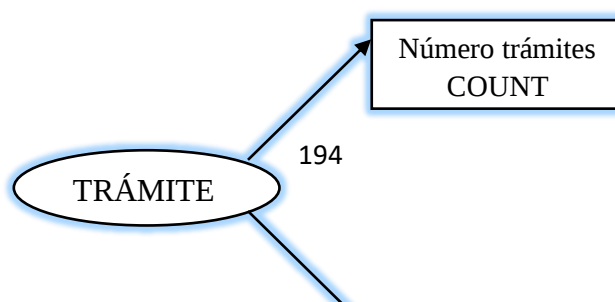
Información del tipo de trámite, datos como: Detalle del tipo de trámite.



ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Key: dim_tip_tipo	Clave subrogada del tipo de trámite	Integer
Pk: dim_tip_id_tipo_tramite	Clave operacional del tipo de trámite	Integer
dim_tip_detalle_tipo_tramite	Detalle del tipo de trámite	Varchar(150)
dim_tip_estado_registro	Estado del registro del tipo de trámite	Char(1)

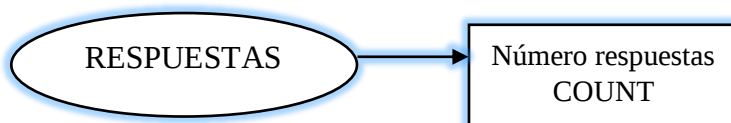
7.3.3. Tablas de hechos

Las tablas de hechos se utilizarán para la toma de decisiones, estas contienen los hechos a través de los que se construirán los indicadores de estudio.



Número de días
DAYDIFF

ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Fk: dim_loc_localidad	Clave operacional del proceso localidad	Integer
Fk: dim_rem_remitente	Clave operacional del proceso remitente	Integer
Fk: dim_tipo_tipo	Clave operacional del proceso tipo	Integer
Fk: dim_per_persona	Clave operacional del proceso de persona	Integer
Fk: dim_est_estado	Clave operacional del proceso estado	Integer
Fk: dim_are_area	Clave operacional del proceso área	Integer
Fk: dim_tie_fecha_documento	Clave operacional del proceso tiempo	Integer
Fk: dim_tie_fecha_proceso	Clave operacional del proceso tiempo	Integer
fact_tra_numero	Número de trámites	Integer
fact_tra_dias	Número de días	Integer
fact_tra_estado_registro	Estado del registro del tipo de trámite	Char(1)



ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Fk: dim_rec_receptor	Clave operacional del proceso receptor	Integer
Fk: dim_med_medioEntrega	Clave operacional del proceso medio entrega	Integer
Fk: dim_tie_fecha_respuesta	Clave operacional del proceso tiempo	Integer
Fk: dim_tie_fecha_entrega	Clave operacional del proceso tiempo	Integer
fact_res_numero	Número de trámites	Integer
fact_tra_estado_registro	Estado del registro de tramite	Char(1)

7.3.4. Uniones

Las uniones entre las tablas de dimensiones y sus tablas de hechos.



7.4. PASO 4) INTEGRACIÓN DE DATOS

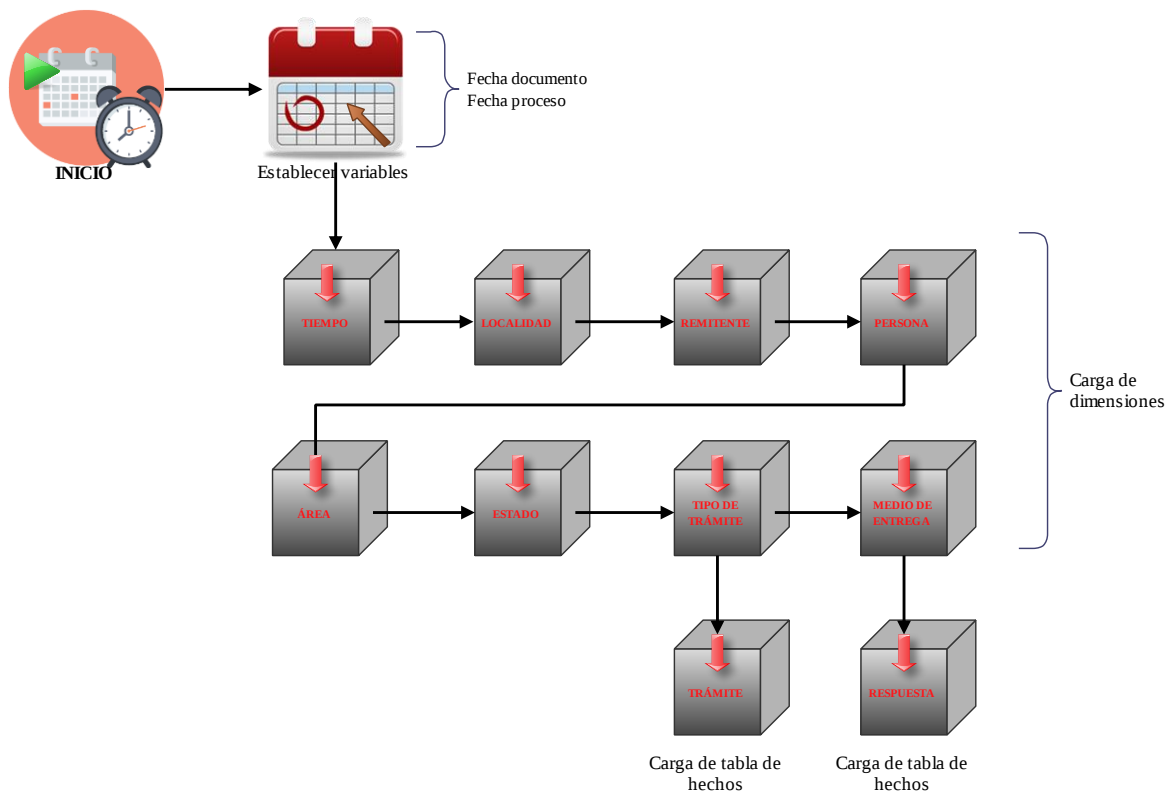
Procedemos a ingresar los datos, para lo cual se usará métodos de limpieza y calidad de datos como ETL; para puntualizar reglas y políticas para actualización.

7.4.1. Carga Inicial

Evitar que el *Data warehouse* sea cargado con valores faltantes o datos no limpios, las tablas dimensionales serán compartidas con varias tablas de hechos, para evitar que algunas restricciones aplicadas sobre una tabla de dimensión en particular para analizar una tabla de hechos, se puedan contraponer con otras restricciones o condiciones de análisis de otras tablas de hechos.

- Carga de datos dimensionales
- Carga de tablas de hechos
- Registrar las acciones llevadas a cabo

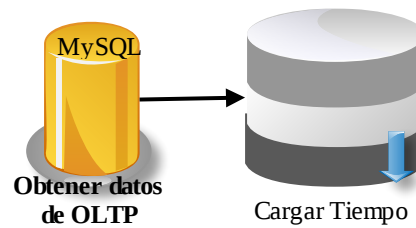
El proceso ETL propuesto para realizar la carga inicial de datos es el siguiente:



Descripción de pasos de la carga inicial:

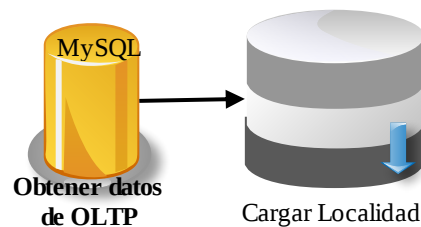
NO	PASOS	DESCRIPCIÓN
1	Inicio	Inicia la ejecución de los pasos en el momento que se indique.
2	Establecer variables	Establece las variables globales que serán utilizadas posteriormente. • Variable fecha_documento: Se obtiene el valor de la fecha que se entrega el documento. • Variable fecha_tramite: Se obtiene el valor de la fecha actual por el proceso y estado a realizar.
3	Carga de Dimensión Tiempo	Ejecuta el contenedor de pasos que cargara la dimensión.
4	Carga de Dimensión Localidad	Ejecuta el contenedor de pasos que cargara la dimensión.
5	Carga de Dimensión Remitente	Ejecuta el contenedor de pasos que cargara la dimensión.
6	Carga de Dimensión Persona	Ejecuta el contenedor de pasos que cargara la dimensión.
7	Carga de Dimensión Área	Ejecuta el contenedor de pasos que cargara la dimensión.
8	Carga de Dimensión Estado	Ejecuta el contenedor de pasos que cargara la dimensión.
9	Carga de Dimensión Tipo de trámite	Ejecuta el contenedor de pasos que cargara la dimensión.
10	Carga de Dimensión Medio de entrega	Ejecuta el contenedor de pasos que cargara la dimensión.
11	Carga de Tabla de Hechos Trámite	Ejecuta el contenedor de pasos que cargara la tabla de hechos.
12	Carga de Tabla de Hechos Respuesta	Ejecuta el contenedor de pasos que cargara la tabla de hechos.

Las tareas que se llevan a cabo por “Carga Dimensión Tiempo”



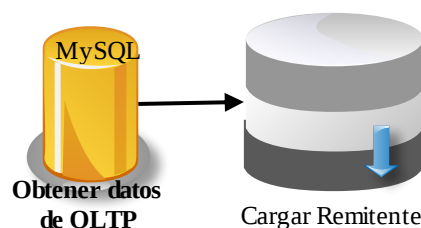
1. Se generan 45000 registros con la (1900-01-01) por defecto
2. Para cada registro se asigna un número secuencial comenzando desde 1
3. Se lleva a cabo las transformaciones 1,2,3
4. Se filtran las fechas para ingresar solo valores hasta un año determinado se ordena las fechas en base a la columna dim_tie_fecha
5. Se cargan los registros en la dimensión

Las tareas que se llevan a cabo por “Carga Dimensión Localidad”



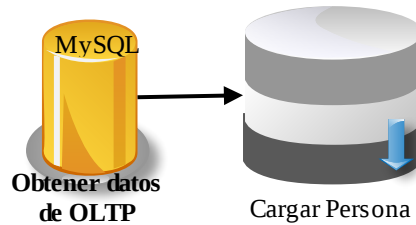
1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión localidad, la extracción se realizó en dos etapas
 - Asignación de códigos y detalles de zonas, regiones, provincias, parroquias
 - Generación de datos mediante script sql
2. Se combinan los registros de la extracción
3. Se llevan a cabo las transformaciones 4,5,6
4. Se asignan la clave subrogada para cada registro en base a la clave operacional:
dim_loc_id_ubicacion

Las tareas que se llevan a cabo por “Carga Dimensión Remitente”



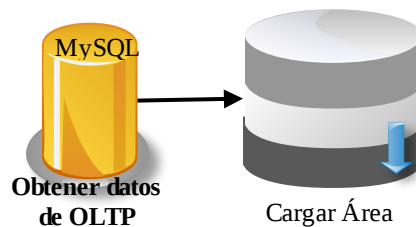
1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional
dim_rem_id_remitente, dim_rem_id_cargo, dim_rem_id_dependencia
3. Se cargan y actualizan los registros en la dimensión

Las tareas que se llevan a cabo por “Carga Dimensión Persona”



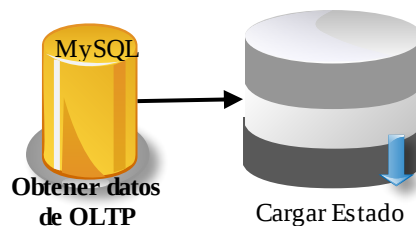
1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional
dim_rem_id_persona_asignada
3. Se cargan y actualizan los registros en la dimensión

Las tareas que se llevan a cabo por “Carga Dimensión Área”



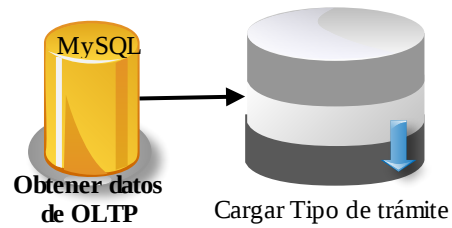
1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional dim_are_id_area
3. Se cargan y actualizan los registros en la dimensión

Las tareas que se llevan a cabo por “Carga Dimensión Estado”



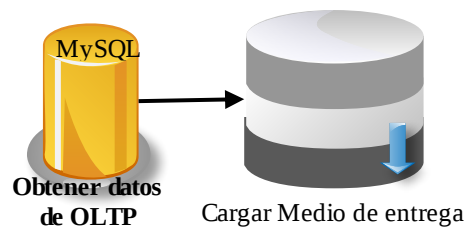
1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional
dim_est_id_estado_tramite, dim_est_detalle_estado_tramite, dim_est_estado_registro
3. Se cargan y actualizan los registros en la dimensión

Las tareas que se llevan a cabo por “Carga Dimensión Tipo de trámite”



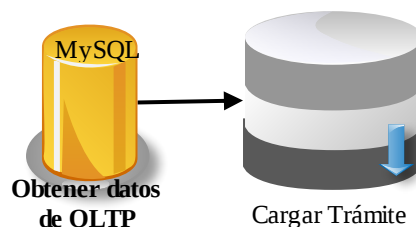
1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional
dim_tipo_id_tipo_tramite, dim_est_detalle_estado_tramite
3. Se cargan y actualizan los registros en la dimensión

Las tareas que se llevan a cabo por “Carga Dimensión Medio de entrega”



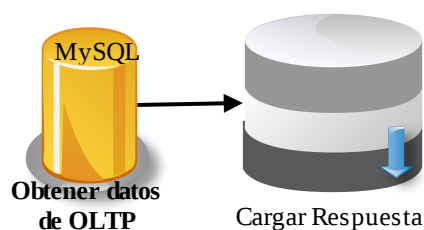
1. Se extraen los datos conforme a lo establecido en los estándares para la dimensión
2. Se asigna la clave subrogada para cada registro en base a la clave operacional
dim_med_id_medio_entrega
3. Se cargan y actualizan los registros en la dimensión

Las tareas que se llevan a cabo por “Carga de la tabla de hechos Trámite”



1. Extracción de datos de las tablas origen
2. Transformaciones y cálculos
3. Obtención de claves dimensionales
4. Validación de claves

Las tareas que se llevan a cabo por “Carga de la tabla de hechos Respuesta”



1. Extracción de datos de las tablas origen
2. Transformaciones y cálculos
3. Obtención de claves dimensionales
4. Validación de claves

7.4.2. Actualización

- La información se refrescará todos los días a las doce de la noche.
- Los datos de las tablas de dimensiones serán cargados en su totalidad simultáneamente.

Para el proceso ETL tenemos las siguientes:

- Inicio: iniciará la ejecución de los pasos todos los días a las doce de la noche.
- Para el cálculo de los números de trámites se basa en el número de registros de trámite.
- Para el número de días se calcula en base a las fechas de recepción de trámite y la fecha de finalización del trámite, se restan y se transforma el resultado a días
- Es importante mantener actualizadas las dimensiones, a través de la carga o actualización de registros. La actualización en las tablas de hechos es poco probable, pero permite mantener actualizados los registros correspondientes a un rango de fechas.
- **Políticas de cambio de dimensiones:** manera como se administran los cambios para cada una de las dimensiones

TIPO CAMBIO	DESCRIPCIÓN
Tipo 0	Sin Cambio
Tipo 1	Sobre escritura
Tipo 2	Creación de nuevo registro en la dimensión

- **Conexiones de base de datos:** Se detalle el nombre de la conexión y cada uno de sus componentes (servidor, DBMS, BD, usuario, password).
- **Transformación de datos:** Identificación de las columnas sobre las cuales se realizan los cambios y cálculos.

- **Flujo de datos de la fuente al destino.** Diagramas de flujo de datos sobre el cada una de las etapas que el proceso debe recorrer (nombre del paso, objeto provisto por la herramienta ETL, descripción).

8. Creación de Cubos Multidimensionales

El análisis OLAP de la solución es realizado con la herramienta *SpagoBI*. Por lo tanto, el cubo refleja el diseño del modelo dimensional del *Data Mart*.

8.1. Estructura del cubo

Para la elaboración del *Data Mart* se establecen los siguientes cubos:

- Trámites Días Tiempo
- Trámite Estado
- Trámite Localidad
- Trámite Remitente
- Trámite Tiempo
- Trámite Tipo

8.1.1. Estructura Trámites Días Tiempo

- Nombre cubo: Estructura Trámites Días Tiempo
- Nombre de dimensiones: Dim tiempo
- Tabla de hechos: Fact tramite

```
<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Numero Dias" column="fact_tra_dias" aggregator="avg"
formatString="#,###"/>
</Cube>
</SCHEMA>
```

```
</Cube>
</SCHEMA>
```

8.1.2. Estructura Trámites Estado

- Nombre cubo: Trámites Estado
- Nombre de dimensiones: Dim estado, Dim tiempo
- Tabla de hechos: Fact tramite

```
<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
  <Dimension name="Dim estado">
    <Hierarchy hasAll="true"
      allMemberName="Estado"
      primaryKey="dim_est_estado"
      name="Estado"
    >
      <Table name="dim_estado"/>
      <Level name="Estado tramite" column="dim_est_detalle_estado_tramite"
        uniqueMembers="false"
      />
    </Hierarchy>
  </Dimension>
  <Dimension name="Dim tiempo">
    <Hierarchy hasAll="true"
      allMemberName="Tiempo"
      primaryKey="dim_tip_tiempo"
      name="Tiempo"
    >
      <Table name="dim_tiempo"/>
      <Level name="Año" column="dim_tip_anio"
        uniqueMembers="false"
      />
      <Level name="Mes" column="dim_tip_nombre_mes"
        uniqueMembers="false"
      />
      <Level name="Fecha" column="dim_tip_fecha"
        uniqueMembers="false"
      />
    </Hierarchy>
  </Dimension>
  <Cube name="Fact tramite">
    <Table name="fact_tramite"/>
    <DimensionUsage name="Dim estado" source="Dim estado"
      foreignKey="dim_est_estado"/>
    <DimensionUsage name="Dim tiempo" source="Dim tiempo"
      foreignKey="dim_tip_tiempo"/>
    <Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
      formatString="#,###"/>
  </Cube>
</SCHEMA>
```

8.1.3. Estructura Trámites Localidad

- Nombre cubo: Trámites Localidad

- Nombre de dimensiones: Dim localidad, Dim tiempo
- Tabla de hechos: Fact tramite

```
<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim localidad">
<Hierarchy hasAll="true"
allMemberName="Localidad"
primaryKey="dim_loc_localidad"
name="Localidad"
>
<Table name="dim_localidad"/>
<Level name="Zona" column="dim_loc_zona"
uniqueMembers="false"
/>
<Level name="Provincia" column="dim_loc_provincia"
uniqueMembers="false"
/>
<Level name="Canton" column="dim_loc_canton"
uniqueMembers="false"
/>
<Level name="Parroquia" column="dim_loc_parroquia"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim localidad" source="Dim localidad"
foreignKey="dim_loc_localidad"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>
```

•

8.1.4. Estructura Trámites Remitente

- Nombre cubo: Trámites Remitente
- Nombre de dimensiones: Dim remitente, Dim Tiempo
- Tabla de hechos: Fact tramite

```
<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim remitente">
<Hierarchy hasAll="true"
allMemberName="Remitente"
primaryKey="dim_rem_remitente"
name="Remitente"
>
<Table name="dim_remitente"/>
<Level name="Nombre Remitente" column="dim_rem_nombre_remitente"
uniqueMembers="false"
/>
<Level name="Identificacion" column="dim_rem_identificacion_remitente"
uniqueMembers="false"
/>
<Level name="Cargo" column="dim_rem_detalle_cargo"
uniqueMembers="false"
/>
<Level name="Dependencia" column="dim_rem_detalle_dependencia"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim remitente" source="Dim remitente"
foreignKey="dim_rem_remitente"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Número Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>
```

8.1.5. Estructura Trámites Tiempo

- Nombre cubo: Trámites Tiempo
- Nombre de dimensiones: Dim tiempo
- Tabla de hechos: Fact tramite

```
<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim tiempo">
<Hierarchy hasAll="true"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>
```

8.1.6. Estructura Trámites Tipo

- Nombre cubo: Trámites Tipo
- Nombre de dimensiones: Dim Tipo, Dim Tiempo
- Tabla de hechos: Fact tramite

```
<?xml version="1.0" ?>
<SCHEMA name="dwinpc negocio">
<Dimension name="Dim tiempo">
<Hierarchy hasAll="True"
allMemberName="Tiempo"
primaryKey="dim_tip_tiempo"
name="Tiempo"
>
<Table name="dim_tiempo"/>
<Level name="Año" column="dim_tip_anio"
uniqueMembers="false"
/>
<Level name="Mes" column="dim_tip_nombre_mes"
uniqueMembers="false"
/>
```

```

<Level name="Fecha" column="dim_tip_fecha"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Dimension name="Dim tipo">
<Hierarchy hasAll="true"
allMemberName="Tipo"
primaryKey="dim_tip_tipo"
name="Tipo"
>
<Table name="dim_tipo"/>
<Level name="Tipo Tramite" column="dim_tip_detalle_tipo_tramite"
uniqueMembers="false"
/>
</Hierarchy>
</Dimension>
<Cube name="Fact tramite">
<Table name="fact_tramite"/>
<DimensionUsage name="Dim tiempo" source="Dim tiempo"
foreignKey="dim_tip_tiempo"/>
<DimensionUsage name="Dim tipo" source="Dim tipo"
foreignKey="dim_tip_tipo"/>
<Measure name="Numero Tramites" column="fact_tra_numero" aggregator="sum"
formatString="#,###"/>
</Cube>
</SCHEMA>

```

Apéndice B

SAS RAPID DATA WAREHOUSE METHODOLOGY

Introducción.....	211
1. Evaluación.....	211
1.1. Objetivos del <i>Data Mart</i>	211
1.2. Situación actual del INPC R3	211
1.2.1. Generalidades.....	211
1.2.2. Mapa orgánico funcional de INPC R3.....	213
1.2.3. Identificación de problemas.....	213
1.3. Desarrollo de cronograma y presupuesto del proyecto	214
2. Requerimientos	214
2.1. Requerimientos del propietario	214
2.2. Requerimientos del arquitecto	214
2.2.1. Fuente de datos	215
2.2.2. <i>Data Mart</i>	215
2.2.3. Acceso y Uso.....	215
2.2.4. Capas del <i>Data Mart</i>	215
2.3. Requerimientos del desarrollador	216
2.3.1. Fuente de datos	216
2.3.2. <i>Data Mart</i>	216
2.3.3. Acceso y Uso.....	216
2.4. Requerimientos del usuario final.....	217
2.4.1. Requerimientos de consulta.....	217
2.4.2. Requerimientos de reporte	217
3. Análisis.....	217
3.1. Dimensiones.....	217
3.2. Tablas de hecho	218
3.2.1. Proceso de medición de trámites.....	218
3.2.2. Medidas	219
3.2.3. Dimensiones.....	219
3.2.4. Modelo Lógico	219
3.2.5. Proceso de medición de respuestas de trámites	219
3.2.6. Medidas	220
3.2.7. Dimensiones.....	220

3.2.8.	Modelo Lógico	220
3.3.	Estrategia de carga y refrescamiento de datos.....	220
3.3.1.	Estrategia de carga de datos	220
3.3.2.	Dimensiones.....	222
3.3.3.	Hechos.....	222
3.3.4.	Estrategia de Refrescamiento de Datos	222
4.	Diseño	223
4.1.	Dimensiones.....	223
4.1.1.	Dimensiones físico de la dimensión dim_tiempo	223
4.1.2.	Dimensiones físico de la dimensión dim_localidad.....	223
4.1.3.	Dimensiones físico de la dimensión dim_remitente	224
4.1.4.	Dimensiones físico de la dimensión dim_persona.....	224
4.1.5.	Dimensiones físico de la dimensión dim_area.....	224
4.1.6.	Dimensiones físico de la dimensión dim_medio_entrega	224
4.1.7.	Dimensiones físico de la dimensión dim_estado	225
4.1.8.	Dimensiones físico de la dimensión dim_tipo.....	225
4.2.	Hechos.....	225
4.2.1.	Modelo físico de la tabla de hechos fact_tramite.....	225
4.2.2.	Modelo físico de la tabla de hechos fact_respuesta_tramite	226
4.3.	Diseño de la arquitectura de la aplicación.....	227
4.3.1.	Fuente de datos	228
4.3.2.	Data Mart.....	228
4.3.3.	Acceso y uso	228
4.3.4.	Componente de visualización.....	228
4.4.	Definición de los niveles de seguridad	229
5.	Implementación.....	229
5.1.	Implementación de la estrategia de carga	230
5.2.	Implementación de la estrategia de refrescamiento	230

Introducción

El almacenamiento de datos es un tema técnico y metodológico que determinan el éxito o el fracaso de los sistemas, por lo que SAS proporciona un almacén de datos que abarca varios ámbitos como son: Gestión de almacenes, organización y explotación, conta de una arquitectura de motor con paso para MySql ya su aporte facilita el desarrollo de sistemas transaccionales legados racionales y otros; su base esta cimentada en la experiencia adquirida en el desarrollo de las diferentes fases lógicas e hitos para verificar que las soluciones tengan bases sólidas.

1. Evaluación

1.1. Objetivos del *Data Mart*

El objetivo que persigue la implementación de un *Data Mart* para el INPC R3, es la construcción de un *Data Warehouse* que apoye la toma de decisiones en el proceso de trámites generado por la institución

1.2. Situación actual del INPC R3

1.2.1. Generalidades

1.2.1.1. Visión

En 2017 el Instituto Nacional de Patrimonio Cultural se consolidará como un centro generador de conocimiento sobre los elementos que constituyen el sustento de nuestra identidad nacional multiétnica y pluricultural. Se fundamentará en la investigación formal, en la conservación y en la difusión multifacética del patrimonio cultural y de la memoria histórica del Ecuador. Este caudal será difundido para concienciar a los actores involucrados y a la ciudadanía sobre la importancia de las raíces y de la identidad. Así, mediante el uso y desarrollo de teorías, métodos y técnicas tanto inter como transdisciplinarias, el INPC R3 facilitará el acceso, el uso social y la preservación de todo el acervo cultural de la Nación.

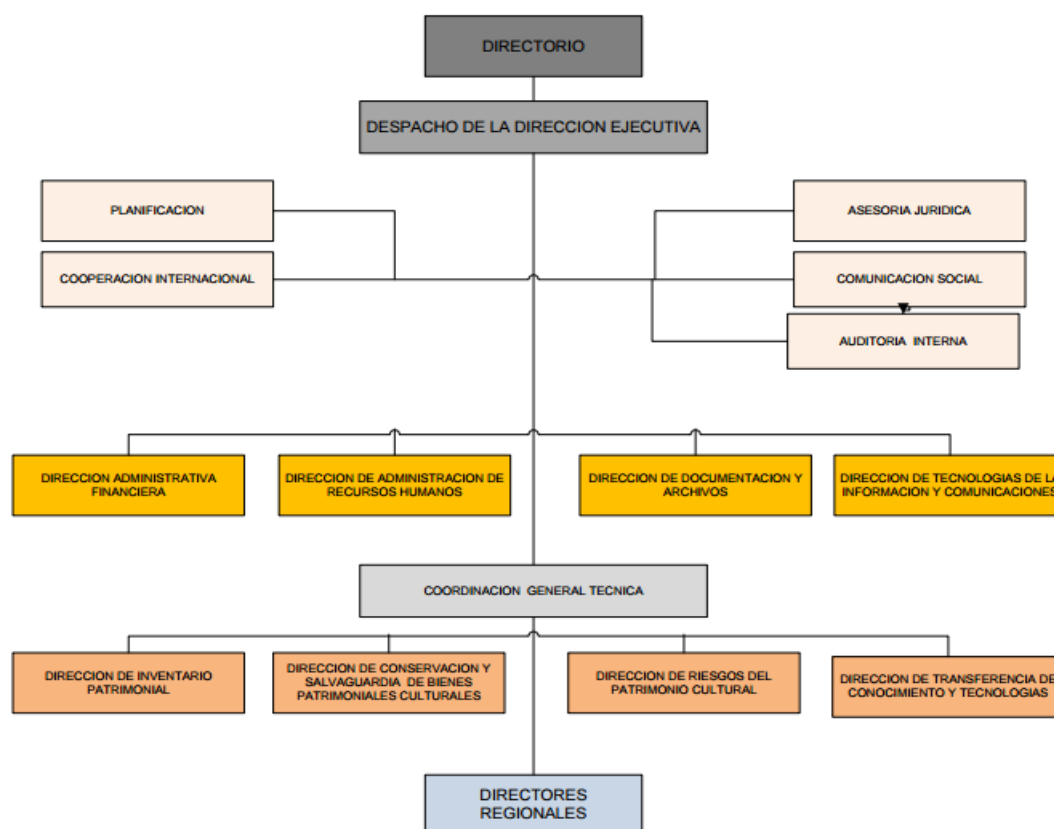
1.2.1.2. Misión

El Instituto Nacional de Patrimonio Cultural es una entidad del sector público con ámbito nacional, encargada de aplicar las políticas sectoriales, así como de investigar, normar, regular y asesorar para la preservación, conservación, apropiación, promoción y uso adecuado del patrimonio cultural material e inmaterial.

1.2.1.3. Objetivos

- Consolidar al Instituto como un Centro de Investigación-Desarrollo y Sistematización del Conocimiento Científico-Técnico del patrimonio cultural, mediante el estudio, análisis y aplicación de teorías, metodologías y técnicas para: catalogar, documentar, conservar, proteger y potenciar el patrimonio material o tangible e inmaterial o intangible, con la finalidad de difundir hacia los actores involucrados las mejores opciones de conservación del patrimonio.
- Desarrollar, sugerir y vigilar la implementación y el cumplimiento de normas, políticas y procedimientos, orientados para la conservación preventiva del patrimonio cultural, a fin de regular y controlar el adecuado manejo del patrimonio por parte de los actores sociales, esto incluye: prevención de riesgo (identificación, conservación y alerta), uso (vulnerabilidad patrimonial) y seguridad (ámbito legal- jurídico).
- Diseñar e implementar sistemas de información para la gestión de los bienes patrimoniales y culturales: El INPC R3 tiene como fin la recopilación de la documentación producida en la institución a nivel nacional, tanto histórica como actual. Toda la información que genere esta base de datos, servirá para la GESTIÓN DEL PATRIMONIO sobre todo para los gobiernos autónomos descentralizados GADs, cuyo mandato constitucional ahora es total. Por otro lado, esta herramienta permitirá conocer nuestros bienes culturales y patrimoniales a través de la WEB y estará al servicio de los estudiosos y de la sociedad en general.
- Sistematizar y comunicar el conocimiento patrimonial material o tangible e inmaterial o intangible mediante la elaboración de publicaciones científicas, las cuales deberán difundirse a través de medios y canales adecuados, así como en el Ministerio de Cultura, mismos que pueden ser conferencias, ponencias y foros públicos, documentos impresos y/o virtuales, así como la realización de cursos propios de sus áreas de especialidad que propendan a la formación de técnicos y personal especializado en materia de patrimonio y seguridad patrimonial.

1.2.2. Mapa orgánico funcional de INPC R3



1.2.3. Identificación de problemas

La siguiente tabla detalla cada una de las necesidades de la institución para con la información requerida de manera frecuente.

PROBLEMA	IMPACTO	PRIORIDAD
No se cuenta con una estrategia de consolidación de la información con respecto a los trámites procesados dentro de la institución, todos estos datos son unificados de forma manual en hojas de cálculo y con muy poca periodicidad	ALTO	ALTO
No posee una herramienta formal, la cual apoye a la toma de decisiones	ALTO	ALTO
No existe seguridad para el acceso a la información	MEDIO	BAJO
Los informes no cuentan con gráficas debido a la complejidad de elaboración.	MEDIO	BAJO
No está disponible la información histórica sobre los trámites que fueron llevado a cabo dentro de la Institución	ALTA	ALTA
Existe redundancia en la información dentro de las hojas de cálculo	MEDIO	MEDIO
Posee una base de datos transaccional con diseño normalizado	BAJO	MEDIO

1.3. Desarrollo de cronograma y presupuesto del proyecto

FASE	TIEMPO ESTIMADO
Planificación	2
Requerimientos	2
Análisis	2
Diseño	2
Implementación	5
Pruebas	3
Duración Total	16

2. Requerimientos

2.1. Requerimientos del propietario

El INPC R3 requiere mantener la información respecto a los trámites de forma transparente y actualizada, ya que, por el momento esto se lo realiza de forma manual, generando así retrasos y poca confiabilidad.

Generar de manera frecuente reportes relacionados con el proceso de trámites

Tener a mano una herramienta que facilite la toma de decisiones, las cuales se base a indicadores y metas planteadas por la institución

2.2. Requerimientos del arquitecto

Los requerimientos del arquitecto están directamente relacionados con la estructura del *Data Warehouse*

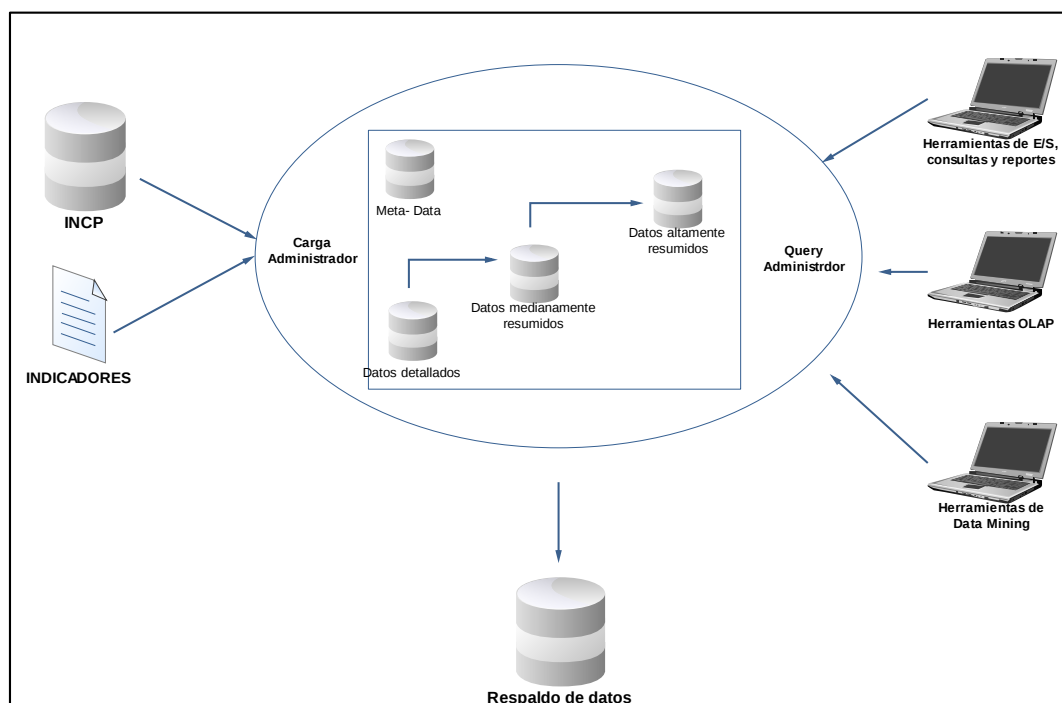


2.2.1. Fuente de datos

Representa los orígenes de los datos, una base de datos transaccional llamada inpc, a continuación, se presenta el modelo físico de la misma.

2.2.2. Data Mart

Es el encargado de realizar el proceso ETL, manejo de datos, listado de consultas y diseño de reportes



2.2.3. Acceso y Uso

El análisis del cubo multidimensional tiene lugar en este proceso, la elaboración de reportes, exploración del *Data Mart* y la obtención de gráficos; todo esto de acuerdo a los perfiles de usuario tomados en cuenta para la ejecución del proyecto.

La herramienta a emplearse es *SpagoBI studio*, que nos permite contar con un módulo de autenticación de usuarios, el cual controlará el acceso a los reportes generados por cada perfil.

2.2.4. Capas del Data Mart

Las capas requeridas para el funcionamiento del *Data Mart* se detallan a continuación:

- Administración de Datos

Encadado de la extracción, transformación y carga de datos desde las fuentes de origen de datos, ubicándolos en el *Data Mart*.

- Administración de Metadatos

Encargado de la definición de los datos que serán almacenados en el *Data Mart*

- Transporte

Se encarga del transporte de datos entre los distintos bloques de la arquitectura del *Data Mart*

- Infraestructura

La infraestructura considerada para esto es cliente – servidor

- Herramientas

Las herramientas que se seleccionan para el diseño, desarrollo y despliegue del *Data Mart* deben considerar los siguientes puntos.

- Definición y diseño de datos
- Administración de base de datos
- ETL
- Generación de cubos
- Reportería
- Acceso para el usuario final

2.3. Requerimientos del desarrollador

Los cuatro principales tipos de requerimientos a tener en cuenta en el desarrollo de este *Data Mart* se describen a continuación:

2.3.1. Fuente de datos

Base de datos relacional MySQL; *Power Designer* para el diseño del *Data Mart* y definición de metadatos.

2.3.2. Data Mart

La construcción del *Data Mart* se lo realizará en MySQL; para la extracción de los datos y la generación de cubos multidimensionales se usará *SpagoBI*

2.3.3. Acceso y Uso

Para la generación de reportes y construcción de gráficos la herramienta *Power BI* será la utilizada.

Para la construcción de la interfaz de usuario se usará

2.4. Requerimientos del usuario final

Los dos grupos de requerimientos que deben ser tratados para cubrir las necesidades de los usuarios son los siguientes:

2.4.1. Requerimientos de consulta

INDICADOR	PREGUNTA
Tramites	Número de trámites receptados en un período de tiempo. Número de trámites por remitente y dependencia dentro de un periodo de tiempo. Identificar número de trámites por estados dentro de un periodo de tiempo. Listar los trámites por prioridad (tipos) en un periodo de tiempo Conocer las personas asignadas durante el proceso del trámite dentro de un período de tiempo. Número de trámites por procedencia geográfica en un periodo determinado de tiempo.
Respuestas	Listar las respuestas de los trámites de un periodo de tiempo su medio de medio de entrega
Número de días	Conocer el número de días transcurridos hasta la finalización del trámite

2.4.2. Requerimientos de reporte

- Acceso autorizado a los reportes en base a permisos
- Interfaz de acceso previa autenticación de usuarios
- Visualización de reportes y gráficos

3. Análisis

Para satisfacer las necesidades del usuario antes descritas se han establecido dos tablas de hechos:

- Tabla de hechos de procesos de trámites
- Tabla de hechos de procesos de las respuestas de trámites

3.1. Dimensiones

- Dimensión Tiempo - dim_tiempo

En esta dimensión comprende las fechas: Fecha Documento (trámite), Fecha Recepción (trámite), Fecha Proceso (proceso trámite), Fecha Respuesta (respuesta trámite). Estos datos están agrupados por: años, semestres, trimestres, meses, semanas, días.

- Dimensión Localidad - dim_localidad

Información descriptiva de la ubicación geográfica, posee datos como las Provincias, Ciudades y Parroquias.

- Dimensión Remitente - dim_remitente

Información descriptiva concerniente a los cargos que un remitente ocupa, datos como: Identificación, Nombres, Cargo, Detalle de la dependencia

- Dimensión Persona Asignada - dim_persona_asignada

Información descriptiva de la persona a quien se le asigna un trámite, datos como Identificación, Nombre

- Dimensión Área - dim_area

Información descriptiva de la dependencia y el área hacia donde es emitida la respuesta de un trámite, datos como: Detalle del área

- Dimensión Receptor - dim_receptor

Información descriptiva del receptor de la respuesta que genera un trámite, datos como: Identificación, Nombres.

- Dimensión Medio Entrega - dim_medio_entrega

Información descriptiva de los medios por los cuales se entrega la respuesta de un trámite, datos como: Detalle del medio de entrega

- Dimensión Tipo Trámite (Tipo) - dim_tipo_tramite

Información del tipo de trámite, datos como: Detalle del tipo de trámite.

- Dimensión Estado Proceso Trámite (Estado) - dim_estado_tramite

Información descriptiva sobre el estado en el cual se encuentra el trámite, datos como: Detalle del estado del proceso de trámite

3.2. Tablas de hecho

3.2.1. Proceso de medición de trámites

Los parámetros del proceso a seguir por parte de la INPC son:

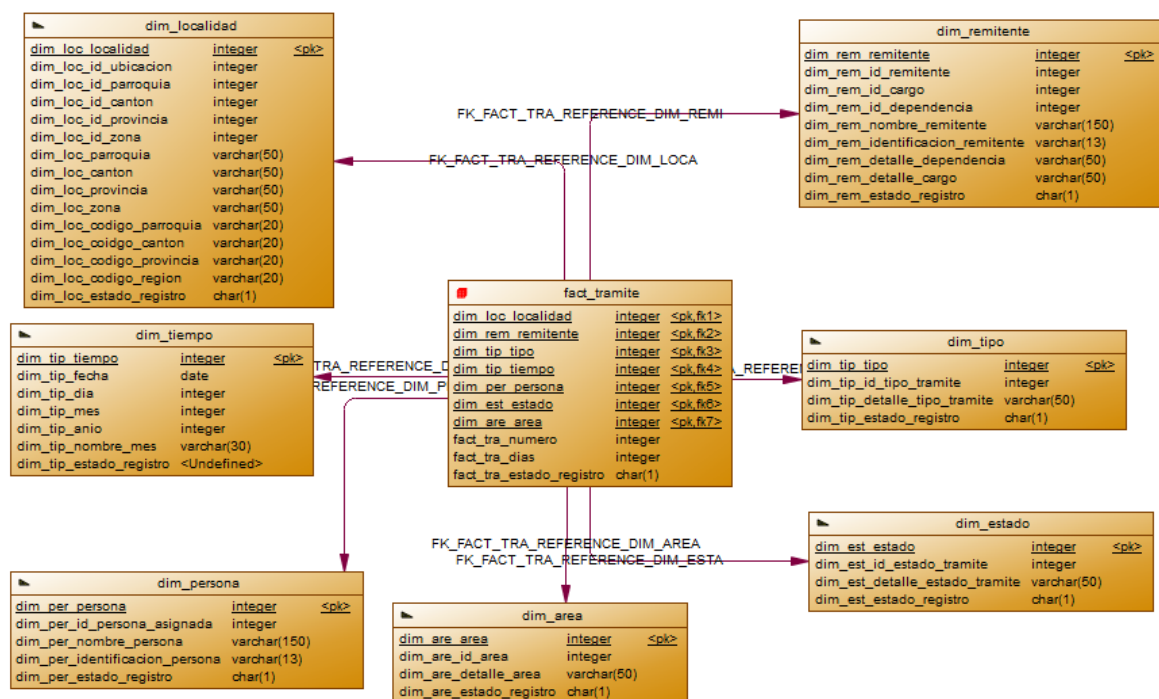
3.2.2. Medidas

- Número de trámites por procesos

3.2.3. Dimensiones

- dim_tiempo
- dim_localidad
- dim_remitente
- dim_persona
- dim_area
- dim_estado
- dim_tipo

3.2.4. Modelo Lógico



3.2.5. Proceso de medición de respuestas de trámites

Los parámetros del proceso a seguir por parte de la INPC R3 son:

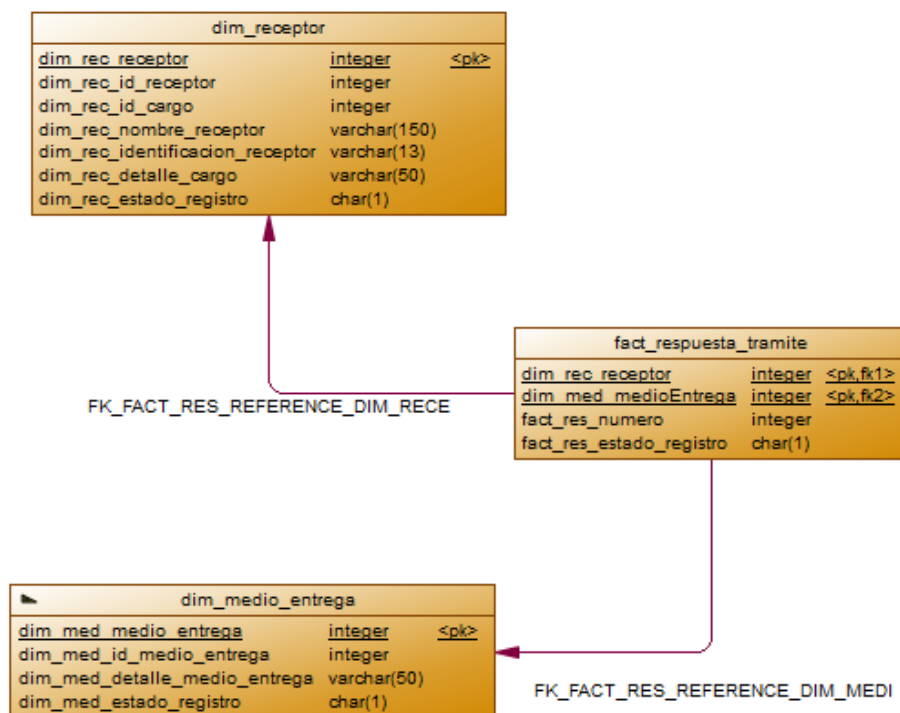
3.2.6. Medidas

- Número de respuestas de trámites

3.2.7. Dimensiones

- dim_receptor
- dim_medio_entrega

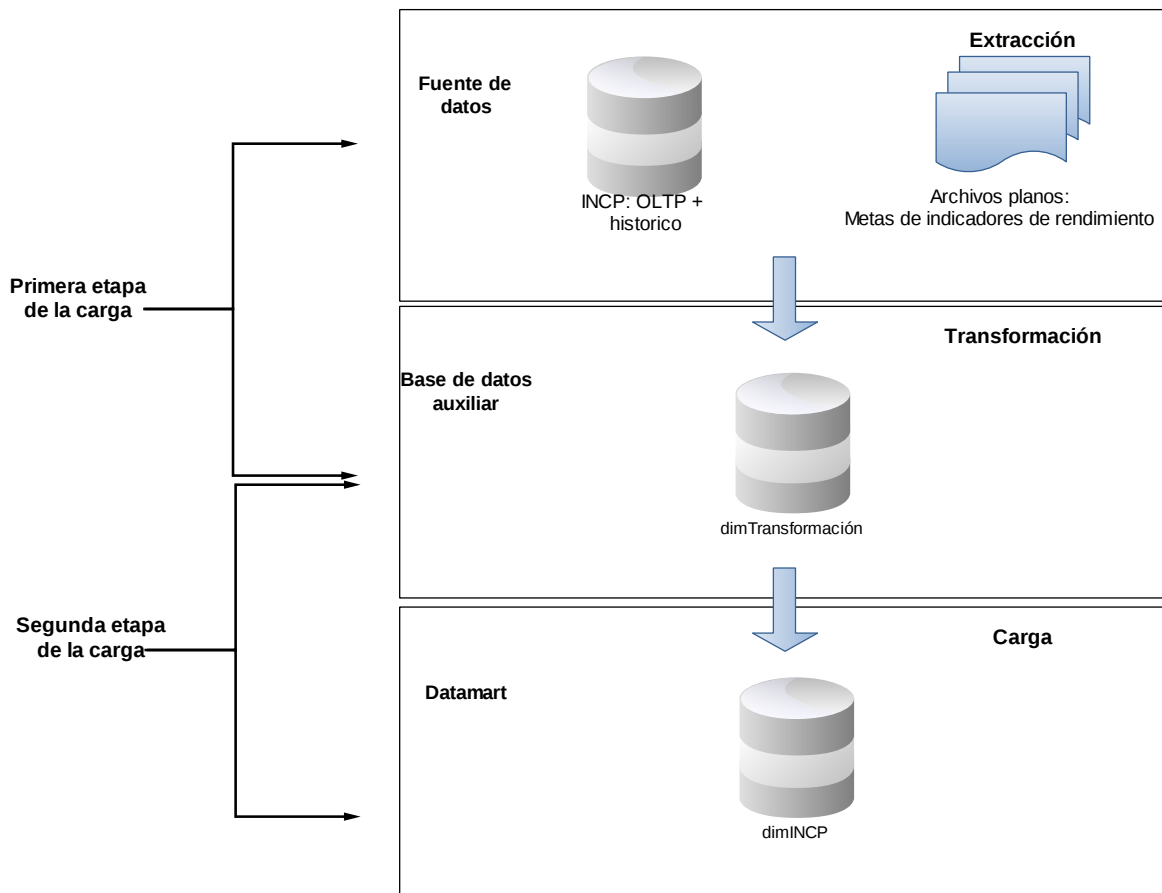
3.2.8. Modelo Lógico



3.3. Estrategia de carga y refrescamiento de datos

3.3.1. Estrategia de carga de datos

Partiendo del análisis de las fuentes realizadas en las fases de planificación, se ha visto la necesidad de la creación de un modelo de datos auxiliar, el cual apoyará en el proceso de carga de información previa la extracción, transformación y carga de datos hacia la base de datos del datamart.



La carga se desarrollará en dos etapas:

- Primera etapa: Consta de la carga y transformación de los datos transaccionales e históricos que provienen de la fuente de datos INPC R3 y las metas de los indicadores almacenadas en las hojas de cálculo hacia una base de datos intermedia denominada dimtransformación, como se muestra a continuación:
 - AUX_TIEMPO: Se tomará la información de la tabla tiempo, concernientes a los periodos de tiempo a utilizar.
 - AUX_LOCALIDAD: Se tomará de la tabla localidad, concerniente al lugar de procedencia de los trámites.
 - AUX_REMITENTE: Se tomará de la tabla remitente, concerniente a la información de la persona que desea realizar el trámite.
 - AUX_PERSONA: Se tomará de la tabla persona, concerniente a los actores directos involucrados.
 - AUX AREA: Se tomará de la tabla área, concerniente a las diversas áreas en que se desenvuelven los procesos.
 - AUX_ESTADO: Se tomará de la tabla estado, concerniente a los estados que alcanzan los trámites.

- AUX_TIPO: Se tomará de la tabla tipo, concerniente a los tipos de trámites que se pueden alcanzar.
 - AUX_RECEPTOR: Se tomará de la tabla receptor, concerniente al receptor involucrado en cada trámite.
 - AUX_MEDIO_ENTREGA: Se tomará de la tabla medio de entrega, concerniente a la manera que se entregan las respuestas de los trámites.
- Segunda Etapa: Comprende en el cálculo y carga de datos desde la base de datos Auxiliar (dimTrenformacion) hacia la base de datos del datamart INPC, tanto las dimensiones como las tablas de hecho según procesos de medición definidos.

3.3.2. Dimensiones

- DIM_TIEMPO: Se tomará la información de la tabla AUX_TIEMPO
- DIM_LOCALIDAD: Se tomará la información de la tabla AUX_LOCALIDAD
- DIM_REMITENTE: Se tomará la información de la tabla AUX_REMITENTE
- DIM_PERSONA: Se tomará la información de la tabla AUX_PERSONA
- DIM_AREA: Se tomará la información de la tabla AUX_AREA
- DIM_ESTADO: Se tomará la información de la tabla AUX_ESTADO
- DIM_TIPO: Se tomará la información de la tabla AUX_TIPO
- DIM_RECEPTOR: Se tomará la información de la tabla AUX_RECEPTOR
- DIM_MEDIO_ENTREGA: Se tomará la información de la tabla AUX_MEDIO_ENTREGA

3.3.3. Hechos

- FACT_TRAMITE: Se tomará la información de las tablas AUX_LOCALIDAD, AUX_TIEMPO, AUX_PERSONA, AUX_AREA, AUX_ESTADO, AUX_TIPO, AUX_REMITENTE.
- FACT_RESPUESTA_TRAMITE: Se tomará la información de las tablas AUX_RECEPTOR, AUX_MEDIO_ENTREGA que permiten la relación que existe entre el remitente y la manera de entregar la documentación.

3.3.4. Estrategia de Refreshamiento de Datos

La carga inicial de los datos se realizará tomando del documento consolidado para el control de trámites a partir del 2013, debido al análisis realizado con anterioridad se conoce que existe información incorrecta y no estandarizada.

El proceso para refrescar los datos del *datamart* dependerá de la programación planificada de forma mensual por el INPC R3; Por lo cual la estrategia de refreshamiento de datos posee una carga transaccional acumulada con frecuencia mensual.

4. Diseño

4.1. Dimensiones

4.1.1. Dimensiones físico de la dimensión dim_tiempo

ATRIBUTO	DESCRIPCIÓN	TIPO DE DATO
Key: dim_tie_tiempo	Clave subrogada de la dimensión	Integer
dim_tie_fecha	Fecha completa: año, mes, día	Date
dim_tie_dia	Día en la fecha	Integer
dim_tie_mes	Mes en la fecha	Integer
dim_tie_anio	Año en la fecha	Integer
dim_tie_nombre_mes	Nombre del mes	Varchar(30)
dim_tie_estado_registro	Estado del registro	Char(1)

4.1.2. Dimensiones físico de la dimensión dim_localidad

ATRIBUTO	DESCRIPCIÓN	TIPO DE DATO
Key: dim_loc_localidad	Clave subrogada de la dimensión	Integer
Pk: dim_loc_id_ubicacion	Clave operacional de la ubicación geográfica	Integer
Id: dim_loc_id_Parroquia	Identificador del atributo parroquia	Integer
Id: dim_loc_id_Canton	Identificador del atributo cantón	Integer
Id: dim_loc_id_provincia	Identificador del atributo provincia	Integer
Id: dim_loc_id_zona	Identificador del atributo zona	Integer
dim_loc_parroquia	Nombre de la parroquia	Varchar(50)
dim_loc_canton	Nombre del cantón	Varchar(50)
dim_loc_provincia	Nombre de la provincia	Varchar(50)
dim_loc_zona	Nombre de la zona	Varchar(50)
dim_loc_codigo_parroquia	Código externo de parroquia	Varchar(20)
dim_loc_codigo_canton	Código externo de cantón	Varchar(20)
dim_loc_codigo_provincia	Código externo de provincia	Varchar(20)
dim_loc_codigo_region	Código externo de región	Varchar(20)
dim_loc_codigo_registro	Estado del registro	Varchar(50)

4.1.3. Dimensiones físico de la dimensión dim_remitente

ATRIBUTO	DESCRIPCIÓN	TIPO DE DATO
Key: dim_rem_remitente	Clave subrogada de la dimensión	Integer
Pk: dim_rem_id_remitente	Clave operacional del remitente	Integer
Pk: dim_rem_id_cargo	Clave operacional del cargo	Varchar(150)
Pk: dim_rem_id_dependencia	Clave operacional de la dependencia	Varchar(13)
dim_rem_nombre_remitente	Nombres y apellidos del remitente	Integer
dim_rem_identificacion_remitente	Identificación del remitente	Varchar(50)
dim_rem_detalle_dependencia	Detalle de la dependencia	Integer
dim_rem_detalle_cargo	Detalle del cargo	Varchar(50)
dim_rem_estado_registro	Estado del registro del remitente	Char(1)

4.1.4. Dimensiones físico de la dimensión dim_persona

ATRIBUTO	DESCRIPCIÓN	TIPO DE DATO
Key: dim_per_persona	Clave subrogada de la dimensión	Integer
Pk: dim_per_id_persona_asignada	Clave operacional del remitente	Integer
dim_per_nombre_persona	Nombres y apellidos del remitente	Varchar(150)
dim_per_identificacion_persona	Identificación del remitente	Varchar(13)
dim_per_estado_registro	Estado del registro del remitente	Char(1)

4.1.5. Dimensiones físico de la dimensión dim_area

ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Key: dim_area_area	Clave subrogada del área	Integer
Pk: dim_area_id_area	Clave operacional del área	Integer
dim_area_detalle_area	Detalle del área	Varchar(150)
dim_area_estado_registro	Estado del registro del área	Char(1)

4.1.6. Dimensiones físico de la dimensión dim_medio_entrega

ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Key: dim_med_medio_entrega	Clave subrogada del medio de entrega	Integer
Pk: dim_med_id_medio_entrega	Clave operacional del medio de entrega	Integer

dim_med_detalle_medio_entrega	Detalle del medio de entrega	Varchar(150)
dim_med_estado_registro	Estado del registro del medio de entrega	Varchar(13)

4.1.7. Dimensiones físico de la dimensión dim_estado

ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Key: dim_est_estado	Clave subrogada del estado de proceso de trámite	Integer
Pk: dim_est_id_estado_tramite	Clave operacional del estado de proceso de trámite	Integer
dim_est_detalle_estado_tramite	Detalle del estado del proceso de trámite	Varchar(150)
dim_est_estado_registro	Estado del registro del tipo de trámite	Char(1)

4.1.8. Dimensiones físico de la dimensión dim_tipo

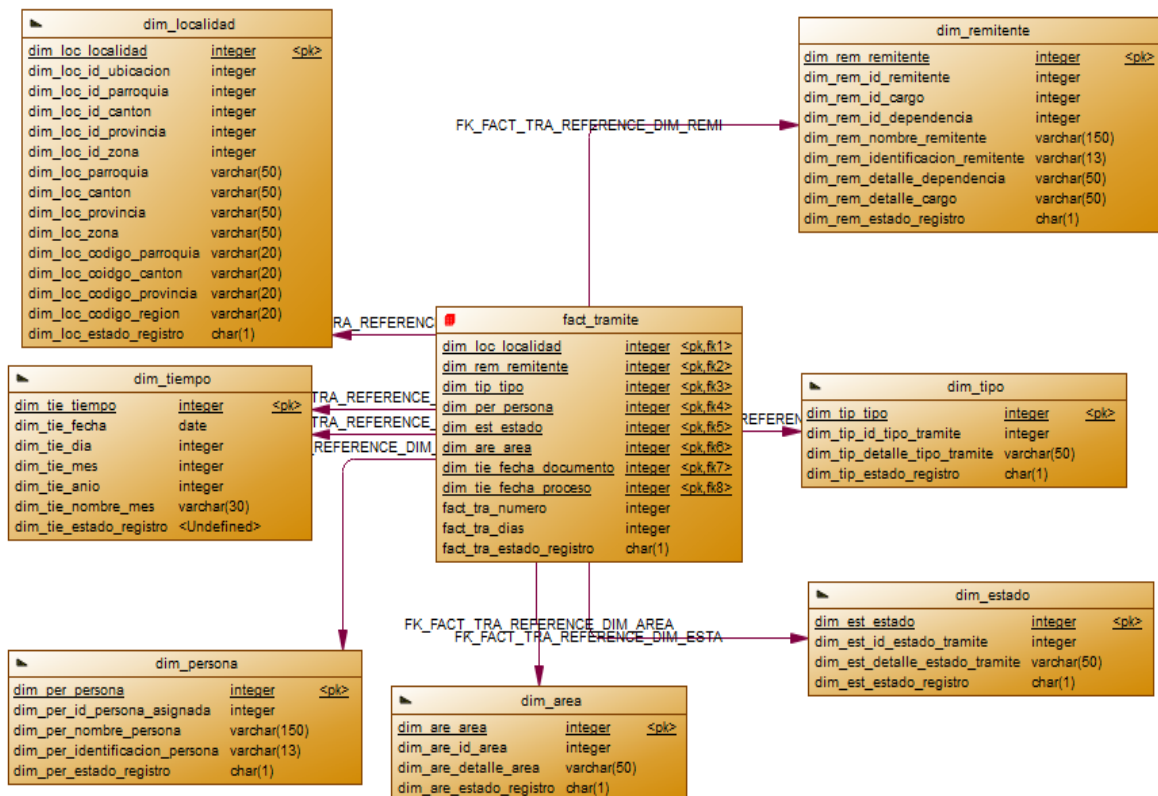
ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Key: dim_tip_tipo	Clave subrogada del tipo de trámite	Integer
Pk: dim_tip_id_tipo_tramite	Clave operacional del tipo de trámite	Integer
dim_tip_detalle_tipo_tramite	Detalle del tipo de trámite	Varchar(150)
dim_tip_estado_registro	Estado del registro del tipo de trámite	Char(1)

4.2. Hechos

4.2.1. Modelo físico de la tabla de hechos fact_tramite

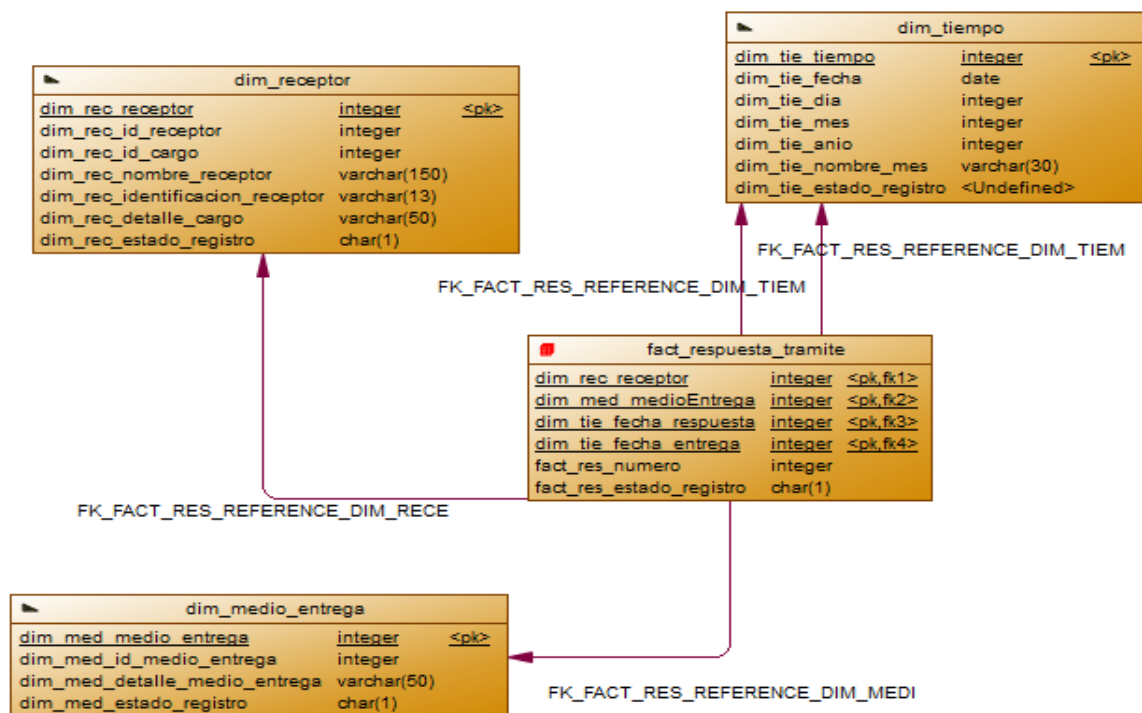
ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Fk: dim_loc_localidad	Clave operacional del proceso localidad	Integer
Fk: dim_rem_remitente	Clave operacional del proceso remitente	Integer
Fk: dim_tipo_tipo	Clave operacional del proceso tipo	Integer
Fk: dim_per_persona	Clave operacional del proceso de persona	Integer
Fk: dim_est_estado	Clave operacional del proceso estado	Integer
Fk: dim_are_area	Clave operacional del proceso área	Integer
Fk: dim_tie_fecha_documento	Clave operacional del proceso tiempo	Integer
Fk: dim_tie_fecha_proceso	Clave operacional del proceso tiempo	Integer
fact_tra_numero	Número de trámites	Integer

fact_tra_dias	Número de días	Integer
fact_tra_estado_registro	Estado del registro del tipo de trámite	Char(1)



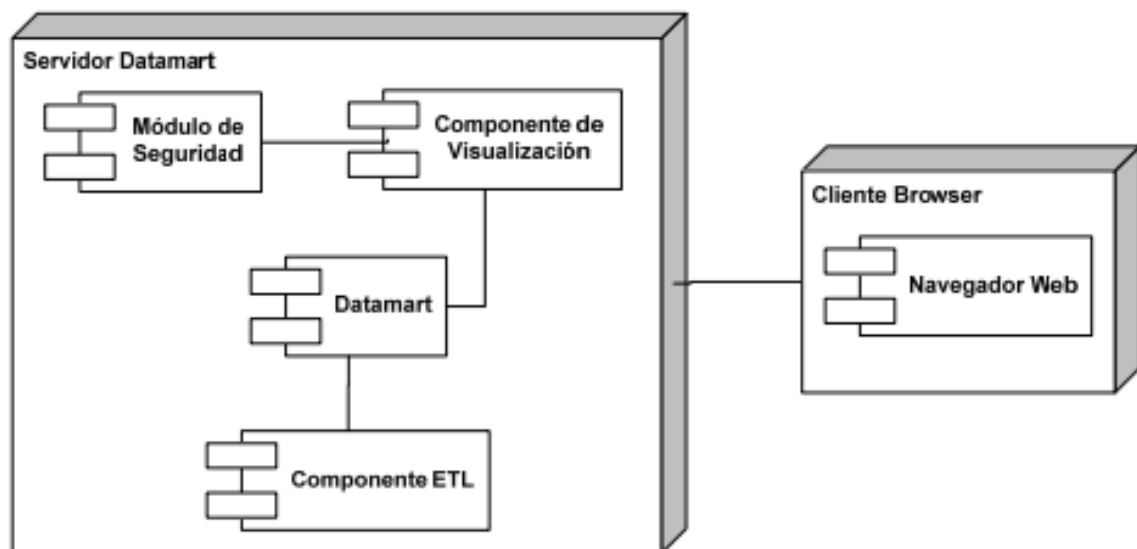
4.2.2. Modelo físico de la tabla de hechos fact_respuesta_tramite

ATRIBUTO	DESCRIPCIÓN	TIPO DATO
Fk: dim_rec_receptor	Clave operacional del proceso receptor	Integer
Fk: dim_med_medioEntrega	Clave operacional del proceso medio entrega	Integer
Fk: dim_tie_fecha_respuesta	Clave operacional del proceso tiempo	Integer
Fk: dim_tie_fecha_entrega	Clave operacional del proceso tiempo	Integer
fact_res_numero	Número de trámites	Integer
fact_tra_estado_registro	Estado del registro de tramite	Char(1)



4.3. Diseño de la arquitectura de la aplicación

La solución se implementará bajo una arquitectura cliente/servidor



4.3.1. Fuente de datos

Componente ETL: este componente está alojado en el servidor, procederá a los procesos de extracción, transformación y carga desde las fuentes de datos hacia el *Data Mart*, a través de la herramienta MySQL.

4.3.2. Data Mart

Componente *Data Mart*: Se encuentra alojado en el servidor realizará los procesos de exploración y explotación del cubo. Se lo implementará a través de *SpargoBI studio*.

4.3.3. Acceso y uso

- **Componente módulo seguridad:** Se ubica este componente en el servidor el cual se encarga del control de accesos de los reportes, se basa en la autenticación previa y asignación de perfiles de acceso.
- **Diseño de Interfaz de autenticación:** El usuario debe autenticarse al ingreso de sus credenciales (usuario y contraseña).

4.3.4. Componente de visualización

Este componente sirve de ayuda para la creación y visualización de reportes generados a partir del análisis realizado al cubo multidimensional.

Este componente se lo implementará con puesto que brinda prestaciones importantes en el momento de elaboración de este tipo de proyectos.

- **Diseño de interfaz de usuario:** Para los criterios de búsqueda se habilitará listas de selección de pendiendo las áreas e información que desee el usuario, con sus respectivos gráficos y tablas de resultados. Las listas o criterios para la búsqueda se ubicarán en la parte superior de la interfaz, en el menú se ubicará el acceso a los diferentes reportes que contiene la solución.
- **Componente navegador web:** Se encuentra alojado por defecto en cada cliente, se encargará del acceso, navegación y visualización de la solución mediante la intranet.

4.4. Definición de los niveles de seguridad

Para el acceso a los reportes se lo hará por medio de un módulo especializado para seguridades, en el cual se definirá los perfiles de acceso en base a la audiencia de destino y los indicadores señalados en los requerimientos de consulta.

INDICADOR	COORDINADOR	SUBCOORDINADOR
Nivel de satisfacción del cliente	x	x
Índice de trámites receptados	x	x
Índice de trámites por dependencia y remitente	x	x
Índice de prioridades de los trámites	x	x
Índice de respuestas	x	
Nivel de crecimiento del INPC respecto a ciertos trámites	x	
Nivel de crecimiento del INPC respecto a las respuestas	x	
Trámites atendidos con mayor fluidez	x	x
Responsables asignados a los trámites	x	x

5. Implementación

El desarrollo del *Data Mart* se realiza utilizando las siguientes herramientas:

1. Especificación de la Base de Datos del *Data Mart*

Plataforma		MySQL
Nombre		Inpc
Número de tablas		15
Owner	Usuario	Root
	Password	11032013
	Privilegios	Administrador

2. Producto seleccionado para la DBMS: Justificación de MySQL

Nombre	MySQL
Versión	5.7.18

3. Producto seleccionado para el área ETL

Nombre	SpagoBI studio
Versión	5.1

4. Producto seleccionado para el motor OLAP

Nombre	SpagoBI studio
Versión	5.1

5. Producto seleccionado como herramienta de acceso a datos

Nombre	SpagoBI studio
Versión	5.1

6. Producto seleccionado para la generación de cubos multidimensionales

Nombre	SpagoBI studio
Versión	5.1

5.1. Implementación de la estrategia de carga

Para la implementación de la esta estrategia se crearon DTS realizados con *SpagoBI studio, MySql* el que estará a cargo de realizar el proceso para facilitar la extracción, transformación, carga y despliegue del cubo multidimensional.

Cargar la base de datos transformación: cumple con las tareas que permiten los procesos de extracción, transformación y carga de los datos que se encuentran en la base de datos histórica y transaccional INPC R3 para su respectiva normalización.

Cargar Data Mart: Realiza el proceso de extracción, transformación y carga de datos desde la base de datos dimTransformacion hacia las dimensiones y tablas de hechos del *data mart* en la base INPC R3.

5.2. Implementación de la estrategia de refrescamiento

Para este proceso que involucra el *data mart* se creó una nueva tarea que ejecute El DTS de manera frecuente en el gestor de base de datos *MySql*, tomando en cuenta esta estrategia para el refrescamiento desde la fase de análisis.

Referencias

- BERNABEU, D. (2009). *Businness Intelligence*. Obtenido de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/1-business-intelligence/14-beneficios>
- BERNABEU, R. (2010). Obtenido de Data warehousing y metodología Hefesto: <http://www.dataprix.com/es/data-warehousing-hefesto>
- BUSTOS S & MOSQUERA, V. (2013). *Diseño e implementación de una solución business intelligence para la generación de indicadores y control de desempeño de la empresa OTECEL S.A; utilizando la metodología Hefesto V2.0 (Tesis de pregrado)*. Escuela Politécnica del Ejército. Obtenido de <https://repositorio.espe.edu.ec/bitstream/21000/6305/1/T-ESPE-047033.pdf>
- CANO, J. (2007). *Business Intelligence: Competir con información*. Obtenido de http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf
- CLARK, P. (2004). *Business Objects Products and Services: the Industry's Leading Suite of Integrated Business Intelligence Products*.
- COMPUTACIÓN. (11 de 2010). *Datos e Información*. Obtenido de <http://compuah.blogspot.com/2010/11/datos-e-informacion.html>
- CUELLAR, G. (2011). *Data warehouse, aspectos técnicos, características, usos, beneficios, componentes, herramientas OLAP*. Obtenido de <http://fccea.unicauca.edu.co/old/datawarehouse.htm>
- DEVLIN, B. (2011). *The data-driven Business and other Lessons from History*. Obtenido de <http://assets.en.oreilly.com/1/event/55/The%20Data-driven%20Business%20and%20Other%20Lessons%20from%20History%20Presentation.pdf>
- DÍAZ, J. (2010). *Introducción a Business Intelligence*. Obtenido de https://books.google.com.ec/books?id=iU3RAXYQXMkC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- ESPINOSA, R. (04 de 2010). *Descubriendo el business intelligence*. Obtenido de Rincon del BI: <https://churriwifi.wordpress.com/2010/04/>
- FERNÁNDEZ, C. (2015). *Fases de implantación de un data warehouse*. Obtenido de <http://www.dataprix.com/fases-de-implantacion-de-un-data-warehouse>
- FRAWLEY W & PIATESTKY-SHAPIO & MATHEUS, C. (2013). *Knowledge Discovery in Databases: An Overview*. Obtenido de <https://pdfs.semanticscholar.org/13a4/e7d48a80413ab49bada9e7c005b0b341600a.pdf>
- GARCÉS, M. (2015). *Metodologías e implmentación de alternativas businness intelligence opensource vs. propietarias en entornos tradicionales; caso prototipo en las pymes en el sector agroindustrial (Tesis de postgrado)*. Universidad de las Américas. Obtenido de <http://dspace.udla.edu.ec/bitstream/33000/2660/1/UDLA-EC-TMGSTI-2015-01.pdf>
- GARCÍA, R. (2012). *Minería de datos para la toma de decisiones e inteligencia de negocios: Aplicación en la Mercadotecnia. (Tesis de pregrado)*. Universidad Nacional Autónoma de México. Obtenido de <http://132.248.9.195/ptd2012/noviembre/098505131/098505131.pdf>
- GOMEZ, A. (2010). *Inteligencia de negocios: Estado del Arte*. Universidad Tecnológica de Pereira. Obtenido de <http://revistas.utp.edu.co/index.php/revistaciencia/article/view/1803/1209>

- HARO V & PÉREZ, W. (2014). *Data Warehouse para el centro de ambientación regional "Juan Bautista Vázquez" (Tesis de pregrado)*. Universidad de Cuenca. Obtenido de <http://dspace.ucuenca.edu.ec/bitstream/123456789/19878/1/tesis.pdf>
- INPC. (2015). *Instituto Nacional de Patrimonio Cultural*. Recuperado el Junio de 2016, de <http://mail.inpc.gob.ec/pdfs/lotaip2015/Enero/estatuto%20inpc.pdf>
- KIMBALL, G. (2016). *Kimball group*. Obtenido de <http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/kimball-reader/>
- KIMBALL, G. (2012). *Enfoques de desarrollo DW*. Obtenido de http://www.interaktiv.cl/blog/wp-content/uploads/2012/04/4.-Metodologia_disegno_DW1.pdf
- KNOBBE, A. (2007). *Multi-Relational Data Mining*. Obtenido de https://books.google.com.ec/books?id=ZwXvAgAAQBAJ&pg=PA111&lpg=PA111&dq=siebes+%26+holzheimer&source=bl&ots=Q7EZjy7qwj&sig=ecKKpuUI1XRnHKrHBTMKZPeZADs&hl=es&sa=X&ved=0ahUKEwjdi_fEg-vUAhUFQYKHbqgDoEQ6AEIMDAC#v=onepage&q=siebes%20%26%20holzheimer&f=false
- LEÓN, M. (2014). *Ciclo de vida de Ralph Kimball*. Obtenido de <http://luisleonin.blogspot.com/2014/02/ciclo-de-vida-de-ralph-kimball.html>
- LEÓN, M. (02 de 2017). *Inteligencia de Negocio*. Obtenido de Ciclo de vida de Ralph Kimball: <http://luisleonin.blogspot.com/2014/02/ciclo-de-vida-de-ralph-kimball.html>
- LONDO, F. (2015). *Desarrollo de un business intelligence en software libre, basado en indicadores de gestión, para una coordinación de salud (Tesis de postgrado)*. Pontificia Universidad Católica del Ecuador. Obtenido de <http://repositorio.pucesa.edu.ec/bitstream/123456789/1484/1/75981.pdf>
- LOZADA X & CRUZ, H. (2014). *Análisis, diseño, construcción e implementación de una data warehouse para la toma de decisiones y construcción de los KPI, para la empresa KRONOSCONSULTING CIA LTDA. (Tesis de pregrado)*. Escuela Superior Politécnica del Ejército. Obtenido de <http://repositorio.espe.edu.ec/xmlui/bitstream/handle/21000/8027/T-ESPE-047750.pdf?sequence=1&isAllowed=y>
- LUNA, M. (2007). *Análisis, diseño e implementación de un Datamart para el manejo de información socio educativa, psicológica para un centro de orientación juvenil (Tesis de pregrado)*. Escuela Politécnica Nacional. Obtenido de Tesis: <http://bibdigital.epn.edu.ec/bitstream/15000/496/1/CD-0434.pdf>
- MALDONADO, J. (2010). *Sistemas Adaptativos en Moodle: módulo de seguimiento de usuarios (Tesis de pregrado)*. Universidad Técnica Particular de Loja. Obtenido de <http://dspace.utpl.edu.ec/handle/123456789/1839?locale=es>
- MOINE J GORDILLO S & HAEDO, S. (2011). *Análisis comparativo de metodologías para la gestión de proyectos de minería de datos*. Obtenido de http://sedici.unlp.edu.ar/bitstream/handle/10915/18749/Documento_completo.pdf?sequence=1
- MURILLO, C. (2008). *Implementación de un Data Mart para el ramo de vehículos de una empresa aseguradora (Tesis de pregrado)*. Escuela Politécnica Nacional. Obtenido de <http://bibdigital.epn.edu.ec/handle/15000/1281>
- PEÑA J & SÚAREZ, J. (2008). *Utilización de información histórica para decisiones empresariales. (Tesis de pregrado)*. Pontificia Universidad Javeriana. Obtenido de <http://hermes.javeriana.edu.co/biblos/tesis/ingenieria/Tesis204.pdf>
- PROAÑO, D. (2013). *Construcción de un Data Warehouse a través de la herramienta de Business Intelligence de ORACLE, para la aplicación de gestión de ventas de la empresa "Electrofertás" (Tesis de pregrado)*.

- Universidad del Azuay. Obtenido de http://biblioteca.uazuay.edu.ec/opac_css/index.php?lvl=author_see&id=40226
- REDIFP. (2013). Catálogo de indicadores para la toma de decisiones de los Institutos de Formación Profesional. Obtenido de http://www.ilo.org/wcmsp5/groups/public/---americas/---ro-lima/---sro-san_jose/documents/publication/wcms_243198.pdf
- ROJAS, A. (2014). *Implementación de un Data Mart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el departamento de finanzas de la contraloría (Tesis de pregrado)*. Universidad de San Martín de Porres. Obtenido de http://www.repositorioacademico.usmp.edu.pe/bitstream/usmp/1061/1/rojas_a.pdf
- SINNEXUS. (2016). *Business Intelligence Informática estratégica*. Obtenido de http://www.sinnexus.com/business_intelligence/
- SINNEXUS. (2016). *Datos, formación, conocimiento*. Obtenido de http://www.sinnexus.com/business_intelligence/piramide_negocio.aspx
- TORRES, J. (2016). *Propuesta metodológica de solución de inteligencia de negocios aplicada al sistema informático integrado de talento humano y registro de contratos y actas de finiquito. Caso: Ministerio de Trabajo. (Tesis de postgrado)*. Universidad de las Américas. Obtenido de <http://dspace.udla.edu.ec/handle/33000/5325>.
- TUBÓN, D. (2016). *Desarrollo de Sistema OLAP para bibliotecas digitales que usan tecnologías Linked Data (Tesis de pregrado)*. Escuela Politécnica Nacional. Obtenido de <http://bibdigital.epn.edu.ec/handle/15000/17022>
- TUFIÑO, J. (2011). *Desarrollo del datamart para el sistema nacional de vigilancia de software libre (Tesis de pregrado)*. Escuela Politécnica Nacional. Obtenido de <http://bibdigital.epn.edu.ec/bitstream/15000/4101/1/CD-3861.pdf>
- TURNERO, I. (2015). *Documentación del SGC: el enfoque basado en procesos*. Obtenido de <http://www.monografias.com/trabajos97/documentacion-del-sgc-enfoque-basado-procesos/documentacion-del-sgc-enfoque-basado-procesos.shtml>
- UVIDIA, M. (2016). *Descubrimiento de conocimiento en base de datos para la toma de decisiones en la unidad de nivelación y admisión de la ESPOCH. (Tesis de postgrado)*. Escuela Superior Politécnica de Chimborazo. Obtenido de <http://repositorio.pucesa.edu.ec/bitstream/123456789/1601/1/76134.pdf>
- VICENTE, A. (03 de 2017). *Dato, Información y Conocimiento*. Obtenido de <http://tecnowalsh.blogspot.com/>
- VILLAREAL, R. (2013). *Estudio de metodologías de Data Warehouse para la implementación de repositorios de información para la toma de decisiones gerenciales (Tesis pregrado)*. Universidad Técnica del Norte. Obtenido de <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&uact=8&ved=0ahUKEwjR0cGz1IPXAhWKMYYKHSIbDvQQFgg5MAU&url=http%3A%2F%2Frepositorio.utn.edu.ec%2Fbitstream%2F123456789%2F7015%2F1%2F04%2520RED%2520080%2520TRABAJO%2520GRADO.pdf&usg=AOvVa>