

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR
FACULTAD DE INGENIERÍA
ESCUELA DE SISTEMAS**



**DISERTACIÓN DE GRADO PREVIA LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN SISTEMAS Y COMPUTACIÓN**

**“ANÁLISIS COMPARATIVO DE HERRAMIENTAS OPEN SOURCE
PARA SOLUCIONES DE INTELIGENCIA DE NEGOCIOS (BI) Y SU
APLICACIÓN PRÁCTICA PARA LA TOMA DECISIONES EN LA
EMPRESA LÁCTEOS DE MARCO’S”**

AUTORES:

**ADRIÁN REMIGIO CARRIÓN ALBÁN
MARK CHRISTOPHER TORRES SOSA**

DIRECTOR:

ING. ANA URGILÉS

QUITO, SEPTIEMBRE 2016

DEDICATORIA

Dedico y agradezco a mi familia por la constante preocupación durante este proceso en especial a mi madre y abuelita, Mónica Sosa y Rebeca Herrera, por su cariño y apoyo incondicional. Además de sus consejos y enseñanzas de todos los días para guiarme por el camino correcto y permitirme ser lo que hoy soy, siempre serán mi orgullo y ejemplo de vida a seguir. A mi abuelito Tomás Sosa porque desde el cielo ha sido mi motor para seguir adelante y luz para cumplir mis objetivos.

Mark Christopher Torres Sosa

Dedico este trabajo a mi madre Alexandra Albán, mi padre Iván Carrión y mis hermanos Valeria y Juan José por ser el pilar fundamental y el apoyo incondicional para seguir en este camino largo, pero fructífero, ellos son mi ejemplo y el motor de mi vida siempre. A mi familia, mis abuelos, tíos, primos por cada momento ameno que he compartido junto a ellos y a todas las personas que en alguna oportunidad pude conocer y me ayudaron a forjar lo que hoy soy. Para terminar, creo firmemente que el éxito no es más que la suma de actitud y perseverancia, nos vemos en 10 años. Mil gracias.

Adrián Remigio Carrión Albán

Finalmente, un agradecimiento especial a Lácteos de Marco's y a la familia Proaño Campaña por su apoyo durante el desarrollo de este trabajo. Además, gracias a todos los amigos, compañeros y profesores que estuvieron durante estos años de esta etapa que termina.

RESUMEN

La presente disertación de grado expone un análisis comparativo de diferentes herramientas Open Source para soluciones de Inteligencia de Negocios, con el objetivo de mostrar a la empresa “Lácteos de Marco’s” los beneficios que conlleva la generación de diferentes reportes personalizados, permitiendo además agrupar y visualizar la información desde diferentes perspectivas de análisis según las necesidades de negocio.

En este proyecto se explora el término Inteligencia de Negocios junto con la aplicación de la Suite de Pentaho Community Edition (CE) para la generación de varios reportes útiles que soporten la toma de decisiones por parte de la alta gerencia. El proceso de construcción del data warehouse se lo realiza bajo la metodología HEFESTO, que incluye el análisis de requerimientos, análisis de los sistemas transaccionales, modelo lógico y la integración de datos.

TABLA DE CONTENIDOS

CAPÍTULO I	1
1. MARCO REFERENCIAL	1
1.1 Planteamiento del Problema.....	1
1.1.1 Formulación del problema	2
1.2 Justificación.....	2
1.3 Objetivos	3
1.3.1 General.....	3
1.3.2 Específicos	3
1.4 Alcance.....	3
CAPÍTULO II.....	4
2. MARCO TEÓRICO	4
2.1 Antecedentes	4
2.2 Fundamentación Teórica.....	5
2.2.1 La información como recurso vital en la empresa	5
2.2.1.1 Adaptación de las empresas a la era digital.....	6
2.2.1.2 Importancia de la información	7
2.2.2 Introducción a Data Warehouse.....	8
2.2.2.1 Introducción a data mart.....	12
2.2.2.2 Arquitectura data warehouse.....	14
2.2.2.3 Introducción al procesamiento analítico en línea (OLAP).....	16
2.2.2.4 Introducción al Data Mining	27
2.2.3 Inteligencia de Negocios (BI)	28
2.2.3.1 Evolución de la Inteligencia de Negocios (BI)	28
2.2.3.2 Definición de Inteligencia de Negocios	30
2.2.3.3 Importancia de Inteligencia de Negocios	32
2.2.3.4 Componentes de Inteligencia de Negocios	34
2.2.3.5 Herramientas para Inteligencia de Negocios (BI).....	38
CAPÍTULO III.....	45
3. ANÁLISIS COMPARATIVO DE HERRAMIENTAS Y METODOLOGÍAS PARA EL DISEÑO DE UN DATA WAREHOUSE.....	45
3.1 Herramientas Open Source para Inteligencia de Negocios (BI)	45

3.1.1	Pentaho Community Edition.....	46
3.1.2	Eclipse Business Intelligence and Reporting Tools (BIRT).....	48
3.1.3	JasperSoft Community Edition.....	49
3.2	Parámetros y Categorías de Comparación	52
3.2.1	Infraestructura.....	52
3.2.2	Administración de datos	52
3.2.3	Análisis y creación de contenido	53
3.2.4	Intercambio de resultados	53
3.3	Estudio comparativo entre las herramientas.....	54
3.3.1	Comparativa de reportería.....	59
3.3.2	Comparativa de reportería Pentaho.....	59
3.3.2.1	Componentes Pentaho.....	59
3.3.2.2	Pentaho Report Designer.....	61
3.3.2.3	Fortalezas y debilidades Pentaho	64
3.3.3	Comparativa de reportería BIRT	65
3.3.3.1	Componentes BIRT.....	65
3.3.3.2	BIRT Report Designer	66
3.3.3.3	Fortalezas y Debilidades BIRT	70
3.3.4	Comparativa de reportería JasperSoft.....	71
3.3.4.1	Componentes JasperSoft	71
3.3.4.2	JasperSoft Studio.....	72
3.3.4.3	Fortalezas y Debilidades JasperSoft.....	74
3.3.5	Resultados del estudio comparativo.....	75
3.4	Metodologías para la creación de un data warehouse	76
CAPÍTULO IV.....		79
4.	DESARROLLO Y APLICACIÓN DE LA SOLUCIÓN.....	79
4.1	Información de la empresa	79
4.1.1	Datos informativos.....	79
4.1.2	Situación actual.....	79
4.1.3	Misión	80
4.1.4	Visión.....	80
4.1.5	Objetivos.....	81
4.1.6	Estrategias	81

4.1.7	Valores	81
4.1.8	Políticas.....	81
4.1.9	Estructura de la organización.....	82
4.1.10	Situación actual de TI	82
4.1.10.1	Hardware	83
4.1.10.2	Software.....	83
4.1.11	Importancia del data warehouse para la empresa y su alcance.....	84
4.2	Metodología HEFESTO.....	85
4.2.1	Pasos y aplicación de la metodología	86
4.2.1.1	Análisis de requerimientos	86
4.2.1.2	Análisis de los OLTP	91
4.2.1.3	Modelo lógico del data warehouse.....	100
4.2.1.4	Integración de datos	106
4.3	Creación de Cubos multidimensionales	114
4.4	Presentación de información y generación de reportes	116
CAPÍTULO V.....		126
5.	CONCLUSIONES Y RECOMENDACIONES	126
5.1	Conclusiones	126
5.2	Recomendaciones.....	127
BIBLIOGRAFÍA		129
ANEXOS		131
Anexo A	131
Anexo B	159
Anexo C	163

ÍNDICE DE FIGURAS

Ilustración 1. Estructura de un data warehouse	16
Ilustración 2. Cubo dimensional	19
Ilustración 3. Ejemplo de esquema estrella.....	24
Ilustración 4. Ejemplo de la tabla dimensión producto (incluyendo instancias)	25
Ilustración 5. Ejemplo de normalización de la dimensión almacén.....	26
Ilustración 6. Necesidades de Información en la Empresa	33
Ilustración 7. Componentes de una arquitectura BI Empresarial madura	36
Ilustración 8. Proceso de Business Intelligence	37
Ilustración 9. Cuadrante Mágico para plataformas de Business Intelligence y analíticas	38
Ilustración 10. Arquitectura BI Tableau	39
Ilustración 11. Arquitectura BI Qlik	40
Ilustración 12. Arquitectura BI SAS Intelligence Platform	41
Ilustración 13. Component Architecture.....	42
Ilustración 14. JasperSoft BI Architecture.....	43
Ilustración 15. Arquitectura BI BIRT	44
Ilustración 16. Pantalla principal Pentaho Report Designer	61
Ilustración 17. Pantalla principal editor de reportes	62
Ilustración 18. Asistente para crear reportes.....	63
Ilustración 19. Vista previa del reporte.....	63
Ilustración 20. Editor de gráficos Pentaho	64
Ilustración 21. Pantalla principal BIRT	68
Ilustración 22. “Asistente” para crear reportes	68
Ilustración 23. “Asistente” para crear reportes	69
Ilustración 24. “Asistente” para crear reportes	69
Ilustración 25. Pantalla principal JasperSoft Studio	72
Ilustración 26. “Asistente” para selección de gráficas.....	73
Ilustración 27. Ventana para configuración de datos en la gráfica.....	74
Ilustración 28. Arquitectura bottom-up de un data warehouse	77

Ilustración 29. Arquitectura top-down de un data warehouse	77
Ilustración 30. Organigrama de Negocio	82
Ilustración 31. Metodología HEFESTO, pasos.....	85
Ilustración 32. Gestión de marketing actua.....	87
Ilustración 33. Modelo conceptual.....	90
Ilustración 34. Modelo conceptual de ventas.....	91
Ilustración 35. Modelo conceptual de compras	91
Ilustración 36. Modelo físico de la sección de venta de productos	94
Ilustración 37. Modelo físico de la sección de compra de productos	95
Ilustración 38. Modelo conceptual ampliado de ventas.....	99
Ilustración 39. Modelo conceptual ampliado de compras	99
Ilustración 40. Tabla dimensión “PRODUCTO”.....	101
Ilustración 41. Tabla dimensión “FECHA”	101
Ilustración 42. Tabla dimensión “CIUDAD”.....	101
Ilustración 43. Tabla dimensión “CLIENTE”	102
Ilustración 44. Tabla dimensión “MATERIAPRIMA”	102
Ilustración 45. Tabla dimensión “FECHA”	103
Ilustración 46. Tabla dimensión “PROVEEDOR”	103
Ilustración 47. Tabla de hechos “VENTAS”	104
Ilustración 48. Tabla de hechos “COMPRAS”	105
Ilustración 49. Uniones del proceso de venta	105
Ilustración 50. Uniones del proceso de compra	106
Ilustración 51. Vista general de la transformación	107
Ilustración 52. Conexión a la base de datos temporal de la empresa.....	107
Ilustración 53. Conexión al data warehouse	108
Ilustración 54. Carga inicial del data mart de ventas	108
Ilustración 55. Carga inicial del data mart de compras.....	112
Ilustración 56. Cubo multidimensional de ventas.....	115
Ilustración 57. Cubo multidimensional de compras	116
Ilustración 58. Cubo de ventas mediante Saiku	117
Ilustración 59. Parámetros para reporte general de ventas	118

Ilustración 60. Reporte general de ventas	118
Ilustración 61. Parámetros para reporte de venta de un producto específico por año	119
Ilustración 62. Reporte de venta de un producto específico por año	119
Ilustración 63. Parámetros para reporte de venta de todos los productos por año y mes	120
Ilustración 64. Reporte de ventas de todos los productos por año y mes	120
Ilustración 65. Tabla reporte de ventas de todos los productos por año y mes.....	121
Ilustración 66. Tabla reporte de ventas de todos los productos por año y mes.....	122
Ilustración 67. Reporte general de compras.....	122
Ilustración 68. Parámetros para reporte de compra de materia prima específica por año	123
Ilustración 69. Reporte de compra de materia prima específica por año.....	123
Ilustración 70. Parámetros para reporte de compra de toda la materia prima por año y mes	124
Ilustración 71. Reporte de compra de toda la materia prima por año y mes.....	124
Ilustración 72. Tabla reporte de compra de toda la materia prima por año y mes.....	125

ÍNDICE DE TABLAS

Tabla 1: Diferencias entre sistemas transaccionales y los basados en data warehouses	11
Tabla 2: Cuadro comparativo de herramientas	51
Tabla 3: Cuadro comparativo.....	54
Tabla 4: Cuadro de escala ponderativa	58
Tabla 5: Cuadro ponderativo	58
Tabla 6: Hardware que dispone la empresa	83
Tabla 7: Campos de las perspectivas de análisis del proceso de venta.....	97
Tabla 8: Campos de las perspectivas de análisis del proceso de compra	98

CAPÍTULO I

1. MARCO REFERENCIAL

1.1 Planteamiento del Problema

En la actualidad la mayoría de organizaciones manejan sistemas de información que diariamente generan grandes volúmenes de datos, los cuales permiten saber la situación actual de las mismas. Con la introducción de las nuevas tecnologías y herramientas para el manejo y análisis de información, es importante que las empresas implementen este tipo de soluciones, para convertir los datos e información en conocimiento útil, que ayude a la toma de decisiones con el fin de lograr el éxito en el competitivo y dinámico mercado de hoy. Los datos se transforman en información cuando han sido interpretados y se aplica a un propósito concreto. La información generada es útil cuando se puede aplicar a la creación de conocimiento dentro de una organización; a todo este proceso se le atribuye el término de Inteligencia de Negocios (BI) que se enfoca en la creación y administración de la información, que permite tomar mejores decisiones a los altos mandos de una organización para promover una ventaja competitiva en el mercado.

Por lo tanto, el problema radica en que la información generada muchas veces no es manejada adecuadamente y no se cuenta con una solución para Inteligencia de Negocios (BI) que facilite el adquirir conocimiento para tomar mejores decisiones dentro de la organización.

En el ámbito de soluciones para Inteligencia de Negocios (BI) existen herramientas propietarias, las cuales lideran el mercado por el alto interés que ponen las organizaciones que las desarrollan en brindar un servicio de soporte rápido y de calidad a los usuarios, además de las características funcionales propias de las herramientas. Sin embargo, su alto costo conlleva a que se busquen otras alternativas de herramientas para Inteligencia de Negocios Open Source, objeto de estudio en este proyecto.

1.1.1 Formulación del problema

¿Cómo optimizar los recursos para mejorar la toma de decisiones del departamento financiero de la Empresa Lácteos de Marco's?

1.2 Justificación

Actualmente, pequeñas, medianas y grandes organizaciones se han orientado por la adquisición de herramientas Open Source o licenciadas para Inteligencia de Negocios (BI), como una alternativa tecnológica y de administración, que permite gestionar la información para la toma de decisiones en todos los niveles de una organización. Desde la extracción, depuración y transformación de datos, hasta la explotación y distribución de la información mediante herramientas de fácil uso para los usuarios.

Lácteos de Marco's cuenta con un software de gestión empresarial llamado "SIGEMP" que automatiza ciertos procesos como: Recepción de leche, Registro de proveedores, Inventario de materia prima y producto terminado, Producción diaria, Facturación, kárdex de productos, Resumen de despachos. Sin embargo, la alta gerencia requiere tener acceso a la información histórica, con estadísticas de proyección organizada por diferentes indicadores, de tal forma que la toma de decisiones sea acertada y sencilla.

Para poder afrontar dichos requerimientos, en la presente investigación se plantea la necesidad de analizar herramientas Open Source para Inteligencia de Negocios (BI) con el fin de aplicar la herramienta seleccionada en la empresa para que proporcione información clave a la alta gerencia.

La aplicación de la herramienta seleccionada, permitirá a Lácteos de Marco's disminuir tiempo y costos en el proceso de generación de consultas, reportes especializados por indicadores y el análisis de la información para la toma rápida y adecuada de decisiones.

1.3 Objetivos

1.3.1 General

Realizar un análisis comparativo de herramientas Open Source para soluciones de Inteligencia de Negocios (BI), con la finalidad de seleccionar una herramienta y aplicarla en el departamento financiero de la empresa Lácteos de Marco's para brindar soporte en la toma de decisiones.

1.3.2 Específicos

- Investigar las herramientas Open Source para Inteligencia de Negocios (BI) más usadas en el mercado.
- Seleccionar la herramienta Open Source y analizar los datos relacionados al departamento financiero de la empresa Lácteos de Marco's.
- Aplicar en la empresa Lácteos de Marco's la herramienta seleccionada como mejor opción que brinde soporte en la toma de decisiones.
- Evaluar resultados de la aplicación funcional de la herramienta para Inteligencia de Negocios (BI) seleccionada.
- Aportar con una fuente de información útil de aprendizaje sobre herramientas Open Source para Inteligencia de Negocios (BI), orientado a personas que deseen incursionar en esta área.

1.4 Alcance

Al finalizar el presente proyecto de disertación de grado se entregará un documento con el análisis comparativo de herramientas Open Source para soluciones de Inteligencia de Negocios (BI), además de la aplicación práctica de la herramienta para la toma de decisiones en el departamento financiero de la empresa Lácteos de Marco's.

CAPÍTULO II

2. MARCO TEÓRICO

2.1 Antecedentes

En el mundo empresarial donde la competencia y la demanda de calidad en los servicios y productos es cada vez mayor, es importante que las organizaciones puedan disponer de la información necesaria y a la orden del día para tomar decisiones acertadas con la finalidad de mantener una competitividad en un mercado cada vez más saturado de opciones para los consumidores. Para ello es necesario estar siempre pendiente de las necesidades actuales y utilizar herramientas que faciliten llevar a cabo los procesos que se vinculen a una organización. Una de las áreas con mayor auge en los últimos tiempos que ha sido punto clave en el éxito de muchas organizaciones es la Inteligencia de Negocios o BI, por sus siglas en inglés. Una interesante definición publicada en el sitio web del TDWI¹(s.f.), define a la Inteligencia de Negocios (BI) como “la combinación de tecnología, herramientas y procesos que permiten transformar los datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial”.

Inteligencia de Negocios (BI), tuvo su origen en la década de los 60's donde los primeros sistemas de información aparecieron. Estos sistemas almacenaban los datos en medios magnéticos de forma secuencial. En la década de los 70's aparecieron los primeros sistemas gestores de base de datos que dieron lugar a la utilización de modelos relacionales para la resolución de problemas de almacenamiento de la información. Este nuevo paradigma en los modelos de base de datos fue presentado por Edgar Codd² en 1969. Durante esta época se pudo apreciar un gran salto en la evolución de las bases de datos, ya que hasta entonces las mismas se basaban en su mayoría en modelos de red, jerárquicos o simplemente archivos estructurados, cuyo problema principal era la

¹TDWI (The Data Warehousing Institute) es la principal institución educativa en el ámbito de Inteligencia de Negocios y data warehousing. Fundada en 1995, se dedica a ayudar a las organizaciones y empresas a aumentar su comprensión y uso de Inteligencia de Negocios mediante el despliegue de estrategias y tecnologías de data warehousing.

² Edgar Codd, Científico informático creador del modelo relacional de base de datos.

inflexibilidad y las relaciones físicas entre las entidades. Si bien las bases de datos tuvieron un gran impulso de la mano del modelo relacional, recién a finales de esta década se construyeron las primeras versiones de sistemas que les daban soporte. A pesar de todo, los informes continuaban siendo estáticos y altamente orientados a la información transaccional.

La base de datos es una entidad en la cual se puede almacenar datos de manera estructurada. Cuando un usuario desea realizar consultas transaccionales, se puede acceder dependiendo del tipo de permisos que tenga y extraer la información que necesite en un período corto de tiempo. ¿Pero qué sucede cuando se manejan grandes cantidades de datos y múltiples usuarios desean acceder a estos? Uno de los factores causales que volvieron a las bases de datos poco operativas es el crecimiento exponencial de las organizaciones que trajo consigo el aumento cuantitativo de los datos a almacenar y la necesidad de manejar información histórica que permita ayudar en la toma de decisiones en una organización. Aquí es donde las herramientas de Inteligencia de Negocios (BI) intervienen y ayudan en la manipulación de los datos de manera dinámica que permita dar más valor a la información que maneja la organización para convertirla en conocimiento.

2.2 Fundamentación Teórica

En esta sección del capítulo se presentan los temas más relevantes entorno a la Inteligencia de Negocios (BI) para el cumplimiento del objetivo de este trabajo. Comienza con una introducción de cómo la información es útil para una organización y cómo esta ayuda en la toma de decisiones. Además, se muestra una introducción a data marts, data warehouse, OLAP y Data Mining; así como aspectos importantes de la Inteligencia de Negocios (BI), como su evolución, importancia, componentes y herramientas.

2.2.1 La información como recurso vital en la empresa

En la sociedad actual, uno de los sucesos más significativos es el manejo de grandes cantidades de información que se pueden obtener de múltiples fuentes. En el mundo empresarial esta afirmación es muy cierta, además; pero gracias al auge

tecnológico que cada vez es mayor, las organizaciones disponen de sistemas informáticos completos que ayudan al manejo de los grandes volúmenes de información proporcionados por los sistemas de información. No debemos confundir los sistemas de información con los sistemas informáticos, ya que según Whitten, Bentley y Dittman (2004), “un sistema de información es un conjunto de personas, datos, procesos y tecnología de información que interactúan para recoger, procesar, almacenar y proveer la información necesaria para el correcto funcionamiento de la organización”. A continuación, se presentan aspectos importantes a considerar acerca de la importancia de la información para la toma de decisiones en las organizaciones.

2.2.1.1 Adaptación de las empresas a la era digital

Sin duda alguna, estamos viviendo una era trascendental en la que la información está al alcance de todos, lo que genera una mayor competencia y demanda de calidad en los servicios y productos; como resultado la información se ha convertido en un recurso vital para que las organizaciones puedan tener éxito y sobresalgan de la competencia. “En la nueva economía, la información en todas sus formas es digital, reducida a bits almacenados en computadoras, desplazándose a la velocidad de la luz a través de las redes. De esta manera, se crea un nuevo mundo de posibilidades que trascienden a las interacciones con base en lo físico” (Schuster Rosa, 2016, p.3).

Esta nueva era se caracteriza por la rapidez en la que cambia la naturaleza de las organizaciones, en la que se desarrollan las transacciones y de cómo el acceso a la información cambia el panorama de consumo y las propias expectativas de las organizaciones. Estos cambios que se están teniendo se ven determinados por dos aspectos: la renovación tecnológica que emerge de antiguos modelos y la adaptación de la sociedad, ya que las personas cambian continuamente sus exigencias y por ende las circunstancias. He aquí el gran reto de las organizaciones actualmente, poder identificar las oportunidades de cómo utilizar la información y tecnología a favor y así llegar a una madurez digital que permita afianzarse en la cima de esta nueva era.

El mundo digital no está diseñado a medida, aunque es cierto que hay acuerdos a nivel mundial sobre el desarrollo de las telecomunicaciones y planes estratégicos que abarcan el diseño de las nuevas infraestructuras y tecnologías para un futuro inmediato. Y debido a la gran movilidad y productividad que permite la nueva era digital, obliga a las organizaciones a remodelar constantemente las estrategias para permitir que se tome decisiones acertadas en bien de la organización.

2.2.1.2 Importancia de la información

En esta época donde la intercomunicación global de un gran número de actividades aumenta a una velocidad antes desconocida, la información adquiere para las empresas un carácter estratégico para el mejoramiento continuo, el buen manejo de la misma puede significar el éxito o fracaso para una organización.

Al pasar los años, el manejo de los recursos dentro de las organizaciones ha cambiado. Hasta hace unos años los activos más importantes para una organización eran los recursos tangibles como puede ser el capital o la mano de obra, pero a comienzos del nuevo siglo esto fue cambiando debido al auge tecnológico y al nuevo modelo de negocio basado en competencias. Es por ello que las organizaciones han optado por darle importancia a otros recursos intangibles como el conocimiento, habilidades, valores y aptitudes de las personas que forman parte de la organización. En realidad, se ven muchos casos de éxito actualmente, en el que muchas organizaciones pequeñas han sabido utilizar y explotar de mejor manera el conocimiento en determinado tema y así generar riquezas.

Vivimos en la era de la información, donde esta tiene un valor económico implícitamente, ya que provee mayor competitividad a una organización a nivel global, siempre y cuando la alta gerencia sepa aprovechar y explotar al máximo la información disponible del entorno; por lo cual la información se ha convertido en un recurso vital y está presente día a día dentro de cualquier organización, en cada actividad que se realiza. Es parte primordial de la naturaleza de cualquier organización, porque en esencia una organización es un conjunto de personas que

se interrelacionan intercambiando información. Por ello es importante que las organizaciones sepan convertir la información en un producto útil, facilitando las condiciones para que esta fluya sin obstáculos entre los distintos departamentos permitiendo así el logro conjunto de metas establecidas para lograr el éxito en el mundo de los negocios.

2.2.2 Introducción a Data Warehouse

A medida que la tecnología avanza, se evidencia la automatización que realizan las organizaciones en sus procesos; manejando gran cantidad de datos con la ayuda de sistemas informáticos y bases de datos. Todos los datos procesados se convierten en información y con el manejo adecuado en conocimiento para la misma organización.

El alto nivel competitivo del mercado actual obliga a las organizaciones a buscar nuevas estrategias de gestión para cumplir con los objetivos planteados y lograr alcanzar el éxito. Hace algunos años, la mayoría de organizaciones se manejaban de forma jerárquica, basada en la existencia de funciones superiores e inferiores. Este tipo de organización empresarial se la conoce como piramidal, donde se refleja el flujo de información, desde el nivel operativo hasta la alta gerencia, donde se encuentran las personas encargadas de la toma de decisiones dentro de la organización. La concepción de una nueva visión de la organización, ha permitido que estas organizaciones se reestructuren y gestionen bajo el concepto de empresa moderna, las cuales se diseñan en base a tareas y buscan nuevas formas de crear escenarios que ayuden a ser más productivos y mejorar continuamente. El mayor logro de este tipo de visión es la toma de decisiones en todos los niveles, lo que permite ser más eficiente porque se asigna responsabilidades a las personas en cada nivel, las cuales conocen a detalle cómo se manejan los procesos, por ello es importante que se cuente con información sólida para influenciar y apoyar las decisiones.

Todo lo mencionado anteriormente nos conduce al concepto de Data Warehouse, que convierten los datos operacionales en información útil disponible en cualquier momento para su consulta y posterior análisis.

En el libro *Introducción al Business Intelligence* publicado por la Universidad Oberta de Catalunya, Díaz (2010) afirma:

En los últimos años han aparecido muchos enfoques de data warehouse tanto a nivel tecnológico como estratégico, ejemplo de ello son: La aparición de bases de datos especializadas en el despliegue de data warehouse en forma de software o appliances, desarrollo de múltiples metodologías que buscan capacitar la organización para lidiar con el reto de un crecimiento exponencial de datos articulado en tres dimensiones: volumen, velocidad y variedad, lo que se conoce como Big Data y las tendencias SaaS o in-memory que buscan reducir el tiempo de creación de estructuras de datos. Se evidencia como el enfoque de la Inteligencia de Negocios ha ido evolucionando y dentro de este marco uno de los aspectos más importantes, el Data warehouse.

Un data warehouse puede ser visto como un depósito donde están almacenados todos los datos necesarios para dar soporte a la toma de decisiones dentro de la organización, proporcionando una visión global e íntegra de su estructura; para ello es importante que el data warehouse sea estable, coherente, fiable y cuente con información histórica. El objetivo del data warehouse entonces es permitir el acceso a los usuarios a enormes cantidades de datos operacionales integrados, para su análisis y posterior apoyo a las necesidades de los usuarios que lo requieran. La forma tradicional de entregar información a los usuarios que la requerían era a través de la generación de reportes impresos que se obtenían de los sistemas de información transaccionales³, también llamados SI de proceso de transacción mediante consultas a las bases de datos y posterior extracción de los datos para que en cierta manera se pueda apoyar a la gestión de las organizaciones. A medida que los avances han ido apareciendo en torno a las tecnologías de la información, las organizaciones han tenido que enfrentar nuevos desafíos que les permita analizar, descubrir y entender toda la información que manejan con el fin de lograr tener ventaja competitiva con

³ En el libro “Uso de los sistemas de información en las organizaciones” de los autores Pastor Collado y Joan Antoni, los sistemas transaccionales son aquellos sistemas del SI global de la organización que se encargan de manera específica de procesar tanto las transacciones de información provocadas por las interacciones formales entre el entorno y la organización como las transacciones generadas en el seno de la organización.

respecto al mercado en el que se encuentran inmersas. Pero debido al gran crecimiento que en los últimos años ha tenido el internet y las aplicaciones, surge el problema de manejar dinámicamente toda la información. Ahí en donde los sistemas transaccionales tienen problemas, debido a que la entrega de información es inconsistente, inflexible y carente de integración a través de la organización.

Los sistemas transaccionales son cambiantes ya que siempre están actualizando sus datos. La extracción de los datos es cuestión de segundos, dependiendo del número de registros y la complejidad de la consulta, pero el análisis de los mismos puede resultar pesado porque implicaría el consumo adicional de recursos de cómputo. Realizar el análisis complejo de un sistema transaccional puede resultar en el mal funcionamiento del sistema, lo que provocaría un retardo evidente en las operaciones del negocio. El data warehouse responde a la necesidad de analizar la información sin sacrificar el rendimiento de los sistemas transaccionales de los cuales se obtienen los datos. El data warehouse genera una base de datos con perspectiva histórica, integrando y depurando información de distintas fuentes que se fusionan de forma acorde a las necesidades de análisis, que pueden ir desde un nivel de detalle hasta un nivel de transacción, con la intención de tener disponible la mayor parte de los datos para analizar y posteriormente realizar reportes dinámicos. Por lo tanto, un data warehouse es un repositorio de datos transaccionales para proporcionar consultas operativas, permitiendo realizar un análisis multidimensional.

Para entender mejor lo citado anteriormente se muestra a continuación algunas diferencias entre los sistemas transaccionales y los que manejan data warehouse.

Tabla 1:*Diferencias entre sistemas transaccionales y los basados en data warehouses*

Transaccionales	Basados en data warehouses
Permiten el acceso concurrente de múltiples usuarios que agregan y modifican datos.	Permiten el acceso concurrente de múltiples usuarios que consultan y no modifican datos.
Reflejan el estado cambiante de una organización sin guardar su historial.	Guardan el historial cambiante del estado de una organización.
Almacenan grandes cantidades de datos, entre datos extensivos utilizados para comprobar transacciones.	Almacenan grandes cantidades de datos, sumariados, consolidados y transformados, los necesarios para el análisis.
Tienen estructuras de base de datos complejas.	Tienen estructuras de Base de datos simples.
Se adaptan para responder a transacciones.	Se adaptan para responder a consultas.
Suministran la infraestructura tecnológica necesaria para admitir las operaciones diarias de una organización.	Suministran la infraestructura tecnológica necesaria para admitir análisis de datos de una organización.
Los analistas no cuentan con la experiencia técnica necesaria para crear consultas "ad hoc" contra la compleja estructura de datos.	Permiten combinar datos de orígenes heterogéneos en una única estructura homogénea y simple, facilitando la creación de informes y consultas.
Las consultas analíticas que resumen grandes volúmenes de datos influyen negativamente en la capacidad de respuesta a las transacciones en línea.	Ordenan los datos en estructuras simplificadas que buscan la eficiencia de las consultas analíticas más que del proceso de transacciones.
El rendimiento del sistema al responder consultas analíticas complejas puede ser lento o impredecible, lo que conlleva a un servicio no eficiente a los usuarios del proceso analítico en línea.	Comprenden datos transformados que son válidos, coherentes, consolidados y con el formato adecuado que permite el análisis sin interferir en la operatoria transaccional diaria.

Los datos que se modifican con frecuencia afectan la coherencia de la información analítica.	Ofrecen datos estables que representan el historial de la organización. Se actualizan periódicamente con datos adicionales, no con transacciones frecuentes.
La seguridad se ve afectada al combinar el análisis en línea con el proceso de transacciones en línea.	Simplifican los requerimientos de seguridad.

Fuente: Nader, J. (2004). Sistema de Apoyo Gerencial Universitario (Magister). Instituto Tecnológico Buenos Aires

El mundo empresarial actualmente demanda una mejor vinculación con el cliente para proveer una respuesta más individualizada. Los avances tecnológicos han permitido que de cierta forma se conozca más de cerca las necesidades de los clientes y se maneje de mejor manera la información dentro de la organización. El data warehouse precisamente ayuda en este aspecto ya que permite almacenar información detallada, procedente de datos internos y externos de la organización, brindando algunos beneficios como reducción de costos, mejora en el modelo de comportamiento de la organización, así como fomentar los procesos de cambio.

A continuación, se explica acerca de una versión especial de los data warehouses denominada data marts.

2.2.2.1 Introducción a data mart

Como se vio anteriormente un data warehouse provee toda la información útil para apoyar a los usuarios en la toma de decisiones, integrando los datos transaccionales obtenidos de fuentes internas y externas de la empresa. Partiendo de esta premisa, es importante considerar que hay información delicada que se puede encontrar dentro del data warehouse que no es necesario que acceda cualquier usuario, por ello es conveniente restringir cierto subconjunto de datos a usuarios determinados; en estos casos se utilizan los data marts.

Por lo general, los data warehouses ofrecen una fuente de datos única para todas las actividades de apoyo en la toma de decisiones. Sin embargo, al

popularizarse fue evidente que frecuentemente los usuarios realizaban grandes operaciones de informes y análisis de datos sobre un subconjunto parcialmente pequeño de todo el data warehouse. De igual manera, era muy posible que los usuarios repitieran las mismas operaciones sobre el mismo subconjunto de datos cada vez que era actualizado.

De manera obvia, la ejecución repetida de tales operaciones sobre el mismo subconjunto de todo el almacén no era muy eficiente; por lo tanto, pareció obviamente una buena idea construir algún tipo de “almacén” limitado de propósito general que estuviera hecho a la medida de ese propósito. Además, en algunos casos sería posible extraer y preparar los datos requeridos directamente a partir de las fuentes locales, lo que proporcionaba un acceso más rápido a los datos que si tuvieran que ser sincronizados con los demás datos cargados en todo el data warehouse. Dichas consideraciones condujeron al concepto de data mart.

De hecho, hay alguna controversia sobre la definición precisa del término data mart. Para nuestros propósitos podemos definirlo como “un almacén de datos especializado, orientado a un tema, integrado, volátil y variante en el tiempo para apoyar un subconjunto específico de decisiones de administración” (Date, 2001, p.710). La principal diferencia entre un data mart y un data warehouse es que el data mart es especializado y volátil. Por especializado se refiere a que contiene datos para dar apoyo únicamente a un área específica de análisis de negocios; por otro lado, volátil quiere decir que los usuarios pueden actualizar los datos e incluso, crear nuevos datos.

Está claro que un data mart es una versión especial del data warehouse, ya que al igual, contienen una visión de los datos transaccionales que al procesarse se convierten en información útil para el apoyo en la gestión de la organización, basados en el análisis de la información histórica. La principal diferencia es que un data mart responde a una necesidad específica, enfatizando el fácil acceso a una información relevante. El objetivo de un data mart por tanto es responder a las necesidades de un determinado departamento dentro de la organización, por lo que podría definirse como un data warehouse departamental.

En la actualidad existen varios enfoques para la creación de un data mart los cuales dan las pautas para saber cuándo es necesario utilizarlos; en el libro *Introducción a los sistemas de bases de datos*, Christopher J. Date (2001) afirma:

- Los datos pueden ser extraídos del data warehouse; con un enfoque de “divide y vencerás” sobre la carga de trabajo general de apoyo para la toma de decisiones, obteniendo un mejor rendimiento y escalabilidad. Por lo general, los datos extraídos son cargados a una base de datos con un esquema físico parecido al subconjunto aplicable del data warehouse; sin embargo, por la naturaleza especializada del data warehouse se lo puede simplificar.
- Aparte de que un data warehouse pretende ofrecer un “punto de control único”, un data mart puede crearse de forma independiente. Dicho enfoque se puede aplicar si no se tiene acceso al data warehouse por cualquier razón operacional, financiera o política.
- En ciertas organizaciones se crean los data marts acorde a las necesidades de la organización, obteniendo al final el data warehouse uniendo varios data marts. (p.710)

En conclusión, un data mart es un subconjunto de data warehouses enfocados en un tema o un área de negocio específico dentro de una organización. A continuación, se especifica la arquitectura de los data warehouses.

2.2.2.2 Arquitectura data warehouse

Una arquitectura de data warehouse es una manera de representar la estructura general de los datos, la comunicación, los procesos y la presentación al usuario final. El término data warehouse se utiliza para hablar en sí de los componentes que lo conforman, específicamente al que se refiere con el almacenamiento físico de los datos.

Un data warehouse incluye varios elementos que se integran de forma conjunta para hacerlo operativo, Luna y García (2006) afirman que estos elementos son los siguientes:

1. Datos operacionales: se refiere a la capa de origen o fuente de datos para poblar el componente de data warehouse. Los datos se obtienen de los sistemas transaccionales que pueden ser internos como externos a la empresa dependiendo de las necesidades de análisis de información.

2. Extracción de Datos: consiste en cargar los datos desde los sistemas transaccionales al data warehouse. El almacenamiento de datos fusiona datos provenientes de distintas fuentes y formatos. La extracción convierte los datos en un formato preparado para iniciar el proceso de transformación.

3. Transformación de datos: es el proceso de aplicar reglas de negocio a los datos operacionales para cubrir las necesidades de análisis y que la integración entre las distintas fuentes sea óptima.

4. Carga de Datos: se realiza la introducción de los datos ya adaptados al formato requerido dentro del sistema destino en este caso el data warehouse; a medida que se van guardando los datos se crea un historial que permite consultas retrospectivas en el tiempo.

5. Data warehouse: almacenamiento físico de datos.

6. Servidor OLAP: Los datos son almacenados en uno o más servidores OLAP⁴, los cuales permiten presentar vistas multidimensionales de los datos hacia herramientas de análisis, consulta, data mining y generación de informes, quienes se encargan de dar el formato adecuado según las necesidades del usuario.

7. Repositorio de meta datos: contiene la información sobre los datos que posea una empresa o que utilizan los sistemas transaccionales. Esta información se refiere a la estructura, significado y procedencia de los datos.

⁴ OLAP (Procesamiento analítico en línea) es una tecnología que permite agilizar las consultas de grandes cantidades de datos.

8. Herramientas front-end: manipulan los datos dependiendo de las necesidades del usuario para realizar análisis, consulta, generación de informes o data Mining. (p.4)

A continuación, se presenta una figura con la arquitectura básica de un data warehouse donde se muestra como los distintos componentes interactúan.

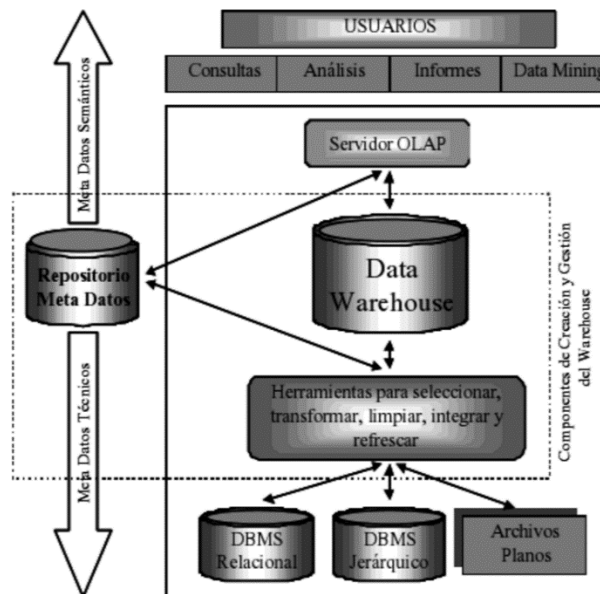


Ilustración 1. Estructura de un data warehouse. Fuente: Luna, M. E., & García, M. F. (2006). El repositorio de metadatos en un data warehouse. Red Revista de Facultad de Ingeniería. Recuperado de <http://www.ebrary.com>

2.2.2.3 Introducción al procesamiento analítico en línea (OLAP)

Por lo general, la información que se quiere utilizar para el almacenamiento en el data warehouse acerca de cierto dominio de la organización, se encuentra en base de datos y otras fuentes tanto internas como externas. Muchas de estas fuentes son las que se utilizan en el diario trabajo al momento de realizar transacciones (conocido como OLTP, Online Transactional Processing). De estas mismas bases de datos de trabajo se puede extraer conocimiento, mediante la utilización de la tecnología de OLAP (Online Analytical Processing). Esta tecnología permite que el data warehouse proporcione respuestas más rápidas a consultas analíticas complejas y repetitivas utilizadas generalmente para sistemas

de soporte para toma de decisiones. Antes de continuar, es necesario introducir una definición formal de OLAP.

Se entiende por OLAP, o proceso analítico en línea, al método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o jerarquía de objetos como en un sistema u organización multidimensional, y cuyo objetivo es recuperar y manipular datos y combinaciones de los mismos a través de consultas o incluso informes. (Díaz, 2001, p.96)

Hay dos funciones de suma importancia que proporcionan los OLAP, la primera se refiere a que OLAP utiliza un modelo de datos intuitivo y natural para la presentación de los datos. Esto permite que los usuarios no se preocupen del proceso en sí, si no de la interpretación que le puedan dar a la información que obtienen. De esta manera las organizaciones pueden reconocer mejor el valor de sus datos. La segunda función, OLAP acelera las respuestas de información debido a que utiliza estructuras denominadas cubos multidimensionales que simplifica a los usuarios las consultas complejas. La combinación de estas dos funciones les permite a los usuarios visualizar y analizar la información más rápida y efectivamente a diferencia de utilizar únicamente sistemas de bases de datos relacionales. Como resultado final, el usuario invierte más tiempo en analizar e interpretar la información que preocuparse por las bases de datos.

2.2.2.3.1 Elementos OLAP

OLAP permite el análisis multidimensional. Lo cual implica que la información está organizada en ejes que son los puntos de vista de análisis y las celdas que se refiere a los valores que se van a analizar. Para el correcto funcionamiento de OLAP, existen diferentes elementos, los cuales se detallan a continuación:

- **Esquema:** es una colección de cubos, dimensiones, tablas de hechos y roles.

- **Cubo:** es una colección de dimensiones asociadas a una tabla de hecho. Un cubo permite intersectar información entre las tablas de hechos cuando las dimensiones son comunes.
- **Tabla de hecho:** dimensión y métrica.
- **Jerarquía:** es la agrupación de miembros estructurados por niveles.
- **Nivel:** es un grupo de datos que tienen los mismos atributos y nivel de profundidad dentro de la jerarquía.
- **Miembro:** es un punto de dimensión de un cubo que pertenece a cierto nivel dentro de una jerarquía. Las métricas llamadas también medidas se consideran un tipo especial de miembro que pertenecen a su mismo tipo de dimensión.
- **Roles:** son los permisos que se dan a un grupo de usuarios.

MDX: es un acrónimo de Multidimensional eXpression. Es el lenguaje de consultas de estructuras OLAP.

2.2.2.3.2 Características de OLAP

Una herramienta OLAP consta de dos partes, un motor y un visor. El motor es el encargado de organizar los datos sobre un objeto como un sistema multidimensional, y su objetivo es recuperar y manipular datos y combinaciones de los mismos a través de consultas. El visor OLAP viene a ser la interfaz que muestra la información a consultar, manipular, reordenar y filtrar, dependiendo de lo que el usuario necesite. Las estructuras OLAP permiten realizar consultas multidimensionales, lo que sería muy complejo hacerlo mediante SQL.

Cuando se utiliza un modelo de datos OLAP, la información se representa como cubos, quienes contienen categorías descriptivas que se denominan dimensiones y valores cuantitativos llamados medidas. Este sistema multidimensional permite simplificar las consultas, filtrar o seccionar los datos dependiendo de las necesidades del usuario. La funcionalidad principal

del modelo OLAP es que los datos se pueden organizar en una jerarquía que representa el nivel de detalle de los mismos dentro de cada dimensión.

OLAP al ofrecer una visión multidimensional de los datos en un concepto de dimensión, permite la visualización de patrones de interés que se evaluarán por separado. De esta manera, no se necesita del concepto de relaciones para la integración y unificación de los datos y además se evitan las filas y columnas. La siguiente figura es un ejemplo de un cubo multidimensional, el cual muestra la distribución de los patrones de interés sin la utilización de filas o columnas.

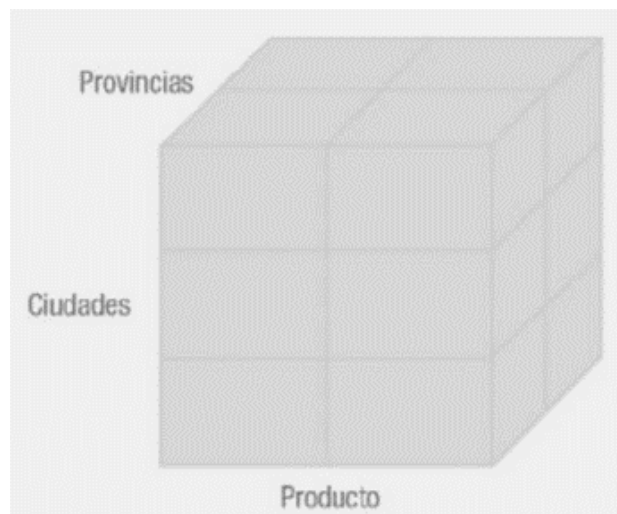


Ilustración 2. Cubo dimensional. Adaptado de Bases de datos, por Reinoso, E. J., Maldonado, C. A., & Muñoz, R., 2012, México, D.F., MX: Alfaomega Grupo Editor, p.163. Recuperado de: <http://www.ebrary.com>

En la figura anterior, la intersección entre las dimensiones genera un valor almacenado que sirve para la asociación de los parámetros presentados por cada dimensión. Entonces, si una dimensión fuese “Productos” y la otra “Ciudades”, en la intersección se almacenaría la cantidad de productos adquiridos por ciudad. Esto contribuye a que el usuario obtenga una visión analítica de la información, lo que facilita el procesamiento porque no debe recurrir a preguntas complejas y permite visualizar mejor las relaciones entre las diferentes tablas.

La funcionalidad de OLAP se caracteriza por permitir realizar un análisis multidimensional de los datos, además de la posibilidad de navegación y

selección de la información que se necesite. La selección de información se ve reflejada en la visualización del cubo multidimensional, los cuales permiten seleccionar el nivel de granularidad de la dimensión o elegir un dato en concreto, pudiendo con ello realizar las siguientes acciones:

- **Rotar:** Consiste en intercambiar las filas por columnas, es decir las dimensiones que se desea analizar.
- **Bajar:** Se disminuye la jerarquía, es decir el nivel de visualización de los datos.
- **Detallar:** Se informa sobre una fila específica a un nivel inferior.
- **Expandir:** Se muestra el dato previo al igual que la información a nivel superior para este y el resto de datos.

Sin duda alguna el OLAP es una parte importante para el cumplimiento de los objetivos que lleva a cabo el data warehouse, ya que esta tecnología abre paso a gran variedad de aplicaciones que van desde reportes dinámicos hasta soporte avanzado para toma de decisiones. Al entender los elementos y características de OLAP, es importante mencionar las distintas implementaciones de OLAP que se detallan a continuación.

2.2.2.3.3 Implementación del OLAP

La implementación OLAP trata de cómo se almacena la información para poder realizar un análisis multidimensional. Aunque para el usuario no debería importarle la manera en cómo se almacena, duplica y oculta la información; es fundamental mencionar que la arquitectura OLAP afecta en gran medida lo que se puede analizar y la forma de hacerlo. Existen diferentes tipos de OLAP, que especialmente se diferencian en la forma en cómo guardan los datos, los cuales se detallan a continuación:

- **MOLAP (Multidimensional OLAP):** forma clásica de OLAP, el cual hace uso de estructuras de bases de datos por lo general optimizadas para recuperar los datos. Estas estructuras se las conoce como bases de

datos multidimensionales. Para la utilización, se crea un fichero donde se encuentra todas las posibles consultas pre calculadas; se diferencian de las bases de datos relacionales, en que la forma de almacenar la información está optimizada para la velocidad de cálculo. Las dimensiones de cada cubo son por lo general atributos como el periodo, el producto, la localización o departamentos de una organización.

- **ROLAP (Relational OLAP):** trabajan con bases de datos relacionales, de manera que se almacenan los datos y tablas dimensionales en forma de tablas relacionales; se crea vistas multidimensionales para la presentación de la información.
- **HOLAP (Hybrid OLAP):** permite un análisis híbrido de la información, es decir que utiliza lo más relevante de cada tipo OLAP mencionados anteriormente. Es un tipo de bases de datos en donde los datos se almacenan de forma relacional y multidimensional. Por ejemplo, las tablas relacionales se las utiliza para guardar cantidad grandes de datos detallados y utilizan el almacenaje multidimensional cuando se trata de cantidades más pequeñas de datos menos detallados o agregados. Esto permite que se necesite menos espacio en disco que en las bases de datos relacionales y además la respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas. El inconveniente surge por el hecho de que los datos multidimensionales deben primero ser cargados antes de realizar una consulta y refrescados cuando se actualizan los datos de la organización.
- **DOLAP (Desktop OLAP):** es un caso particular que está orientado a equipos de escritorio. Consiste en extraer la información de bases de datos relacionales y guardarla en el escritorio. La consulta y el análisis se realiza sobre los datos guardados en el escritorio.
- **In-memory OLAP:** consiste en que la construcción de la estructura dimensional se la haga solo en memoria y se guarde el dato original en

un formato que potencia su visualización de esta manera por ejemplo comprimiendo el dato o mediante bases de datos lógicas asociativas.

Todos estos tipos se inclinan a la explosión de la base de datos, esto significa que se exceda en la cantidad de espacio de almacenaje cuando se presentan ciertas condiciones como por ejemplo un alto número de dimensiones, resultados calculados de antemano y escasos de datos multidimensionales.

La dificultad de implementar OLAP viene de la complejidad de las consultas, la elección de los datos base y el desarrollo del esquema. Es por ello que muchas suites actuales de Inteligencia de Negocios (BI) traen implementadas bibliotecas con enormes consultas pre configurados. Además, se debe tener en cuenta que los datos deben ser completos y constantes.

2.2.2.3.4 Modelo Multidimensional

Un data warehouse se apoya en el modelamiento de bases de datos multidimensionales, el cual permite ofrecer una visión de la operación del negocio. La idea principal es que el usuario pueda ver claramente la relación entre los distintos elementos del modelo.

Tomando como ejemplo un punto en el espacio definido por los ejes x, y, z con sus respectivos valores. El eje x representa una entidad, el eje y representa un indicador y el eje z corresponde al tiempo. Se podría tener la siguiente combinación: entidad = 'Sucursal Cayambe', indicador = 'Total de ventas', tiempo = 'diciembre-2015'. La intersección de estos tres valores definirá el punto en el espacio, que se traduce como *la cantidad de ventas que tuvo la sucursal de Cayambe en diciembre del 2015* en este caso.

En el modelo multidimensional cada eje corresponde a una dimensión específica. Por lo tanto, la dimensión de los cubos OLAP estará representada por la cantidad de ejes que se le asocie.

2.2.2.3.4.1 Estructura del modelo multidimensional

La estructura básica de un modelo multidimensional está conformada por dos elementos: las tablas y los esquemas. Es posible encontrar dos tipos de tablas:

- **Tabla de hechos:** es el componente básico de un data warehouse y hace referencia a lo que se desea medir a partir de los datos presentes. Es la tabla central de un esquema multidimensional y en ella se guardan las mediciones numéricas del negocio. Las medidas se hacen sobre granularidad de la tabla que está determinada por el nivel de detalle que se almacenará en la tabla. Por ejemplo, se tiene las dimensiones producto, mercado y tiempo, la granularidad puede ser la cantidad de productos vendidos a la semana. Cada medición es tomada de la intersección de las dimensiones. La clave de la tabla de hechos se la denomina clave compuesta o concatenada porque se forma de la composición de las llaves primarias de las tablas dimensiones unidas a ella. Por lo tanto, se distinguen dos tipos de columnas en la tabla de hechos: columnas fact y columnas key. La columna fact guarda una medida de negocio específica y la columna key es la clave compuesta.
- **Tabla de dimensiones:** contienen los valores asociados a la tabla de hechos. Una tabla de dimensiones está compuesta por una llave primaria que identifica una fila en la tabla junto con un conjunto de atributos, y dependiendo del diseño del modelo pueden existir llaves foráneas que refleja su relación con otra tabla de dimensión. Para decidir si un campo es un atributo o un hecho es importante analizar la variación de la medida en el tiempo. Si varía continuamente significa que se debe tomar como un hecho, caso contrario como un atributo.

Los esquemas son la colección de tablas en el data warehouse. Tenemos dos categorías básicas de esquemas:

- **Esquema estrella:** El esquema Estrella es una técnica de modelamiento de datos para reflejar una base de datos relacional en un modelo multidimensional. El diseño estructural consiste en una tabla central (tabla de hechos) y un conjunto de tablas de dimensión, como se visualiza en la siguiente imagen.

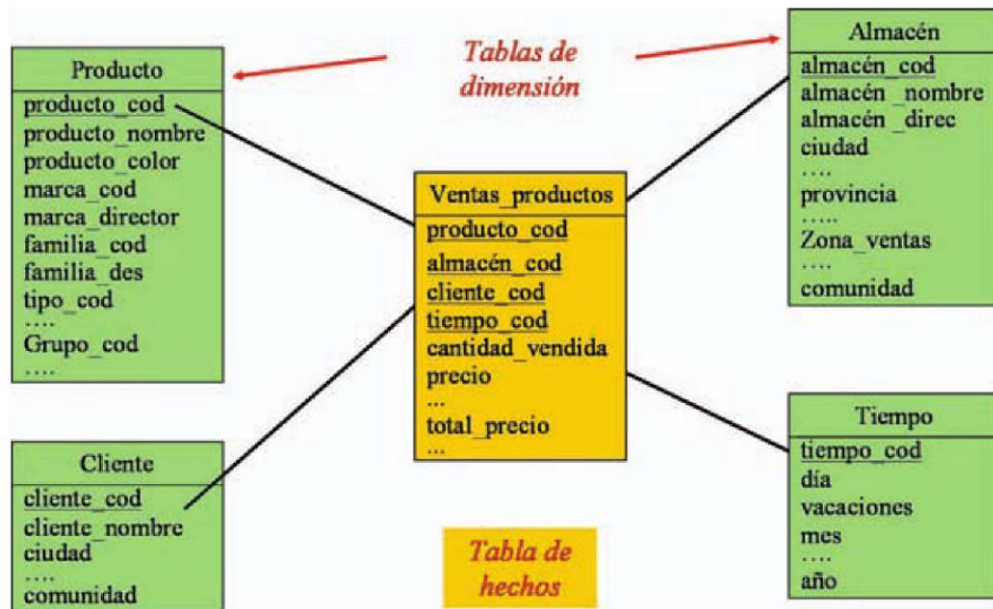


Ilustración 3. Ejemplo de esquema estrella. Adaptado de Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional, por Trujillo, J. C., 2013, Alicante, ES: ECU., p.35. Recuperado de: <http://www.ebrary.com>

Las características principales de este esquema son:

- Cada tabla de dimensión posee una clave primaria no compuesta.
- La clave primaria de la tabla de hechos está compuesta por las claves de las tablas de dimensiones, lo que significa que la tabla de hechos tiene una relación muchos a muchos con las tablas de dimensiones.

Es importante mencionar que la tabla de hechos guarda datos que se crearon por eventos que ocurrieron en el pasado y que no cambiarán con el tiempo. En cambio, las tablas de dimensiones almacenan información

descriptiva que puede estar relacionada con varias tablas de hechos y que se utiliza para poder delimitar las consultas de la información de la tabla de hechos. Es por ello, que las tablas de hecho ocupan más espacio que las tablas de dimensión.

Una característica relevante a tomar en cuenta cuando se diseña un modelo multidimensional es la denominada granularidad. La granularidad está vinculada a la cantidad de dimensiones que definen el alcance de la tabla de hechos y el nivel de detalle en que se guardan los datos. Por ejemplo, resúmenes diarios, mensuales, anuales o cualquier otro tipo de periodo de tiempo. La granularidad afecta el tamaño del repositorio y el grado de análisis y flexibilidad de los datos, ya que cuanto más granulares sean los datos mayor espacio de almacenamiento se necesitará, aunque con ello se tendrá la ventaja de poder realizar un análisis con mayor nivel de detalle de los mismos.

Las tablas de dimensión contienen datos alfanuméricos que se encuentran almacenados en forma desnormalizada, lo que significa que existe mucha redundancia de datos. En definitiva, cada tabla de dimensión corresponde a una sola tabla, como se evidencia en la siguiente figura:

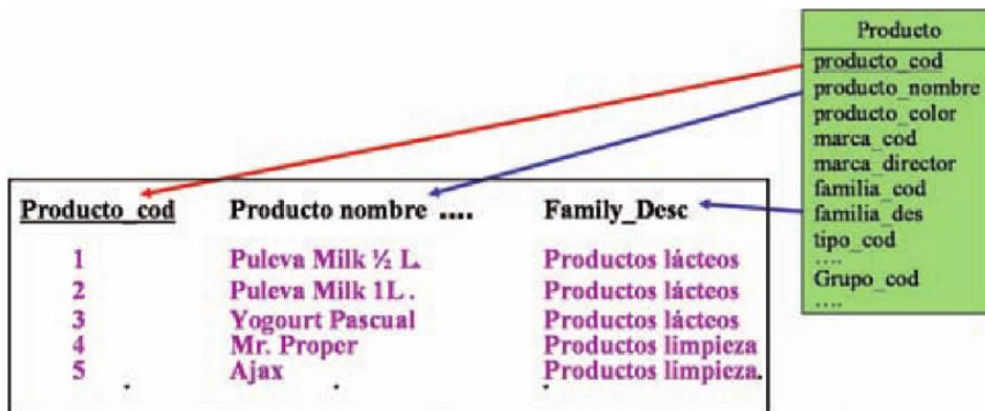


Ilustración 4. Ejemplo de la tabla dimensión producto (incluyendo instancias). Adaptado de Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional, por Trujillo, J. C., 2013, Alicante, ES: ECU., p.36. Recuperado de: <http://www.ebrary.com>

La desnormalización permite que el contenido de varias tablas se una en una sola con el propósito de aumentar el rendimiento de las consultas, ya que se elimina el número de relaciones entre las tablas y es más fácil consultar por parte del usuario no experto a expensas de incrementar la redundancia de datos.

- **Esquema copo de nieve:** El esquema de copo de nieve se diferencia del esquema de estrella en la estructura de algunas tablas de dimensiones, ya que se normalizan para mejorar el rendimiento y la utilización del espacio de almacenamiento. Una ventaja de la utilización de este esquema es que la consulta de atributos simples es más rápida debido a que se recorre tablas con menor número de instancias. Referente a inconvenientes es evidente que al aumentar el número de tablas aumenta el número de operaciones de unión que se deben realizar, por lo que algunas consultas que implican más de una tabla pueden tomar más tiempo. En cuanto a la estructura resulta complejo el diseño y mantenimiento del modelo multidimensional, ya que se requiere una clave primaria más por cada nivel de jerarquía normalizado.

Este tipo de esquemas es recomendable cuando el espacio de almacenamiento es un problema. Sin embargo, hay que tomar en cuenta que se debe normalizar únicamente las tablas de dimensiones que sean demasiado grandes. A continuación, se presenta un ejemplo de normalización de la dimensión almacén:

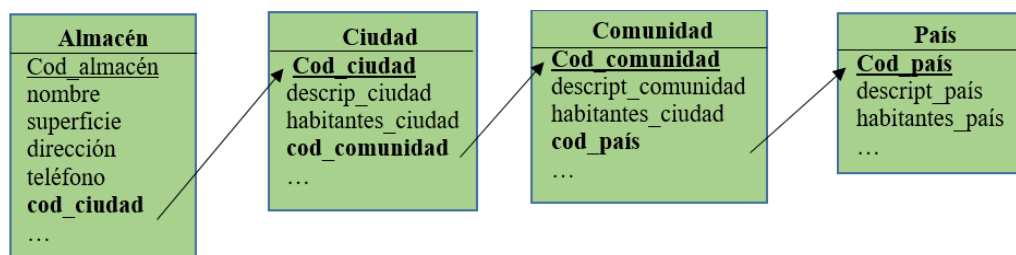


Ilustración 5. Ejemplo de normalización de la dimensión almacén. Adaptado de Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional, por Trujillo, J. C., 2013, Alicante, ES: ECU., p.42. Modificado por: Adrián Carrión y Christopher Torres

2.2.2.4 Introducción al Data Mining

El uso de los sistemas de información transaccionales, la aparición de aplicaciones de internet, el empleo de tecnología de hardware para lectura de datos, el avance de tecnología en almacenamiento de datos, entre otros ha dado lugar a la acumulación de grandes cantidades de datos que deben manejar las organizaciones y esta tendencia continúa a un ritmo acelerado. Sin embargo, todo este tipo de información forma parte de un activo muy importante para una organización, en la medida en que se pueda transformar en conocimiento para la empresa. Y surge la pregunta, ¿cómo manejar grandes volúmenes de datos? La respuesta es precisamente el tema que se explica a continuación.

En el libro “Introducción a los sistemas de bases de datos”, su autor Christopher J. Date (2001) afirma:

Data mining se describe como un “análisis de datos exploratorio”. Con el propósito de encontrar patrones relevantes en los datos, los mismos que pueden ser usados para especificar la estrategia del negocio o para identificar comportamientos extraños. Las herramientas de minería de datos utilizan técnicas estadísticas a una gran cantidad de datos almacenados para buscar tales patrones. (p.722)

Las técnicas que se utilizan para el data mining son el resultado de un proceso de investigación amplio. Este desarrollo empezó con los datos operacionales, cuando fueron almacenados por primera vez en computadoras, para continuar con mejoras en el acceso a los datos, y tecnologías para permitir a los usuarios navegar entre los datos en tiempo real. Data mining está soportado por tres tecnologías que con el tiempo han ido madurando: recolección de grandes volúmenes de datos, potencia computacional y los algoritmos para búsqueda y navegación entre los datos que utilizan.

Los elementos esenciales de data mining han estado bajo desarrollo por largo tiempo en áreas de investigación como inteligencia artificial y estadística. Hoy, la madurez de los algoritmos ha permitido que esta tecnología fuera tomando mayor valor dentro de los entornos de data warehouse actuales.

2.2.3 Inteligencia de Negocios (BI)

La Inteligencia de Negocios (BI) en la actualidad ha ido creciendo dentro de las organizaciones que la implementan ya que representa una ventaja competitiva del negocio al transformar los datos disponibles en conocimiento para mediante la toma de decisiones pertinentes posicionarse de mejor manera en su mercado.

2.2.3.1 Evolución de la Inteligencia de Negocios (BI)

Normalmente las personas son las encargadas de la toma de decisiones, tomamos decisiones todos los días, pero muchas veces tomamos malas decisiones. Entonces, la Inteligencia se basa en decisiones, decisiones que son tomadas por personas, si añadimos información con inteligencia se obtendrá una buena decisión.

La información se ha usado a través del tiempo para compartir historias, para aprender mediante libros, para descubrir y hacer un seguimiento de grandes cosas. Con el paso del tiempo la información proveía de respuestas para descifrar problemas, innovar y crear cosas de tamaños inmensamente grandes hasta muy pequeñas. Entonces, la información tiene todo que ver con respuestas que ayudan a la gente a tomar mejores decisiones.

Es importante aclarar cómo la alta gerencia de las organizaciones obtiene sus respuestas, cuentan con gran cantidad de datos ya sean de personas, productos, lugares. Además, también tienen muchas preguntas, qué producto es el mejor, cómo están sus ventas, cómo está trabajando su equipo, antes de que se puedan obtener todas las respuestas que necesitan es necesario que analicen y entiendan los datos con los que disponen, sino son simplemente datos.

Los datos son el recurso más valioso de una organización, es por esto que deben ser almacenados. Antes de la década del 60, los datos eran almacenados en anaqueles o gabinetes; con la aparición de las primeras computadoras los datos empezaron a guardarse en cintas magnéticas y su acceso era secuencial. El problema de este tipo de acceso es que se debe recorrer todos los registros almacenados antes de poder encontrar el registro que se está buscando. Además, a medida que crecía la cantidad de datos operacionales, se iba acumulando una

enorme cantidad de cintas y medios de almacenamiento que hacía difícil su administración.

En 1969, Edgar Codd inventó la base de datos, cuyo objetivo era almacenar datos. Sin embargo, esto no era tan fácil se necesitaba de un nivel de experticia, además que se necesitaba entender el negocio para que estos datos puedan ser ingresados.

En los 70, se crearon las primeras aplicaciones para gestión empresarial, que brindaban una mejor manera de recolectar los datos de la organización y proveían una mayor facilidad para almacenarlos, los datos juntos con el acceso a ellos proveen respuestas. Sin embargo, en los comienzos el acceso a los datos no era sencillo, ya que los datos almacenados en la base de datos venían de diferentes aplicaciones de gestión empresarial. Se intentaba acceder a los datos uno por uno mediante reportes, pero estos eran unidimensionales, y ofrecían acceso a los datos de forma fragmentada en diferentes sitios. Se necesitaba un acceso más efectivo a través de varios de estos sitios, así que los datos fueron movidos a un data warehouse.

A inicios de los 80, los data warehouses fueron popularizados por Bill Inmon y Ralph Kimball, ahora los datos podían ser organizados y podían venir de diferentes sitios para ser administrados y accedidos, entonces los datos podían ser consumidos.

En 1989, Howard Dresner acuña el término Business Intelligence (BI), es ahora cuando aparece Business Intelligence 1.0, las herramientas para Inteligencia de Negocios (BI) permiten obtener reportes y analizar los datos, entonces las diferentes marcas que ofrecían aplicaciones para Inteligencia de Negocios (BI) ofrecían mayor acceso a través de más sitios. Sin embargo, las organizaciones necesitaban los datos de manera más rápida, entonces Inteligencia de Negocios (BI) se trataba de desempeño.

Con la aparición de WWW (World Wide Web), todo empezó a moverse de manera mucho más rápida y los datos seguían aumentando, esto significaba que los

negocios necesitaban respuestas de manera más rápida. En los 90, las herramientas de Inteligencia de Negocios (BI) crecieron notablemente, había herramientas de gestión de información, de informes de consulta y análisis, de gestión de rendimiento y de visualización, pero no eran las mejores opciones a pesar de sus altos costos de adquisición y de mantenimiento, así que los vendedores necesitaban una manera de ofrecer mayor funcionalidad a menos costo, aquí aparecen las plataformas de Inteligencia de Negocios (BI).

Ahora aparece Business Intelligence 2.0, en donde hay muchas más herramientas con mayor funcionalidad, que permiten el acceso online y también offline, convirtiéndose en una prioridad para las organizaciones. A pesar de esto la alta gerencia en las organizaciones no terminaban de entender completamente qué era Inteligencia de Negocios (BI), ya que aún no obtenían las respuestas que necesitan y esperaban. Ya que las herramientas no eran intuitivas y no eran datos puntuales.

Inteligencia de Negocios (BU) se trataba de usabilidad para transformar los datos en el formato correcto para que pueda ser fácilmente accedido y consumido por las personas, además de ser fácilmente administrable por el personal de IT.

La Inteligencia de Negocios (BI) propone una relación entre las personas encargadas de tomar decisiones y las herramientas que se emplean; es por ello que la tecnología está claramente vinculada al análisis de la información de las organizaciones permitiendo obtener ventajas competitivas, producto de decisiones mejor tomadas. Según Bitam (2002), “Inteligencia de Negocios como una combinación de tecnología y desarrollo de negocios”.

2.2.3.2 Definición de Inteligencia de Negocios

Dentro de las muchas definiciones que existen para Inteligencia de Negocios (BI), la primera definición formal propuesta por Howard Dresden (1989) dice: “la Inteligencia de Negocios son conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos”.

La definición más acertada y entendible según Luis Méndez Del Río (2006) afirma:

“La Inteligencia de Negocios (BI) es un conjunto de herramientas y aplicaciones para la ayuda a la toma de decisiones que posibilitan acceso interactivo, análisis y manipulación de información corporativa de misión crítica. Estas aplicaciones contribuyen a un conocimiento valioso sobre la información operativa identificando problemas y oportunidades de negocio. Con estas, los usuarios son capaces de tener acceso a grandes volúmenes de información para establecer y analizar relaciones y comprender tendencias que posteriormente soportarán decisiones de negocios.” (p.24)

Nader (2004) dice también lo siguiente: “Es el conjunto de tecnologías que permiten a las empresas utilizar la información disponible en cualquier parte de la organización para hacer mejores análisis, descubrir nuevas oportunidades y tomar decisiones más informadas” (p.52).

Díaz (2010) afirma: “La Inteligencia de Negocios (BI) es el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización” (p.18).

Luego de analizar las definiciones anteriores, se puede decir que la Inteligencia de Negocios (BI) es un término que describe al conjunto de metodologías, prácticas, herramientas y aplicaciones cuyo enfoque es crear y administrar información útil para organizar, mejorar y optimizar la toma de decisiones de una organización.

La Inteligencia de Negocios (BI) es además de gran ayuda para conocer la situación actual de la organización y su forma de trabajar, optimizando su capacidad para obtener respuestas de manera eficiente para la prevención y la rápida respuesta a problemáticas que vayan apareciendo. Entonces, La Inteligencia de Negocios (BI) proporciona una ventaja competitiva dentro del

mercado, ya sea en operaciones complejas o simples a partir de la toma de decisiones.

2.2.3.3 Importancia de Inteligencia de Negocios

En la actualidad las organizaciones buscan posicionarse primeros frente a su competencia y buscan la mejor forma de tener una ventaja competitiva es la implementación de Inteligencia de Negocios (BI), mediante la explotación de los datos y de la información que disponen para que los ayude en la toma de decisiones.

Una vez generado el conocimiento a partir la información disponible dentro de la organización, los procesos dentro de la misma se mejoran y optimizan para que sean más eficientes. El tener acceso a la información de manera clara y entendible para la gestión del conocimiento es un factor muy relevante y productivo para la organización.

Edison Medina La Plata (2012) menciona algunos elementos importantes que caracterizan asociados a la Inteligencia de Negocios.

- **Integración y universalización de la información:** Se busca buscar respuestas a necesidades analíticas. Es decir que todos los departamentos de la empresa proporcionan información y de la misma manera son capaces de acceder a ella sin importar de donde sea obtenida.
- **Conversión de datos en información valiosa:** La cantidad de datos que genera una organización puede ser incontable. Sin embargo, lo que se busca es obtener información relevante a partir de dichos datos que sea de utilidad para quien lo requiera.
- **Apoyo en un conjunto de herramientas para el análisis de datos:** Actualmente las herramientas ayudan y optimizan a la extracción, la depuración, el análisis y el almacenamiento de los datos generados en una organización para que se conviertan en información y así apoyen la toma de decisiones.

- **Capacidades de análisis multidimensional:** Esta capacidad permite ver los datos que se necesiten utilizando varias dimensiones dependiendo de su importancia, permitiendo obtener información importante y también un seguimiento de los indicadores más importantes de la organización.



Ilustración 6. Necesidades de Información en la Empresa. Fuente: *Business Intelligence*, una guía práctica. Segunda edición. UPC. Por Edison Medina La Plata (2012) (p.29)

Ahora que se tiene claro la importancia de la Inteligencia de Negocios (BI) para el manejo y gestión de la información disponible en una organización para la toma de decisiones, es importante mencionar los objetivos potenciales según Josep Curto Díaz (2010):

- **Crear un círculo virtuoso de la información:** Trata precisamente sobre el proceso en el que se obtiene información y conocimiento a partir de datos para obtener más datos, y así poder tomar mejores decisiones en la organización.
- **Permitir una visión más a detalle de la información:** Hace referencia a una forma diferente de ver la información, con el fin de obtener el mayor potencial de la misma de una forma única, histórica y sólida.
- **Manejar y mantener métricas dentro de la organización:** Es importante que dentro de la organización existan métricas que básicamente son indicadores ya sean de rendimiento o de metas de todos

los procesos organizacionales, los cuales permiten tener una mejor visión para poder plantearse estrategias para trabajar de la mejor manera.

- **Aportar información actualizada en todos los niveles:** Lo importante es manejar información en tiempo real, esto hace que la misma se encuentre constantemente actualizada y así se podrá trabajar con la misma desde cualquier nivel en la organización.
- **Mejorar la comprensión de los sistemas de información dentro de la organización:** La idea es poner a la disposición la información necesaria que se requiera por los usuarios para que puedan analizarla y gestionarla a través de diferentes tecnologías, y de esta forma ayudar a los usuarios a entender mejor los procesos operativos de la organización.
- **Mejorar la competitividad de la organización en el mercado:** El buen posicionamiento en el mercado es el principal objetivo de las organizaciones. Ya que, al optimizar y agilizar la toma de decisiones en la empresa, se vuelven más ágiles y tolerantes a cambios repentinos con la ayuda de la información que disponen.

Hay que tomar en cuenta que la idea de dichos objetivos ayuda a las organizaciones mediante diferentes herramientas de Inteligencia de Negocios (BI) a establecer objetivos estratégicos, que permiten optimizar y mejorar el rendimiento de la organización.

2.2.3.4 Componentes de Inteligencia de Negocios

En la actualidad las organizaciones tienen acceso a una cantidad inimaginable de datos. Ellas producen, recogen y almacenan grandes volúmenes de datos. La Inteligencia de Negocios (BI) se refiere a una serie de metodologías que pone estos datos a trabajar, ayudando a las empresas a ser más eficaces para que las ganancias aumenten. Mediante el uso específico de metodologías y análisis de software, la alta gerencia puede aprovechar el poder de los datos en bruto y apoyar la planificación estratégica, que puede ayudar a un movimiento de organización por delante de la competencia. Según Oriol Amat Salas y Fernando

Campa (2013) define las principales componentes de Inteligencia de Negocios (BI):

- **Componente de diseño:** Principalmente usado para diseñar el data warehouse. En donde se utilizan tablas agregadas donde se guardan los indicadores de las tablas de detalle a un nivel superior. Por ejemplo, no es necesario calcular el total de ventas de un producto en una determinada zona para un periodo de tiempo determinado porque se podría tener pre calculado las ventas a nivel mensual, o por cliente, o por producto en el caso de ser necesario. Un buen diseño es clave a la hora de facilitar la información necesaria para la toma de decisiones.
- **Componente de carga o ETL (Extract, Transform and Load):** Usado principalmente para la extracción de datos desde los entornos transaccionales (ERP), su transformación para que pueda ser adoptada bajo a los estándares del data warehouse y su posterior carga en este último.
- **Componente de acceso a datos:** Permite a los usuarios finales consultar y analizar la información contenida en el data warehouse. Estas herramientas permiten que el modelo de datos sea entendible y no sea complejo para que el usuario realice consultas de manera sencilla y en lenguaje natural, y así pueda elaborar sus propios informes de manera sencilla y rápida. Según la complejidad de los informes obtenidos, estas herramientas pueden dividirse en:
 - Herramientas para informes predefinidos básicos.
 - Herramientas de análisis multidimensional. También conocidas como cubos OLAP, son una estructura de datos que proporciona un análisis rápido de datos. Los cubos pueden mostrar y sumar grandes cantidades de datos, a la vez que proporcionan a los usuarios acceso mediante búsqueda a los puntos de datos.

- **Herramientas complejas de “data mining”:** Se apoyan en análisis estadísticos para buscar patrones de comportamiento y dependencias ocultas dentro de grandes volúmenes de datos.
- **Data Marts:** Se trata de una partición del data warehouse orientado a un determinado departamento de la organización. Es una forma de segmentar datos y permitir a cada usuario centrarse en su especialidad, optimizando costos con resultados más fáciles.
- **Executive Information Systems (EIS):** Son herramientas de presentación de datos, que presentan cuadros de mando muy visuales que resumen el estado y las tendencias de la organización.

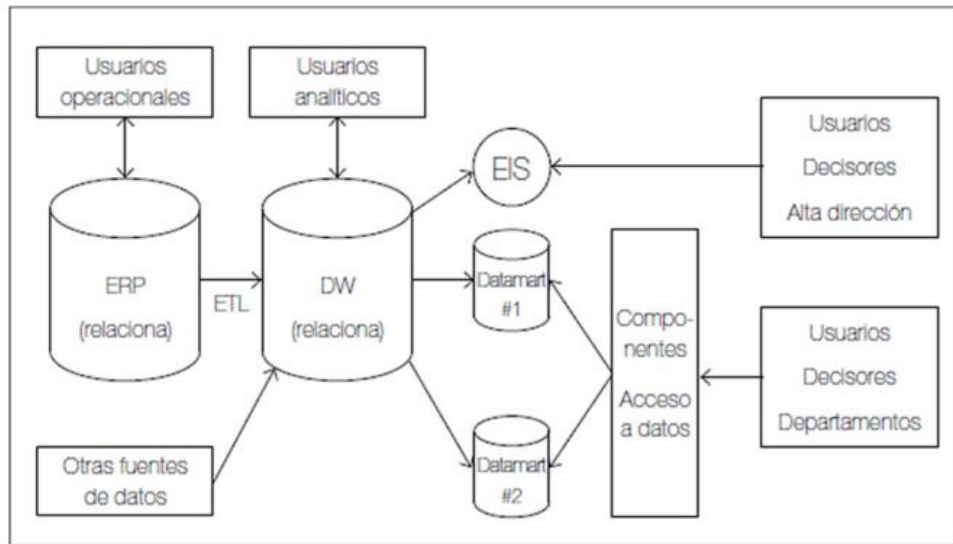


Ilustración 7. Componentes de una arquitectura BI Empresarial madura. Fuente: *Manual del Controller*. Primera Edición. Bresca (Profit Editorial). Por Oriol Amat Salas y Fernando Campa (2013)

Ahora, se explica el proceso de Inteligencia de Negocios (BI) con el fin de entender cómo la organización crea inteligencia para ofrecer a los usuarios acceso a la información. Según Darío Bernabéu en su artículo *Data Warehousing y metodología Hefesto (2009)*, publicado en el sitio web [dataprix](http://dataprix.com)⁵, menciona cinco fases:

⁵ Dataprix – Portal de referencia sobre Tecnologías de la Información.

- **Dirección y planificación:** Es la fase inicial donde se realiza una recolección de requerimientos de información específicos de cada usuario para poder entender las principales necesidades de la organización, y así poder generar las preguntas que ayudarán a alcanzar los objetivos de la misma.
- **Recolección de Información:** Es aquí donde se extraen desde las diferentes fuentes de información internas y externas los datos que servirán para realizar un análisis y encontrar las respuestas a las preguntas que se plantean en la fase anterior.
- **Procesamiento de Datos:** En esta fase se integran y guardan los datos con un mismo formato para que puedan ser analizados y procesados.
- **Análisis y Producción:** Se trabaja con los datos almacenados en el repositorio mediante la utilización de herramientas y técnicas propias para Inteligencia de Negocios (BI), se realiza un análisis y se genera reportes, indicadores de rendimiento, cuadros de mando, gráficos estadísticos, etc. Se obtendrán las respuestas a todas las preguntas de anteriores fases.
- **Difusión:** Finalmente, mediante la utilización de la herramienta que sea necesaria se mostrará a los usuarios la información de forma que les permita explorar los datos de manera sencilla e intuitiva.

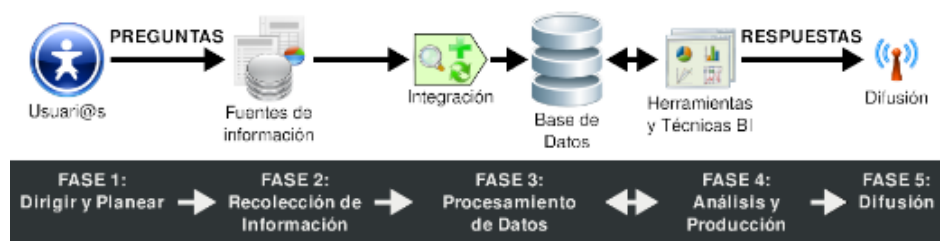


Ilustración 8. Proceso de Business Intelligence. Fuente: Artículo *Data Warehousing y metodología Hefesto* (2009), publicado por el Ingeniero Darío Bernabéu en el sitio web dataprix.com

2.2.3.5 Herramientas para Inteligencia de Negocios (BI)

2.2.3.5.1 Herramientas Propietarias

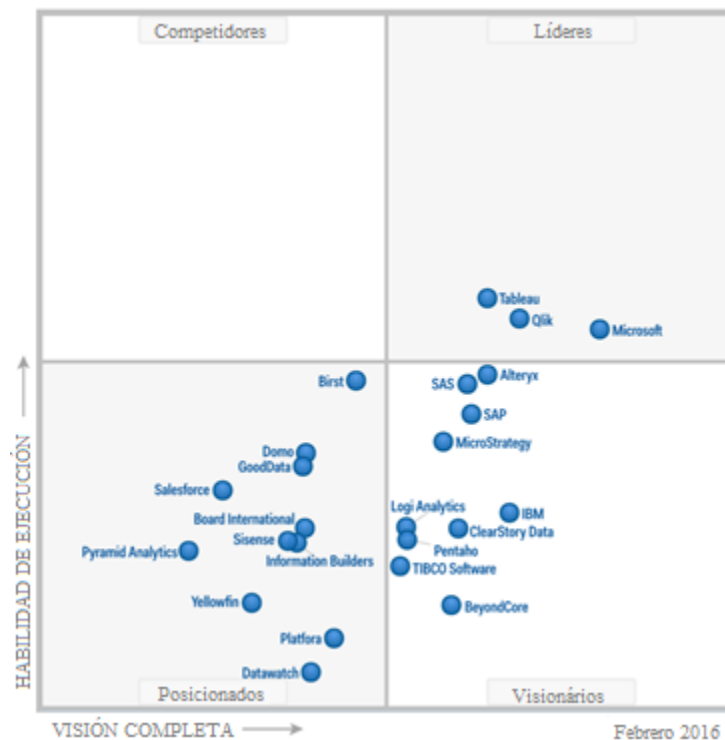


Ilustración 9. Cuadrante Mágico para plataformas de Business Intelligence y analíticas.

Fuente: Investigadora Gartner (2016), Traducción: Christopher Torres y Adrián Carrión

Según la investigadora Gartner⁶ en su informe de investigación “*Magic Quadrant for Business Intelligence and Analytics Platforms*” (2015), podemos ver que las plataformas para Inteligencia de Negocios (BI) que se encuentran delante de su competencia son las propietarias, destacándose Tableau, Qlik, Microsoft, MicroStrategy, SAS.

A continuación, se detalla la arquitectura de las suites para Inteligencia de Negocios (BI) propietarias más importantes según el Cuadrante mágico de Gartner:

⁶Gartner - Empresa consultora y de investigación de tecnologías de la información con sede en Stanford, Estados Unidos.



Tableau ha sido por cuatro años consecutivos líder en el informe Cuadrante Mágico para plataformas de análisis e Inteligencia de Negocios (BI). Entre sus principales productos destaca Tableau Desktop, que permite encontrar patrones visuales con rapidez, combina fuentes de datos sin necesidad de escribir código, permite crear presentaciones interactivas, ofrece análisis de tendencias, entre otras que la hacen destacar y ser una de las herramientas más potentes.

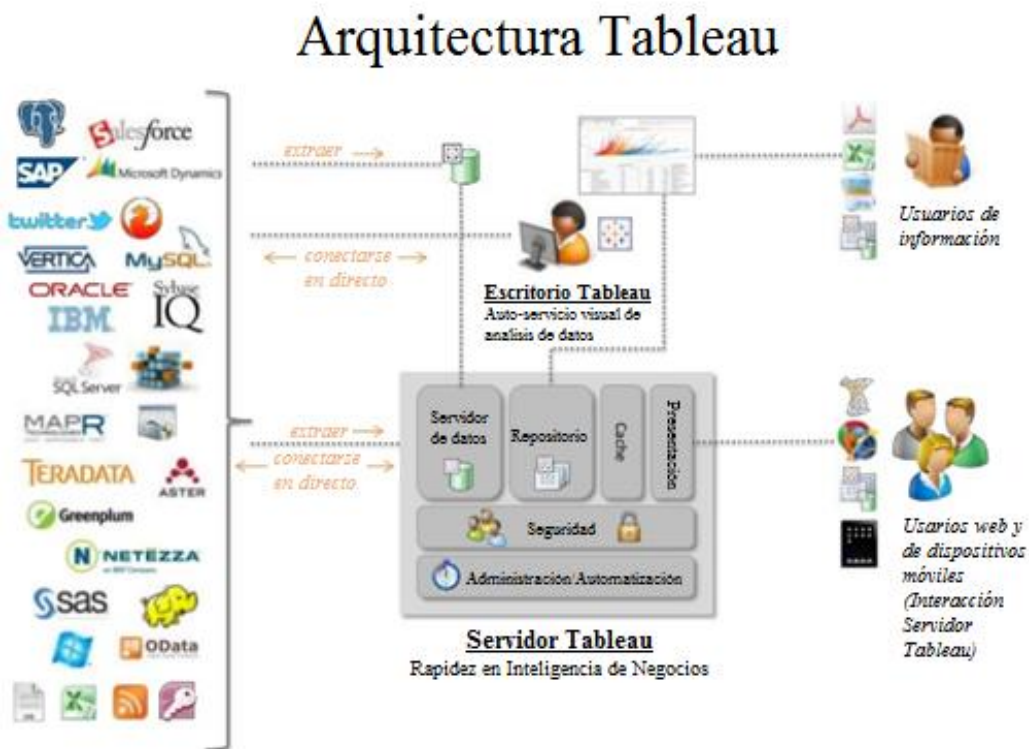


Ilustración 10. Arquitectura BI Tableau. Fuente: *Tableau Solutions Overview* (2015), publicado por BSD Solutions en el sitio web slideshare.net, Traducción: Christopher Torres y Adrián Carrión



Qlik ha sido posicionado durante seis años consecutivos entre los líderes en el informe Cuadrante Mágico para plataformas de análisis e Inteligencia de Negocios (BI). Destaca por tener un producto revolucionario de auto-servicio de descubrimiento y visualización de datos diseñada para individuos, grupos y empresas. Qlik Sense le permite extraer con rapidez visualizaciones, explorar los datos a fondo, revelar conexiones al instante y ver oportunidades desde cualquier ángulo.

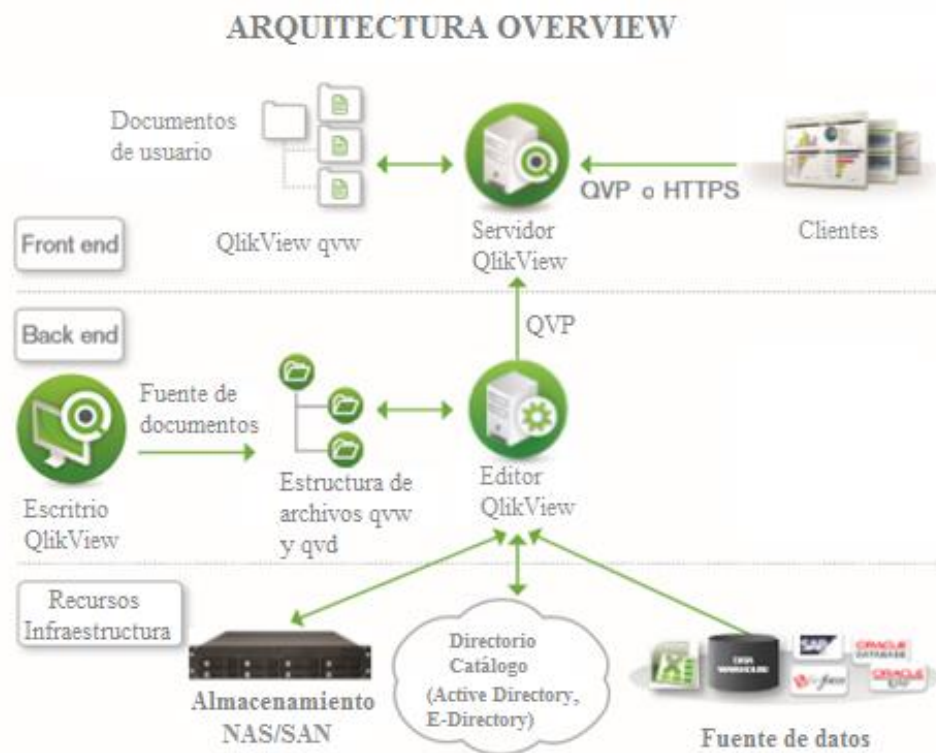


Ilustración 11. Arquitectura BI Qlik. Fuente: *QlikView architectural overview (2014)*, Traducción: Christopher Torres y Adrián Carrión



La herramienta para Inteligencia de Negocios (BI) SAS Visual Analytics, proporciona capacidades de autoservicio de fácil uso y la implementación de análisis en tiempo real directamente en dispositivos móviles y aplicaciones de Microsoft. Esto con la finalidad de ayudar a la alta gerencia a tomar decisiones desde cualquier lugar, Con características tales como gobernabilidad, metadatos centralizados y escalabilidad que la hacen adecuada también para la Tecnología de Información de las organizaciones.

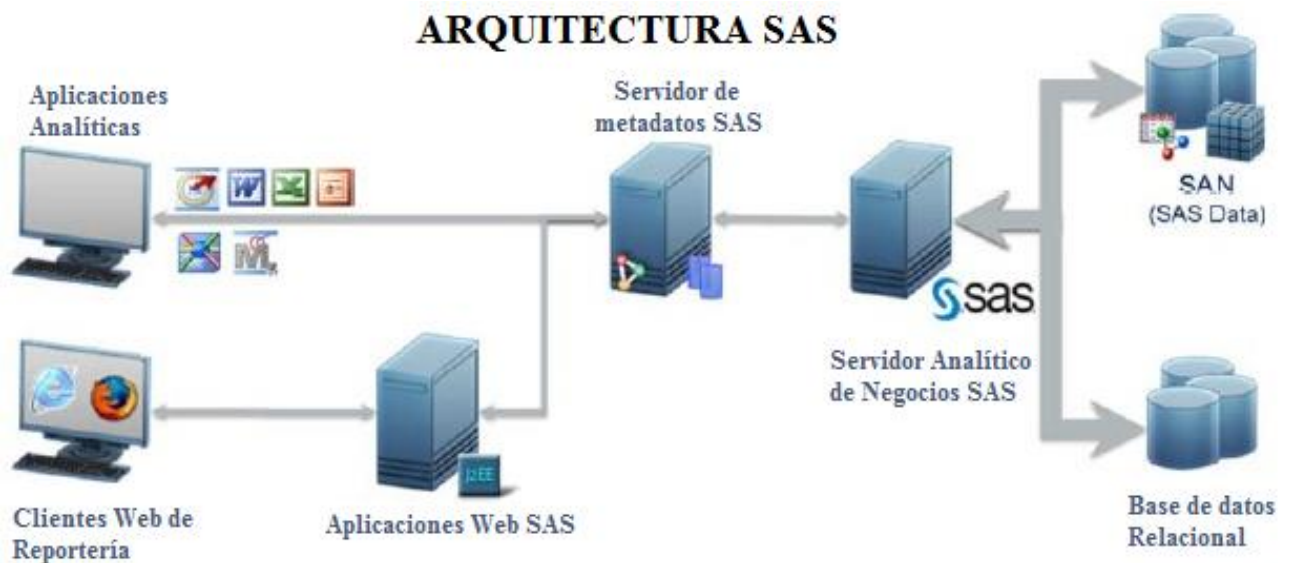


Ilustración 12. Arquitectura BI SAS Intelligence Platform. Fuente: <http://metier.jakarman.nl/saswrk02.html>, Traducción: Christopher Torres y Adrián Carrión

2.2.3.5.2 Herramientas Open Source

En actualidad las herramientas informáticas Open Source son incontables y muchas veces hasta mejores que el software pagado. Dentro de estas también encontramos herramientas para Inteligencia de Negocios (BI), que se encuentran en el “Magic Quadrant for Business Intelligence and Analytics Platforms”.



Es una plataforma compuesta de diferentes programas que satisfacen los distintos requerimientos para Inteligencia de Negocios (BI), ofreciendo soluciones para el tratamiento y análisis de información, que incluye funcionalidades como la construcción de modelamiento multidimensional OLAP, presentación de informes dinámicos, data mining y creación de cuadros de mando para una mejor visualización y acceso fácil para el usuario.

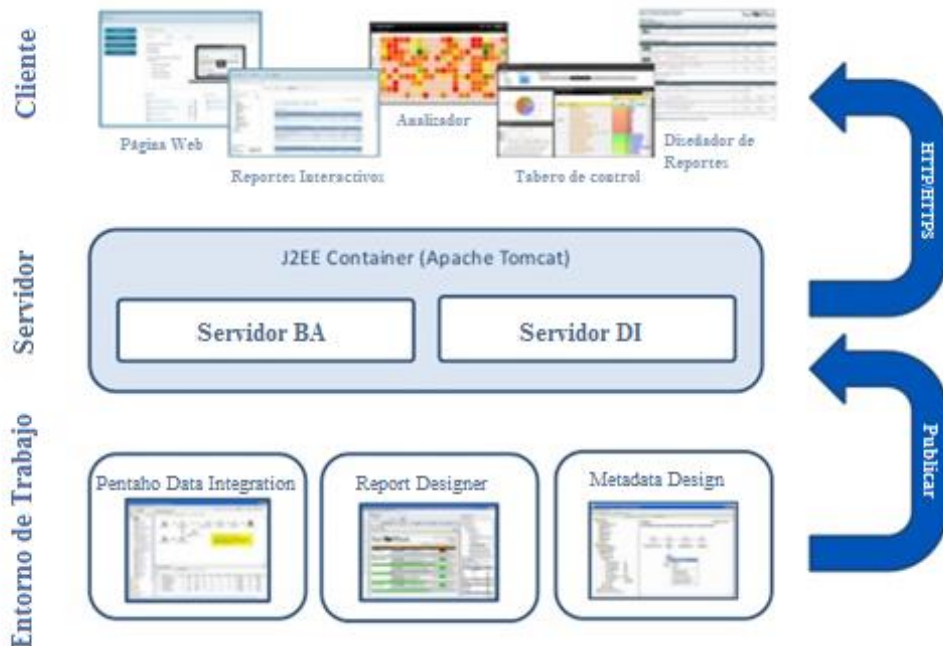


Ilustración 13. Component Architecture. Fuente: *Quick Start Your Pentaho Evaluation* (2013), publicado por Pentaho en el sitio web slideshare.net, Traducción: Christopher Torres y Adrián Carrión

La plataforma de JasperSoft por ser de fácil integración y asequible de informes y análisis, permite a los usuarios crear informes y análisis integrados en su aplicación y de su negocio, además compartir informes y cuadros de mando atractivo, acceso mediante dispositivos móviles, análisis de Big Data, por un valor adicional análisis en la nube.

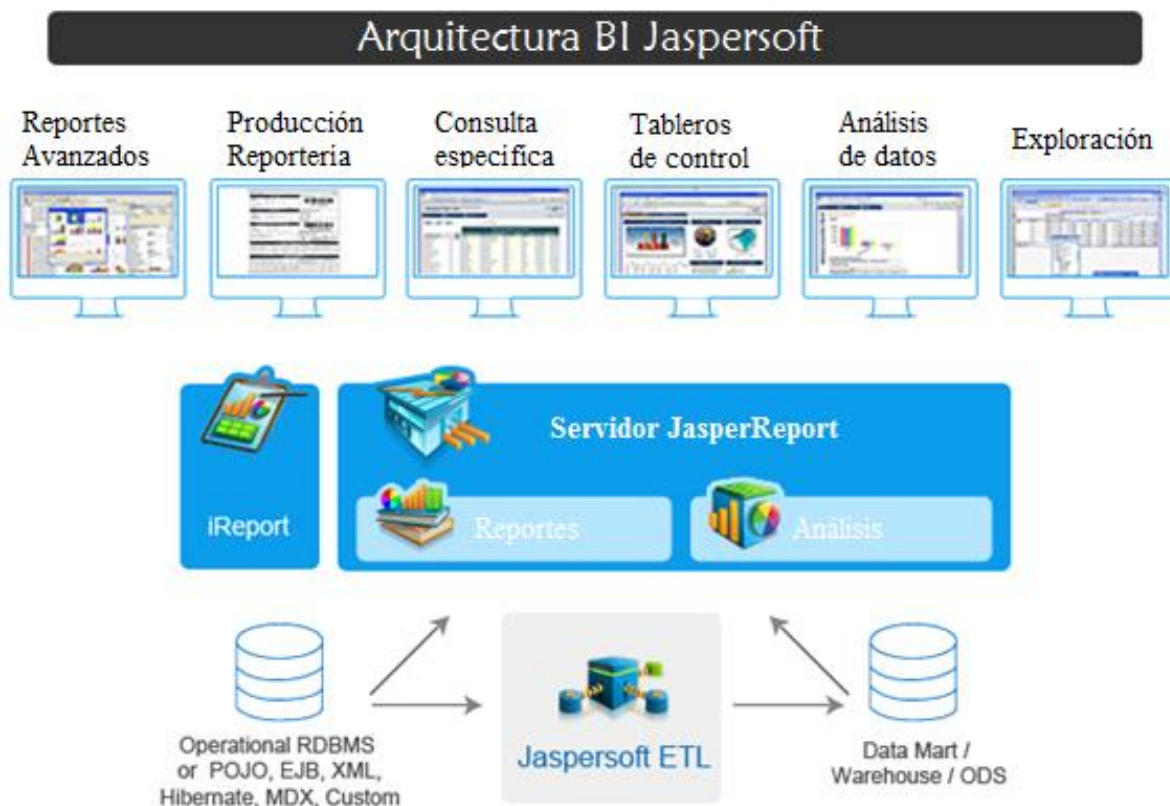


Ilustración 14. JasperSoft BI Architecture. Fuente: JasperSoft (2015), Traducción: Christopher Torres y Adrián Carrión



BIRT es un proyecto de software Open Source que ofrece la plataforma tecnológica BIRT para crear visualizaciones de datos e informes que se pueden integrar en aplicaciones cliente y web grandes, especialmente los basados en Java y Java EE. BIRT es un proyecto de software de alto nivel dentro de la Fundación Eclipse, un consorcio independiente sin fines de lucro de los proveedores de la industria de software y una comunidad de código abierto.

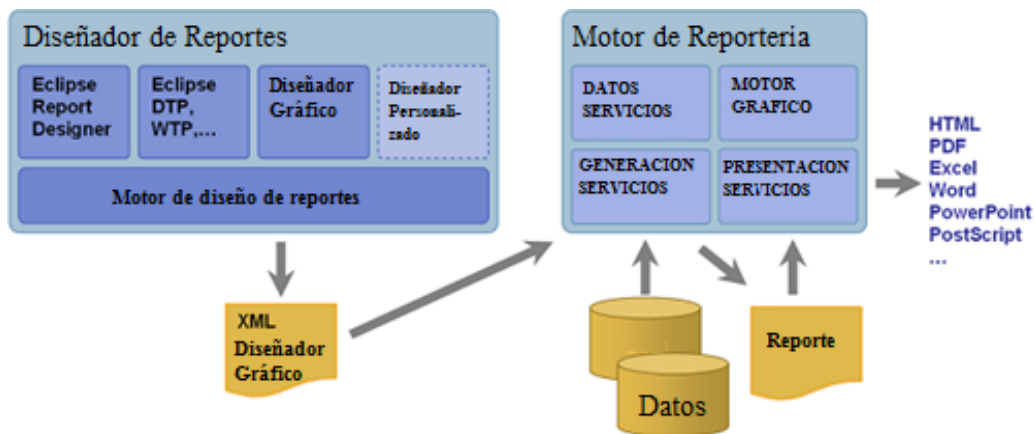


Ilustración 15. Arquitectura BI BIRT. Fuente: Eclipse (2015), Traducción: Christopher Torres y Adrián Carrión

CAPÍTULO III

3. ANÁLISIS COMPARATIVO DE HERRAMIENTAS Y METODOLOGÍAS PARA EL DISEÑO DE UN DATA WAREHOUSE

En este capítulo, se analiza y compara tres herramientas Open Source, con enfoque en reportería. Esta área en los proyectos de Inteligencia de Negocios (BI) en las organizaciones, es utilizada en el análisis y la presentación de la información. Adicional, participa como recurso de apoyo en la toma de decisiones.

3.1 Herramientas Open Source para Inteligencia de Negocios (BI)

El término Open Source se originó en los años 90, por parte de los usuarios de la comunidad de Software Libre, para usarlo como reemplazo al ya tradicional término Software Libre. Sin embargo, no tuvo mucha acogida por algunos usuarios debido a que dicho término eliminaba la idea de libertad, que era confundida generalmente con el término gratuidad (GNU, 2016).

Se utiliza la expresión Open Source, para denominar a ciertos tipos de software, los cuales son distribuidos a los usuarios bajo una licencia, que permita al mismo, si es capaz, lo siguiente: analizar, estudiar, modificar y mejorar el código fuente de un Software. A la vez, permite agregar opciones y corregir errores, con la posibilidad de redistribuirlo, conceptualizándose mejor que el Software original (GNU, 2016).

Por otro lado, el Software Libre hace referencia a la libertad de los usuarios para: copiar, distribuir, cambiar, estudiar y mejorar el software. De igual manera, respeta las cuatro libertades establecidas por la Free Software Foundation⁷, fundada por Richard M. Stallman⁸ (GNU, 2016), que son:

- Libertad de usar el programa con cualquier propósito.
- Libertad de estudiar, modificar y adaptar el programa a necesidades propias.

⁷ FSF – Organización creada en octubre de 1985 por Richard M. Stallman, para promover el movimiento Software Libre.

⁸ Considerado el padre del movimiento Software libre y fundador de la FSF.

- Libertad de distribuir copias del programa.
- Libertad de mejorar el programa y hacer públicas esas mejoras, de modo que toda la comunidad se beneficie.

Es importante aclarar que, el Software Libre propicia y establece muchas libertades, aunque esto no signifique que deba ser necesariamente gratuito. Es decir que, conservando y respetando las libertades que el Software Libre conlleva, se lo puede distribuir comercialmente, garantizando sus derechos de modificación y redistribución (GNU, 2016).

Entre las herramientas a evaluar se encuentran: Pentaho Community Edition, JasperSoft Community Edition y Eclipse Business Intelligence and Reporting Tools (BIRT). Estas son consideradas entre las mejores opciones de Open Source, retribuido a una mayor potencialidad, opciones y capacidades involucradas en la rama de la Inteligencia de Negocios (BI) (Gartner, 2016).

3.1.1 Pentaho Community Edition

El sitio web oficial presenta a la plataforma Pentaho (2016) como: “una compañía líder en integración de datos y análisis de negocio que dispone de una plataforma de clase empresarial Open Source para diversas implementaciones orientadas a Big Data”.

Adicional a esto, se mencionan cuatro aspectos, por los cuales la plataforma Pentaho se considera la mejor elección para las organizaciones líderes e innovadoras del mercado, estas son (Pentaho, 2016):

- **Control en la entrega de datos:** presenta la capacidad de combinar los datos confiables y oportunos, potencializando el análisis en diferentes escalas, para todos los usuarios de las diferentes áreas en las organizaciones.
- **Facilidad para ajustarse:** gracias a su arquitectura multi-propietario, facilita la integración para el análisis. También contribuye en el flujo de trabajo o de diferentes tipos de implementación, como la nube o en modelos híbridos o móviles.

- **Poder para acceder, preparar y combinar datos:** la plataforma de integración de datos, pone a disposición los datos listos desde cualquier fuente de datos, para ser usados por cualquier usuario final. Con una interfaz más amigable, se elimina la complejidad y el código, poniendo al alcance la información en tiempo real.
- **Integración con grandes fuentes de datos:** permite la transformación de grandes volúmenes en grandes ideas, sin importar el tipo de análisis, el entorno de despliegue o requisitos de análisis. La adaptabilidad que se dispone, permite un acceso seguro en el futuro a todas las fuentes de datos.

Pentaho Community Edition pone a disposición herramientas las cuales permiten al usuario empezar con una solución de Inteligencia de Negocios (BI), estas son (Pentaho, 2016):

- **Pentaho Data Integration (Kettle):** es una herramienta que dispone de capacidades ETL. Es decir, permite la extracción, la transformación y la carga de datos. Estas tareas son típicas en procesos de migración, integración con terceros y explotación de big data. Es por esto que, se las aplica en cualquier proyecto de grande o mediana escala. Kettle facilita este trabajo, debido a que, no requiere entrar en detalle en el proceso de realizar estas tareas, por el contrario, determina una tarea específica.
- **Pentaho Analysis Services (Mondrian):** permite analizar grandes cantidades de datos de manera fácil. Es un servidor de procesamiento analítico en línea (OLAP) open source escrito en Java, que responde de manera eficiente, permitiendo una exploración de datos interactiva, sin importar el volumen de estos. Incluye el análisis multidimensional, permitiendo a los usuarios navegar y examinar entre los datos de negocio, explorando y entrelazando información.
- **Pentaho Data Mining (Weka):** permite indagar sus propios datos sobre usuarios, clientes y negocio; transformándola en información útil y relevante de manera sencilla. Ayuda también a encontrar patrones en grandes conjuntos

de datos y extraer la información. Gracias a que está implementado en Java, soporta gran cantidad de tareas de data mining, agregando una gran capacidad de portabilidad.

- **Report Designer:** permite el acceso a cualquier fuente de datos, entre estos: relacional, OLAP o XML. Estos datos pueden ser transformados en información valiosa, la cual es mostrada por reportes en diferentes formatos como: DF, Excel, HTML, Texto plano, XML and CSV. Todos estos reportes pueden ser manejados a través de BI Server.

3.1.2 Eclipse Business Intelligence and Reporting Tools (BIRT)

BIRT menciona en su sitio web oficial, que es proyecto de software open source que proporciona la plataforma de tecnología BIRT, para crear visualizaciones y reportes de datos que, pueden ser adaptados en grandes aplicaciones para cliente o web, especialmente aquellas basadas en Java y en Java EE. Adicional a esto, BIRT es un proyecto de software de alto nivel dentro de la Eclipse Foundation⁹, un consorcio independiente sin fines de lucro de los proveedores de la industria de software y de la comunidad open source (InnovateSolutions, 2016).

El proyecto BIRT es patrocinado por Actuate¹⁰ junto con contribuciones de IBM e Innovent Solutions. En la actualidad, BIRT es una de las tecnologías de visualización y reportería de datos adoptada, con más de 12 millones de descargas y gran cantidad de desarrolladores, en más de 157 países. Entre las grandes empresas que han incorporado BIRT a su negocio tenemos: IBM, Cisco, S1, ABS Nautical Systems, entre otras. Entre las principales componentes de BIRT tenemos (InnovateSolutions, 2016):

- **BIRT Report Designer:** es una perspectiva de Eclipse que se utiliza para crear diseños de informes. Estos diseños son almacenados en un formato XML abierto. Esta herramienta puede ser descargada como una Rich Client

⁹ Lidera el desarrollo de Eclipse, el IDE y plataforma de código abierto para el desarrollo de aplicaciones en Java.

¹⁰ Compañía dedicada al desarrollo y comercialización de software de reportería y análisis de datos.

Platform¹¹ (RCP), mediante un conjunto de plug-ins, y se puede añadir la perspectiva de BIRT Report Designer a un Eclipse IDE existente, como también, se lo puede descargar en un todo en uno que incluye Eclipse.

- **Design Engine:** este motor es el responsable de crear y modificar el diseño de los reportes. El Design Engine API (DE API) contiene la funcionalidad de este motor y puede ser usado con cualquier proyecto Java / Java EE. El BIRT Report Designer utiliza este API internamente para crear los diseños XML.
- **Report Engine:** este motor utiliza los archivos de diseño de reportes para poder generar y renderizar los reportes. El Report Engine API (RE API) este motor puede ser añadido a cualquier aplicación Java / Java EE. El BIRT Web Viewer utiliza este API para ejecutar y mostrar reportes.
- **Charting Engine:** este motor permite diseñar y generar gráficos o tablas, independientes o adaptado a un reporte BIRT. El Charting Engine API (CE API) permite a los desarrolladores Java /Java EE añadir capacidades gráficas o de tablas dentro de sus aplicaciones.
- **BIRT Viewer:** BIRT provee de una vista de “espectador”, utilizada para obtener un modelo preliminar del reporte dentro de Eclipse. Una vez empleada esta vista preliminar BIRT, invoca al servidor incluido Apache Tomcat.

3.1.3 JasperSoft Community Edition

JasperSoft, software de TIBCO, proporciona a ejecutivos, desarrolladores y usuarios de negocio, una solución para negocio rápida, que pone a disposición los datos precisos en el momento preciso, para obtener respuestas más rápidas, mejores decisiones y acciones más inteligentes (InnoventSolutions, 2016).

JasperSoft utiliza la Inteligencia como herramienta que ayuda a proporcionar la información correcta a los usuarios en el tiempo y lugar que se requiera, empleando

¹¹RCP – Herramienta que permite la integración de componentes de software independientes, facilitando el desarrollo de aplicaciones del lado del cliente.

Inteligencia en una aplicación, portal, proceso o sistema de preferencia que regularmente es usado por el usuario (InnoventSolutions, 2016).

JasperSoft en su Community Edition ofrece diferentes herramientas que se adaptan a diferentes organizaciones dependiendo de las capacidades, características y nivel de rendimiento que estas necesiten (InnoventSolutions, 2016):

- **JasperReports Server:** Es un servidor de informes independiente e integrable, que proporciona análisis e informes que pueden ser utilizados en aplicaciones web o móvil, así como también puede actuar como eje central de la organización mediante la entrega de información crítica en tiempo real o de forma programada a un navegador, dispositivo móvil, una impresora o buzón de correo electrónico en diferentes formatos de archivo. JasperReports Server está optimizado para compartir, proteger y gestionar de forma centralizada los informes de JasperSoft y vistas analíticas.
- **JasperReports Library:** La biblioteca JasperReports es el motor de reportería open source más popular del mundo. Está escrito completamente en Java y es capaz de utilizar los datos procedentes de cualquier tipo de fuente de datos y mostrar documentos de alta calidad que se pueden ver, imprimir o exportar en una variedad de formatos incluyendo HTML, PDF, Excel, Open Office y Word.
- **JasperSoft Studio:** Es el diseñador de reportes open source basado en Eclipse para JasperReports y JasperReports Server. Permite crear reportes con un diseño sofisticado que contiene gráficos, imágenes, informes integrados, tablas cruzadas y mucho más. Permite el acceso a los datos a través de JDBC, TableModels, JavaBeans, XML, Hibernate, CSV, y fuentes personalizadas, además de permitir publicar los reportes en formato PDF, RTF, XML, XLS, CSV, HTML, XHTML, texto, DOCX, u Open Office.
- **JasperSoft ETL:** Es la herramienta open source más flexible, potente y asequible para satisfacer necesidades de integración de datos. Diseñado para escalar los más altos niveles de volúmenes de datos y la complejidad de los

procesos. Los desarrolladores pueden gráficamente diseñar, programar y ejecutar los movimientos y transformaciones de datos para diferentes proyectos de Inteligencia de Negocios (BI), como la carga de un Operational Data Store (ODS)¹², data mart, o data warehouse.

Tabla 2:

Cuadro comparativo de herramientas

	BIRT	JasperSoft	Pentaho
Sitio Web CE	eclipse.org/birt	community.jaspersoft.com	community.pentaho.com
Sitio Web Comercial	developer.actuate.com	www.jaspersoft.com	pentaho.com
Licencia	EPL ¹³	LGPLv3 ¹⁴ - AGPLv3 ¹⁵	GPLv2 ¹⁶ - LGPLv2 - Apache License
Componentes incluidos CE	BIRT Report Designer Design Engine Report Engine Charting Engine BIRT Viewer	JasperReports Server JasperReports Library JasperSoft Studio JasperSoft ETL	Pentaho Data Integration BI Server Report Designer
Plataformas	Windows, Linux, Mac	Windows, Linux, Mac	Windows, Linux, Mac
Plug-in de Eclipse	✓	✓	X
Plug-in de NetBeans	X	X	X

Autor: Christopher Torres y Adrián Carrión, agosto 2016

¹² ODS – es donde se almacenan los datos antes de su almacenamiento en un Data Warehouse. Aquí los datos pueden ser borrados, modificados para evitar redundancias y verificados, el análisis y la reportería puedan llevarse a cabo sin detener las operaciones del negocio.

¹³ Eclipse Public License - Licencia utilizada por la Fundación Eclipse para su software.

¹⁴ Lesser General Public License - Licencia que garantiza la libertad de compartir y modificar el software, asegurando que sea libre para sus usuarios.

¹⁵ Affero General Public License - Licencia copyleft que asegura la cooperación con la comunidad en el caso de software que corra en servidores de red.

¹⁶ General Public License - Licencia que ofrece la libertad de usar, estudiar, compartir y modificar software por usuarios finales.

3.2 Parámetros y Categorías de Comparación

Según la Investigadora Gartner (2016) una plataforma de Inteligencia de Negocios (BI) debe ofrecer 14 capacidades divididas en cuatro categorías: infraestructura, administración de los datos, análisis y creación de contenido e intercambio de resultados. Estas capacidades serán analizadas y comparadas en las tres herramientas open source de Inteligencia de Negocios (BI) mencionadas anteriormente.

3.2.1 Infraestructura

- **Administración de la plataforma BI:** Capacidades que permiten escalar la plataforma, optimizando su rendimiento y asegurando una alta disponibilidad y recuperación frente a desastres.
- **BI en la nube:** Capacidades que permiten la construcción, despliegue y gestión de análisis y aplicaciones analíticas en la nube, aplicando la plataforma y las aplicaciones como un servicio.
- **Seguridad y administración de usuarios:** Capacidades que habilitan la seguridad de la plataforma permitiendo la gestión de usuarios y el control de acceso y utilización.
- **Conexión con fuentes de datos:** Capacidad que permite conectar la herramienta con diferentes bases de datos.

3.2.2 Administración de datos

- **Gobierno y gestión de metadatos:** Herramientas que permiten compartir el mismo modelo semántico y metadatos del sistema de almacenamiento de información. Buscando centralizar la búsqueda, captura, almacenamiento, reutilización y publicación de objetos de metadatos, como dimensiones, jerarquías, medidas de rendimiento, parámetros, métricas.
- **Autonomía en ETL y almacenamiento de datos:** Capacidades de la plataforma para acceder, integrar, transformar y cargar datos a una propia capa de almacenamiento, permitiendo indexar los datos y manejar la carga de los mismos.

- **Preparación de datos:** Cuenta con características como “arrastrar y soltar”, combinación de diferentes fuentes de datos, y la creación de modelos de análisis definidos por el usuario en medidas, grupos y jerarquías. Ofrece también capacidades avanzadas como el descubrimiento automático de semántica, Uniones inteligentes, perfiles inteligentes, generación de jerarquías, el linaje y mezcla de datos en varias fuentes de datos, incluyendo datos multiestructurados.

3.2.3 Análisis y creación de contenido

- **Integración de análisis avanzado:** Acceso sencillo a capacidades avanzadas para análisis de datos integrados en la herramienta o mediante la importación e integración de modelos externos.
- **Tableros analíticos:** Permite crear cuadros de mando interactivos junto con contenidos de exploración visual de análisis avanzado y análisis geoespacial que pueda ser consumido por otros.
- **Exploración visual interactiva:** Permite la exploración de datos a través de gráficas con distintos colores, tamaños, formas y movimiento de los objetos que visualmente representan el conjunto de datos analizados. Incluyendo varias opciones de visualización como gráficos de barras, líneas, mapas geográficos, gráficos de dispersión entre otros. Esto permite a los usuarios el análisis de los datos con una perspectiva visual.
- **Exploración móvil:** Permite el desarrollo y entrega de contenido a través de dispositivos móviles aprovechando las diferentes capacidades que integran como pantalla táctil, cámara, geolocalización además de consultas en lenguaje natural.

3.2.4 Intercambio de resultados

- **Integración de contenido analítico:** Incluye un kit de desarrollo con APIs y soporte de estándares abiertos que permiten crear y modificar contenido analítico, visualizaciones y aplicaciones. Además de su incorporación a un proceso de negocio, aplicación o portal.

- **Publicación de contenido analítico:** Capacidades que permiten publicar y desplegar contenido analítico a través de diversos tipos de salida y métodos de distribución.
- **Colaboración e integración social:** Permite a los usuarios compartir y discutir a través de foros, chats, anotaciones sobre información, análisis, contenido analítico o decisiones.

InnovateSolutions. (2016)

3.3 Estudio comparativo entre las herramientas

A continuación, se presenta un cuadro comparativo con los parámetros y categorías señalados anteriormente entre las herramientas Pentaho, BIRT y JasperSoft:

Tabla 3:

Cuadro comparativo

Parámetros y categorías de comparación	Plataformas de Inteligencia de Negocios open source		
	Pentaho	BIRT	JasperSoft
INFRAESTRUCTURA			
Administración de la plataforma BI	Pentaho Business Intelligence Suite Community Edition (CE))	Business Intelligence and Reporting Tools (BIRT)	JasperSoft Business Intelligence Suite Community Edition (CE)
BI en la nube	Pentaho BI Server	X	JasperReports Server
Seguridad y administración de usuarios	✓	X	✓

<p>Conexión con fuentes de datos</p>	<ul style="list-style-type: none"> • JDBC's genéricos • AS/400 • Apache Derby • Borland Interbase • Calpont InfiniDB • Exasol 4 • ExtenDB • Firebird SQL • Greenplum • Gupta SQL Base • H2 • Hadoop Hive • Hadoop Hive 2 • Hypersonic • IBM DB2 • Infobright • Informix • Ingres • Ingres VectorWise • Intersystems Cache • Kettle thin JDBC Driver • KingbaseES • LucidDB 	<ul style="list-style-type: none"> • JDBC's genéricos • Drivers nativos instalados desde la interfaz gráfica 	<ul style="list-style-type: none"> • JDBC's genéricos • Cloudbase • Derby • Firebird • Hadoop Hive • H2 • HSQLDB • IBM DB2 • Inetdae7 • Informix • Ingres • JDBC - ODBC Bridge • MS SQLServer • MariaDB • Mondrian • MySQL • OLAP4J • Oracle • PostgreSQL • SQLite • Sybase • Vertica
--------------------------------------	---	--	---

	<ul style="list-style-type: none"> • MS Access • MS SQLServer • MS SQL Server (Native) • MaxDB (SAP DB) • MonetDB • MySQL • Native Mondrian • Neoview • Netezza • Oracle • Oracle RDB • PostgreSQL • Remedy Action Request System • SAP ERP System • SQLite • Sybase • SybaseIQ • Teradata • UniVerse database • Vertica • Vertica 5+ • dbase III/IV/5 		
--	--	--	--

ADMINISTRACIÓN DE DATOS			
Gobierno y gestión de metadatos	Pentaho Metadata Editor	X	X
Autonomía en ETL y almacenamiento de datos	Pentaho Data Integration (Kettle)	No aplica	JasperSoft ETL
Preparación de datos	<ul style="list-style-type: none"> • Mondrian • Schema Workbench 	BIRT Report Designer	JasperSoft OLAP
ANÁLISIS Y CREACIÓN DE CONTENIDO			
Integración de análisis avanzado	✓	✓	✓
Tableros analíticos	Community Dashboard Editor (CDE)	X	JasperSoft Dashboards
Exploración visual interactiva	Pentaho Report Designer	BIRT Report Designer	JasperSoft Studio
Exploración móvil	X	X	JasperSoft Mobile
INTERCAMBIO DE RESULTADOS			
Integración de contenido analítico	X	X	X
Publicación de contenido analítico	<ul style="list-style-type: none"> • HTML • XHTML • PDF • Excel • XML • Texto Plano • RTF • CSV 	<ul style="list-style-type: none"> • HTML • XHTML • PDF • Excel • XML • Texto Plano • RTF • PowerPoint • CSV 	<ul style="list-style-type: none"> • HTML • XHTML • PDF • Excel • XML • Texto Plano • RTF • PowerPoint • CSV

	<ul style="list-style-type: none"> • Formatos personalizados 	<ul style="list-style-type: none"> • PostScript • Tipos de reportes OpenOffice • Formatos personalizados 	<ul style="list-style-type: none"> • PostScript • Tipos de reportes OpenOffice • Flash (SWF) • Formatos personalizados
Colaboración e integración social	Pentaho Community forums	Eclipse Community Forums	JasperSoft Community

Autor: Christopher Torres y Adrián Carrión, agosto 2016

Tabla 4:

Cuadro de escala ponderativa

Cumple totalmente	Cumple mayormente	Cumple parcialmente	No cumple
4	3	2	1

Autor: Christopher Torres y Adrián Carrión, agosto 2016

Tabla 5:

Cuadro ponderativo

Parámetros y categorías de comparación	Plataformas de Inteligencia de Negocios open source		
	Pentaho	BIRT	JasperSoft
INFRAESTRUCTURA	4	2	3
ADMINISTRACIÓN DE DATOS	4	2	3

ANÁLISIS Y CREACIÓN DE CONTENIDO	3	2	4
INTERCAMBIO DE RESULTADOS	3	3	3
Promedio	3,5	2,25	3,25

Autor: Christopher Torres y Adrián Carrión, agosto 2016

3.3.1 Comparativa de reportería

El análisis comparativo, centrado en reportería, de las tres herramientas para Inteligencia de Negocios (BI) antes mencionadas: Pentaho, BIRT y JasperSoft; se realiza en base a sus componentes, puntos fuertes y débiles. Por último, de manera general, un breve análisis de la componente, para crear reportes de cada herramienta.

3.3.2 Comparativa de reportería Pentaho

Actualmente la Suite Open Source para Inteligencia de Negocios (BI), Pentaho Community Edition (CE), se encuentra en la versión 6.1. Incluyendo las siguientes componentes para reportería (Pentaho, 2016):

3.3.2.1 Componentes Pentaho

- **Pentaho Report Designer:** Es una herramienta que permite crear reportes reemplazando la programación o manipulando XML, a través de una interfaz gráfica. Los reportes creados pueden ser corridos desde otras herramientas como: Pentaho Reporting Engine o Pentaho BI Server. Es una aplicación de escritorio y no está disponible como plug-in para Eclipse o NetBeans.
- **Pentaho Reporting Engine:** Conocido formalmente como JFree Report, es un conjunto de clases Java y APIs que ejecutan informes de Pentaho, basados en XML. Pentaho Reporting Engine proporciona diseños de informes frente a las fuentes de datos y, ofrece salida de reportes en formatos HTML, PDF, Excel, entre otros. Puede ser integrado en diferentes aplicaciones Java, sin embargo, no es necesario utilizar Pentaho BI Server.

- **Pentaho Data Integration (Kettle):** Es una herramienta gráfica para integración de datos, que permite a los desarrolladores crear trabajos y transacciones que pueden ser usados para extraer, transformar y cargar datos una gran cantidad de fuentes de datos distintas.
- **Pentaho Reporting SDK:** Incluye Pentaho Reporting Engine, además de la documentación y librerías, que son necesarias para que los desarrolladores pueden integrar esta herramienta en sus aplicaciones.
- **Pentaho BI Server:** Permite a los usuarios ejecutar reportes y cubos OLAP, a través de una interfaz web. Adicional, provee una infraestructura para reportes y análisis de cubos, para acceder a los datos y metadatos por medio de Pentaho
- **Data Integration (Kettle):** Proporciona ciertas capacidades como: servidor de autenticación, manejo de usuarios, inicio de sesión, notificaciones por correo electrónico, programación de reportes, entre otros. Existen dos consolas en las que diferentes usuarios pueden trabajar:
 - **Pentaho User Console:** Permite a los usuarios finales explorar, ejecutar y ver reportes ya sea en HTML o PDF, y descargarlos en otros formatos.
 - **Pentaho Administrator Console:** Permite a los desarrolladores y administradores implementar reportes, administrar usuarios, administrar seguridades y crear flujos de trabajo.

La componente Pentaho Report Designer destaca su importancia, retribuido a la gran capacidad que ofrece en la elaboración de reportes. Estos son expuestos al usuario final (gerentes, empresas, altos directivos, entre otros), y son utilizados por el mismo en la toma de decisiones para la organización. A continuación, se analiza su funcionalidad.

3.3.2.2 Pentaho Report Designer

Pentaho ofrece una solución completa para la integración de datos y análisis de negocio, que incluye soluciones para la extracción, transformación y carga (ETL), informes básicos, análisis, exploración, visualización y minería de datos. Esta visión tan amplia de la herramienta de cierta forma explica la falta de funcionalidad en informes como tablas cruzadas y gráficos robustos, por esto crear reportes complejos sería más complicado (Pentaho, 2016).

Pentaho es la herramienta más fácil de aprender en creación de reportes básicos, en su interfaz de usuario no se encuentran funcionalidades sofisticadas ni poco usadas, esto facilita el aprendizaje de la misma y cumplimiento de las expectativas de los usuarios (Pentaho, 2016).

Es una herramienta de reportería muy completa, cuya usabilidad y estabilidad para crear reportes simples o moderados no deja nada que desear respecto de funcionalidad frente a sus competidores (Pentaho, 2016).

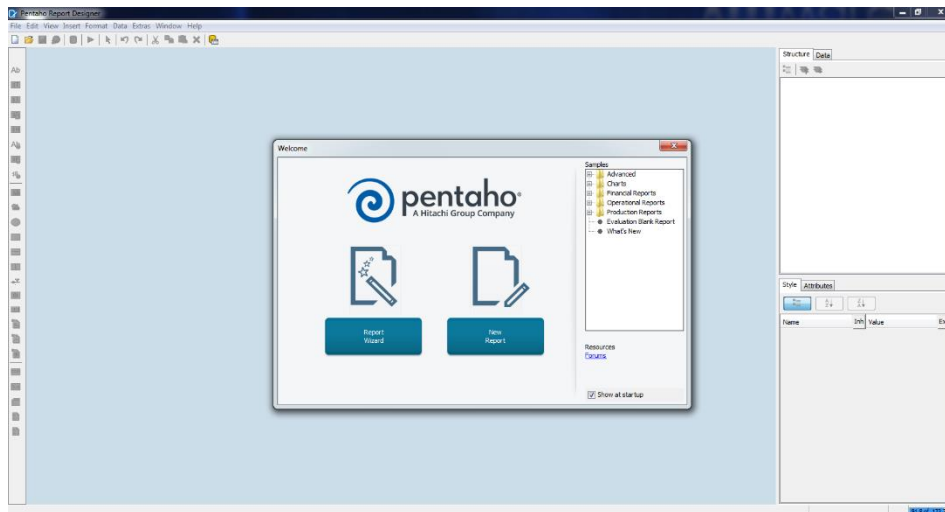


Ilustración 16. Pantalla principal Pentaho Report Designer. Fuente: Herramienta Pentaho Report Designer

Pentaho Report Designer ofrece “posicionamiento de píxeles”, que permite a los usuarios especificar de manera precisa donde será mostrado cada elemento del

reporte y que su adaptabilidad a diferentes tamaños de pantalla, teniendo un control absoluto sobre el aspecto del informe (Pentaho, 2016).

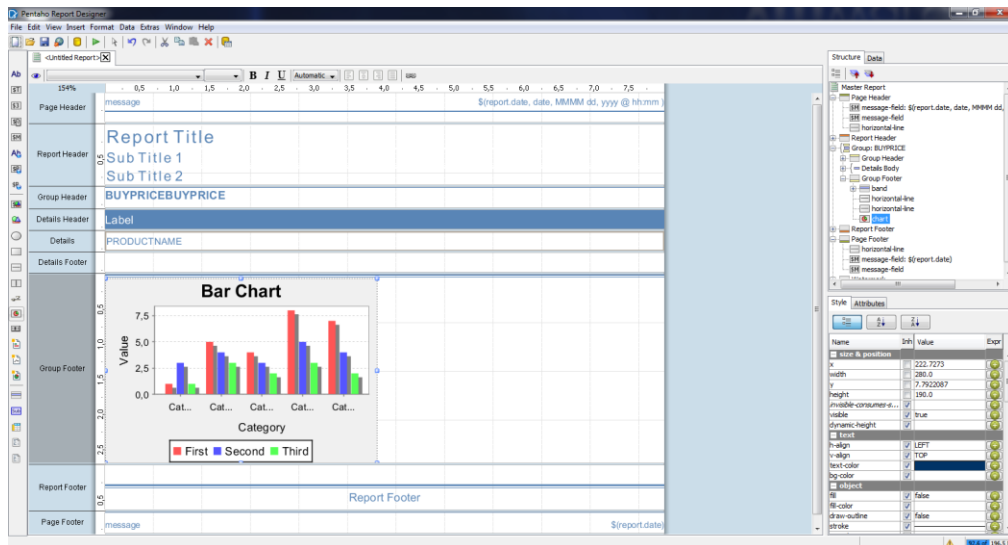


Ilustración 17. Pantalla principal editor de reportes. Fuente: Herramienta Pentaho Report Designer

Una gran característica que ofrece Pentaho Report Designer es el “asistente” para crear reportes en especial para usuarios sin experiencia, permitiendo la creación y edición rápida de reportes, a partir de varias fuentes de datos y con un diseño en estilo de lista (Pentaho, 2016).

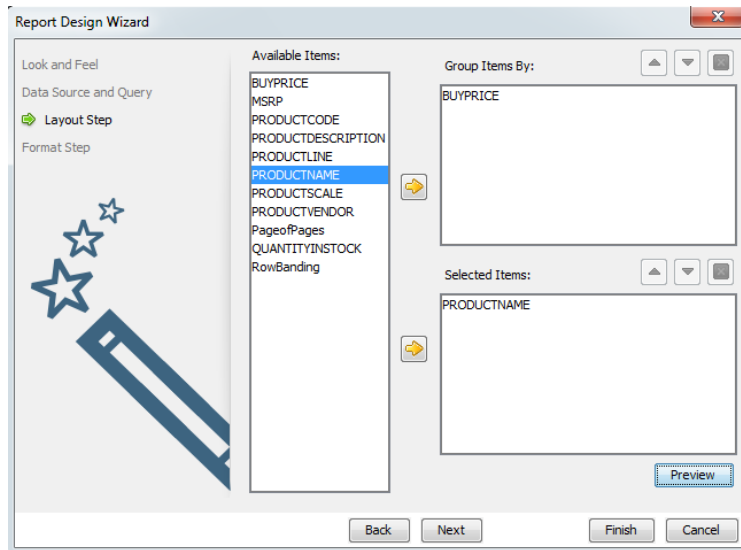


Ilustración 18. Asistente para crear reportes. Fuente: Herramienta Pentaho Report Designer

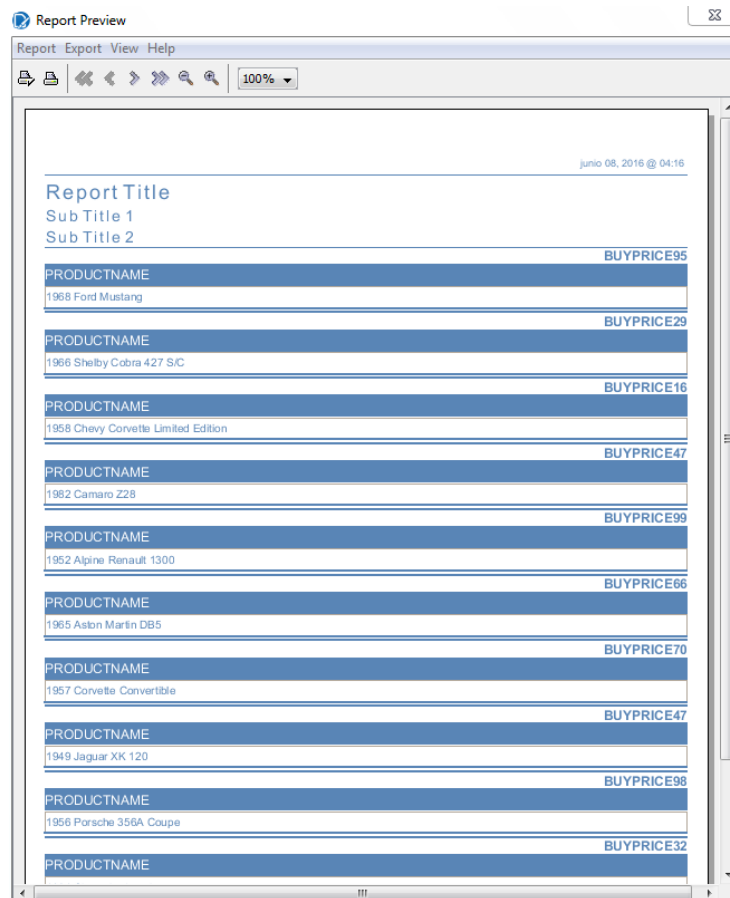


Ilustración 19. Vista previa del reporte. Fuente: Herramienta Pentaho Report Designer

Pentaho Report Designer no cuenta con un “asistente” para la creación de gráficos. Sin embargo, cuenta con un editor de gráficos que le permite escoger el modelo del gráfico a mostrar, pero no permite personalizar el contenido, apariencia y comportamiento que deberá ser configurado manualmente y no podrás verlo hasta ejecutar el reporte ya que no cuenta con una opción de vista previa (Pentaho, 2016).

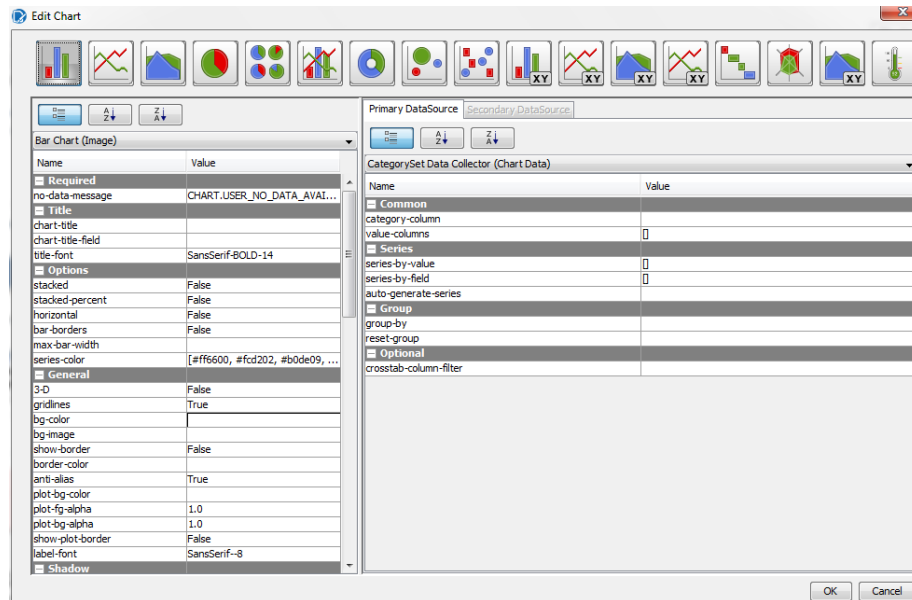


Ilustración 20. Editor de gráficos Pentaho. Fuente: Herramienta Pentaho Report Designer

3.3.2.3 Fortalezas y debilidades Pentaho

Fortalezas:

- Es de las herramientas de diseño la más atractiva visualmente. Se centra principalmente en la experiencia de usuario y usabilidad.
- Cuenta con “asistente” integrado que ayuda en la creación de reportes.
- Con la simple activación de una casilla, permite crear reportes con filas de colores alternos.
- No es necesario compilar el reporte antes de poder ejecutarlo.
- Los reportes creados están en formato XML facilitando un control de revisión.

- El lenguaje de expresión parecido a Excel es mucho más fácil de entender para los no programadores, que a diferencia de las otras herramientas se necesita de conocimiento avanzado de JavaScript.

Debilidades:

- Las tablas cruzadas se encuentran en una fase “experimental”, esta componente no es del todo funcional y además debes activarla manualmente desde Preferencias.
- La creación de gráficas es difícil. No cuenta con un “asistente” para personalizar el contenido, apariencia y comportamiento, esto complica su usabilidad.
- Se requiere que la consulta haga el trabajo de agrupación, filtración, ordenamiento y agregación. Si los datos no llegan al reporte de manera adecuada, Pentaho tiene menos capacidad para manipular que BIRT.
- No soporta reportes con un diseño en columnas como JasperSoft, ni tampoco texto vertical.

3.3.3 Comparativa de reportería BIRT

Eclipse Business Intelligence and Reporting Tools (BIRT), es un proyecto enfocado en herramientas y capacidades para: extraer datos de diferentes fuentes de datos, procesarla y mostrar la información procesada al usuario final. Actualmente, se encuentra en su versión 4.5.0, incluyendo dos componentes principales y tres adicionales que pueden ser usadas por separado (InnoventSolutions, 2016).

3.3.3.1 Componentes BIRT

- **BIRT Report Designer:** Es un diseñador de informes gráfico diseñado para ser fácil de usar. Que puede ser instalada junto con el IDE Eclipse o como un RCP de escritorio. La interfaz gráfica está orientada a los desarrolladores para crear diseños de informes, conectar las fuentes de datos, y producir diseños de informes basados en XML.

- **BIRT Report Engine:** Es un conjunto de clases Java y API's que, hace que los reportes en XML corran al consultar sus fuentes de datos, combinan los datos de la consulta en el diseño del informe y luego produce reportes en diferentes formatos como: HTML, PDF, Excel entre otros.
- **BIRT Chart Engine:** Permite incorporar gráficas a una aplicación Java mediante el uso de Charting Engine API (CE API).
- **BIRT Chart Designer:** Es un diseñador de reportes basado en OSGi¹⁷, que puede ser desplegado y usado junto con aplicaciones que soporten arquitectura OSGi.
- **BIRT Viewer:** Permite la pre-visualización de reportes mediante un plug-in de Eclipse. También se encuentra disponible como una aplicación Java EE que, tras ser montado en un servidor Java, ofrece al usuario final una interface web para ver y navegar a través de reportes, además de crearlos en diferentes formatos como: HTML, PDF, Excel, PowerPoint, Word entre otras.

La componente BIRT Report Designer destaca por una funcionalidad enfocada en elaborar reportes digitales. A continuación, se analiza con detalle la componente.

3.3.3.2 BIRT Report Designer

Es una herramienta robusta y potente para la creación de reportes, de fácil uso, a excepción de determinados casos. Permite crear reportes desde muy simples hasta muy complejos, con una personalización de código mínima. BIRT cuenta con excelente soporte para características de reportería como: gráficos, tablas cruzadas desde varias fuentes de datos y el reuso de consultas para el mismo informe, añadir código personalizado, entre otras. BIRT hace a estas características de fácil uso y, su implementación es mucho más desarrollada.

¹⁷ Open Services Gateway Initiative - sistema y plataforma modular de servicios para Java que implementa un modelo de componente completo y dinámico.

BIRT Report Designer cuenta con algunas características importantes que lo hacen diferente a sus competidores, como las siguientes (InnoventSolutions, 2016):

- Ofrece un diseño orientado a la web para colocar los elementos del reporte en la página, Permitiendo que automáticamente, el reporte se adapte al navegador o pantalla en donde es visualizado. El enfoque de BIRT es la visualización de reportes a través de la web y no en hojas de papel impresos, el llamado “posicionamiento de pixeles” sería la mejor opción si ese es el objetivo de los informes.
- BIRT se considera intimidante para desarrolladores novatos y con experiencia por su gran capacidad. Sin embargo, es la herramienta para reportería más fácil de usar con una funcionalidad de las más completas, junto con su interfaz gráfica permite al usuario acceder a todas las funcionalidades ya que es muy intuitiva una que vez que se trabaje con la misma. Además, la documentación que incluye en el paquete de BIRT es la más completa y gratuita a diferencia de los competidores que ofrecen más documentación por un valor monetario.
- Agregar gráficos a los reportes es sencillo, gracias a que su interfaz guía al usuario paso a paso en todo el proceso de creación de un reporte. La capacidad de previsualizar el reporte en tiempo real a medida que se agregan los parámetros junto con varias opciones de diseño e interacción del usuario con el reporte son una característica única de BIRT.

El “asistente” para crear reportes de BIRT, es muy dinámico y cuenta con una gran variedad de gráficos para escoger y darle el formato deseado. Además, permite escoger la fuente de datos de donde se obtendrán los parámetros a mostrar todo esto visualizándolo en tiempo real en la misma pantalla.

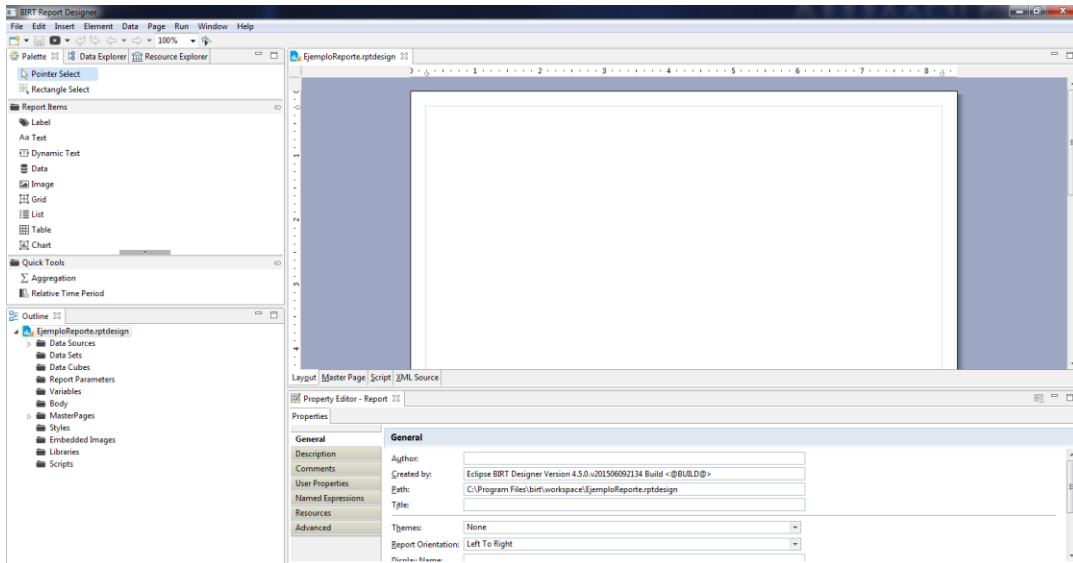


Ilustración 21. Pantalla principal BIRT. Fuente: BIRT Report Designer

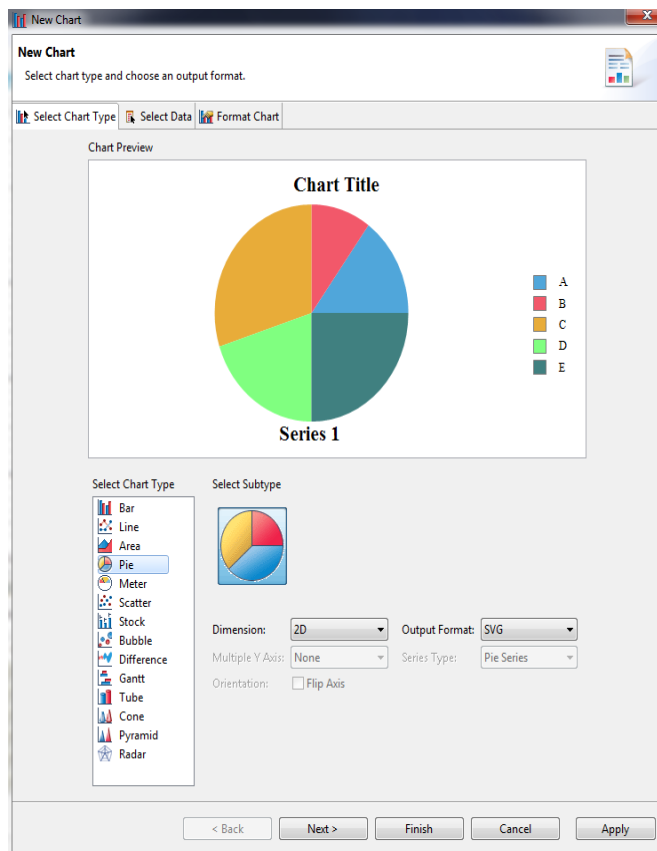


Ilustración 22. “Asistente” para crear reportes. Fuente: BIRT Report Designer

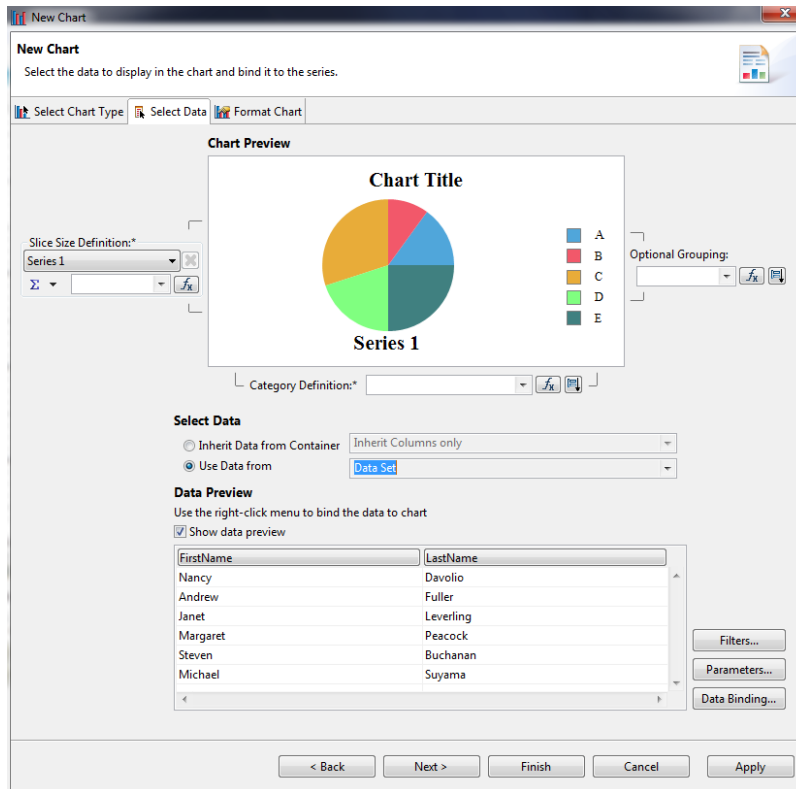


Ilustración 23. “Asistente” para crear reportes. Fuente: BIRT Report Designer

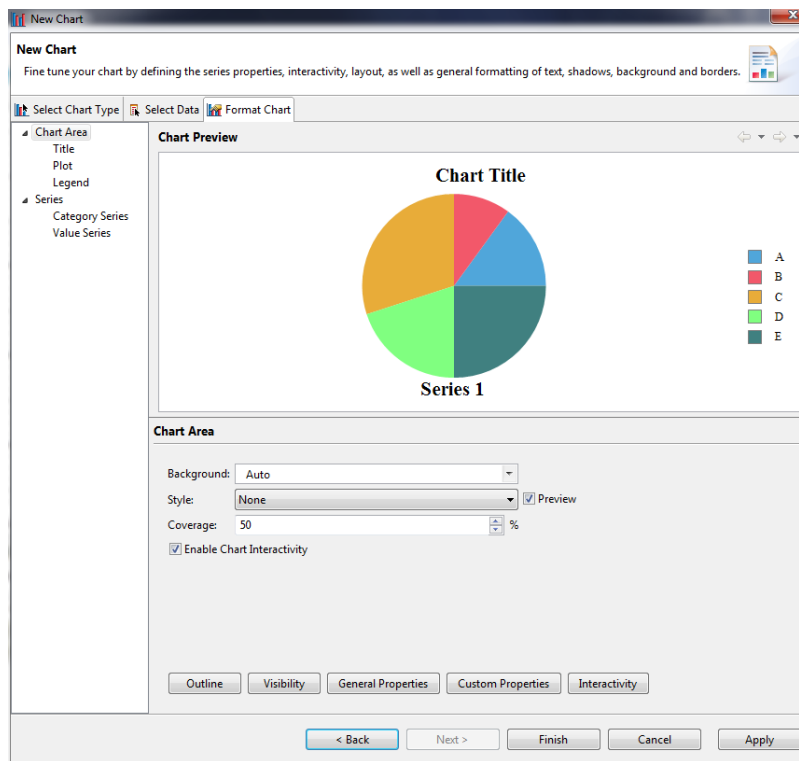


Ilustración 24. “Asistente” para crear reportes. Fuente: BIRT Report Designer

3.3.3.3 Fortalezas y Debilidades BIRT

Fortalezas:

- Es una de las mejores para desarrolladores de aplicaciones Java al estar incluida en el IDE Eclipse, permite una mayor familiaridad con el ambiente de desarrollo.
- Cuenta con la más extensa y completa documentación gratuita.
- Ofrece el mejor y más completo “asistente” para creación de reportes, obteniendo reportes más atractivos e interactivos.
- Cuenta con el mejor soporte en tablas cruzadas.
- Los sub reportes en BIRT son opcionales, incluso para tablas cruzadas o con el uso de varias fuentes de datos.
- Sin necesidad de código por parte de los desarrolladores BIRT permite crear reportes de mediana y grande complejidad con una gran funcionalidad desde la interfaz de usuario.
- Permite la reutilización de componentes, creando librerías de recursos compartidos, que pueden ser usados dinámicamente por varios reportes. Esto ayuda a que si necesitas cambiar una componente que se encuentra en varios reportes se lo hace una vez y no en todos los reportes.

Debilidades:

- La interfaz de usuario puede resultar un poco confusa y desordenada para usuarios novatos de BIRT.
- BIRT es una herramienta para reportería no tan centrada en Inteligencia de Negocios (BI) como Pentaho y JasperSoft.
- No cuenta con integración con mapas.
- No cuenta con un instalador de la herramienta como tal, la instalación de los diferentes drivers para las bases de datos se las debe hacer uno mismo.

- Ordenar las componentes para la salida de un reporte impreso es más difícil de hacerlo y toma tiempo, BIRT está orientado a mostrar reportes a través de la web.
- BIRT no incluye una gran variedad de fuentes de datos preconfiguradas como sus competidores.

3.3.4 Comparativa de reportería JasperSoft

JasperReports Library, JasperSoft Studio y JasperReports Server son parte de la Suite de TIBCO JasperSoft. Estas aplicaciones tienen como núcleo un API y motor de reportería basado en Java llamado JasperReports. JasperSoft Studio es una herramienta open source de diseño para crear reportes, actualmente se encuentra en la versión 6.2.2 y fue un reemplazo para JasperSoft iReport (InnovateSolutions, 2016).

3.3.4.1 Componentes JasperSoft

Dentro de la Suite Community Edition de JasperSoft, se encuentran las siguientes componentes para reportería (InnovateSolutions, 2016):

- **JasperReports Library:** Es una extensa biblioteca de clases Java y API's que permiten diseñar reportes, ejecutarlos a través de fuentes de datos, y generarlos en diferentes formatos.
- **JasperSoft Studio (JSS):** Plug-in de Eclipse para diseñar reportes que reemplaza a iReport.
- **JasperReports Server:** Permite a los usuarios ejecutar reportes y cubos OLAP a través de una interface web, utiliza el servidor Apache Tomcat, sin embargo, se puede ejecutar en cualquier servidor de aplicaciones Java.
- **JasperSoft ETL:** Es un motor de para integración de datos.
- **JasperSoft OLAP:** son un conjunto de herramientas comerciales y open source que permiten el análisis de datos.

De manera similar, que las anteriores componentes mencionadas, JasperSoft se destaca por la elaboración de reportes para organizaciones. A continuación, se analiza su funcionalidad.

3.3.4.2 JasperSoft Studio

Es una herramienta WYSIWYG¹⁸ que permite crear reportes usando una interfaz gráfica, al contrario de crearlos mediante código usando JasperReports Library. Estos reportes luego pueden ser ejecutados mediante JasperReports Engine o JasperReports Server. JasperSoft Studio cuenta con una interfaz bien diseñada con diferentes vistas que permiten al desarrollador ajustar las diferentes de los controles en el reporte (InnovateSolutions, 2016).

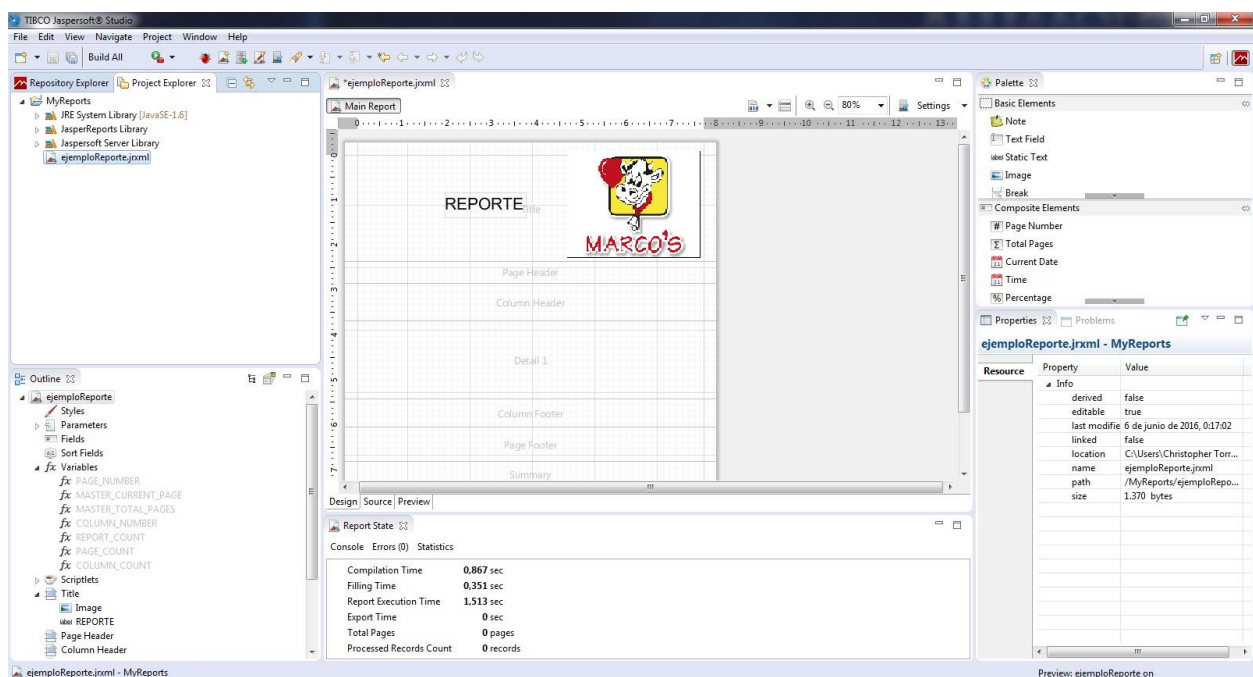


Ilustración 25. Pantalla principal JasperSoft Studio. Fuente: JasperSoft Studio

JasperSoft Studio cuenta con muchas de las mismas funcionalidades que las demás herramientas comparadas. Sin embargo, ofrece algunas características que lo diferencian, por ejemplo: JasperReports utiliza “posicionamiento de pixeles”, haciendo énfasis en reportes paginados, en lugar de un diseño web. La desventaja

¹⁸ What You See Is What You Get – Hace referencia a un editor que permite escribir un documento viendo directamente el resultado final.

de este enfoque es que limita de cierta manera la habilidad para adaptarse a diferentes tamaños de pantalla, donde el diseñador de informes debe colocar explícitamente cada componente en la página y darle formato de forma individual (InnovateSolutions, 2016).

JasperReports se basa en sub reportes para apoyar a la generación de reportes que utilicen varias consultas o fuentes de datos. Estos sub reportes son muy útiles para reusar componentes en diferentes reportes. Sin embargo, elaborar sub reportes puede resultar difícil y complejo en el proceso de creación de reportes, ya que es necesario agregar parámetros y controlar dependencias manualmente (InnovateSolutions, 2016).

JasperReports Studio en su versión open source utiliza como librería para gráficas JFreeChart¹⁹. Las gráficas son construidas a través de un “asistente” fácil de usar, que guía al usuario en el proceso de selección de la fuente de datos, especificación de los campos y en general el diseño del reporte (InnovateSolutions, 2016).

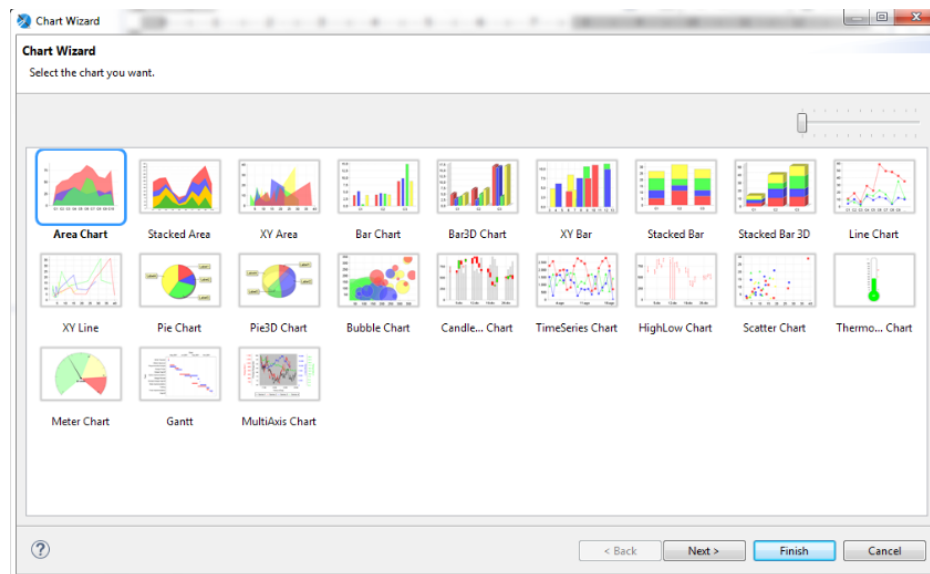


Ilustración 26. “Asistente” para selección de gráficas. Fuente: JasperSoft Studio

¹⁹ Librería gráfica de Java gratuita que permite el mostrar gráficas de calidad profesional en aplicaciones.

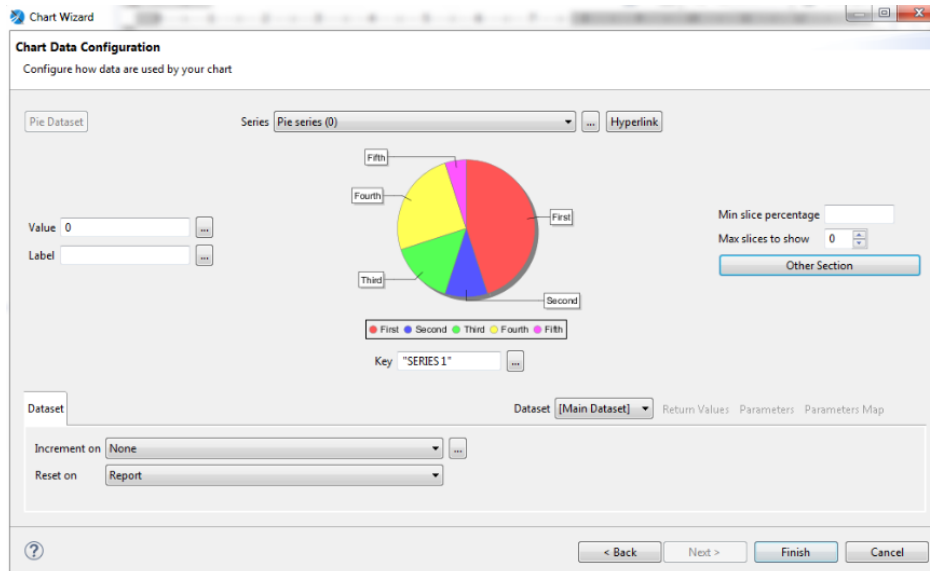


Ilustración 27. Ventana para configuración de datos en la gráfica. Fuente: JasperSoft Studio

Los reportes elaborados en JasperReports necesariamente deben ser compilados antes de poder ejecutarlos, por ende, se necesita que los desarrolladores tengan instalado todo el SDK de Java en sus ordenadores (InnovateSolutions, 2016).

3.3.4.3 Fortalezas y Debilidades JasperSoft

Fortalezas:

- Incluye por defecto varios drivers de diferentes bases de datos instalados, con la posibilidad de agregar más fuentes de datos a través de la interfaz de usuario.
- Permite la creación de reportes con estilo periódico en columnas.
- La documentación de JasperReports es muy completa y entendible, a pesar de que no es gratuita.
- JasperReports es mucho mejor desarrollado en el aspecto de gráficas y tablas cruzadas.
- Es la única herramienta que dispone de componentes para incluir Google Maps en reportes con agregar un campo de la base de datos.

- JasperReports es simple de instalar y de conectar con bases de datos comunes. Al instalar JasperReports Server incluye JasperReports Studio ya configurado para publicar directamente los reportes.
- Es una herramienta ampliamente desplegable que demuestra flexibilidad y estabilidad.

Debilidades:

- JasperReports Studio es el diseñador de reportes más complicado de aprender, en especial al elaborar sub reportes.
- Necesita obligatoriamente tener instalado el SDK de Java, ya que los reportes deben ser compilados antes de ejecutarlos.
- Existen varias componentes para personalizar los reportes a través de la interfaz de usuario. Sin embargo, este proceso es complicado y necesita de un conocimiento medio por parte de los desarrolladores.
- Cuenta con un editor SQL que permite crear consultas, pero es difícil de usar para unión de tablas, y el cambio entre modos de texto y diagramas muchas veces hace variaciones que pasan desapercibidos en el texto SQL.

3.3.5 Resultados del estudio comparativo

Tras el estudio realizado, se evidencia que las herramientas open source para Inteligencia de Negocios (BI) cumplen mayormente las capacidades propuestas por la investigadora Gartner para la elección de la mejor herramienta. Se concluye que Pentaho y JasperSoft destacan por ser herramientas bastante desarrolladas que cumplen en gran medida con funcionalidades y características para el desarrollo de un proyecto de Inteligencia de Negocios (BI) para empresas PYMES²⁰.

Pentaho es la herramienta que alcanzó un mayor puntaje en el cuadro ponderativo con un promedio de 3,5/4 debido a que cumple la mayoría de capacidades destacando

²⁰ PYMES, Hace referencia a pequeñas y medias empresas.

la compatibilidad con gran variedad de fuentes de datos y el proceso de administración de los mismos gracias a las herramientas que proporciona.

Es importante mencionar que herramientas como Pentaho y JasperSoft ofrecen una documentación completa con un costo, a diferencia de BIRT que ofrece la documentación gratuita y de mayor detalle.

Como conclusión de la comparativa de reportería, BIRT es la herramienta que ofrece una mayor potencialidad y facilidad debido a que su orientación a la reportería y no al análisis de datos, permitiendo la presentación de información y creación de reportes de forma más flexible y dinámica.

3.4 Metodologías para la creación de un data warehouse

Existen diferentes metodologías propuestas para el desarrollo de un data warehouse, algunas que van desde el análisis de los requerimientos del usuario final hasta la presentación de reportes y tableros dinámicos.

Ciertas metodologías son extensiones clásicas de bases de datos y en otros casos proponen un nuevo enfoque. Para explicar el esquema de trabajo de estas metodologías, se las puede clasificar en tres categorías: metodologías orientadas por datos, metodologías orientadas por procesos y metodologías compuestas. Las metodologías orientadas por datos tienen como objetivo construir el esquema conceptual del data warehouse a partir de la descripción de las bases de datos, las metodologías orientadas por procesos parten de los requisitos del usuario y las metodologías compuestas son una combinación de las dos anteriores. A continuación, se presenta algunas de las metodologías que se recopiló de la página oficial del SENA²¹:

- **Modelo de Barry Devlon y Paul Morphy:** Inicialmente se analiza minuciosamente el modelo de negocio con el propósito de definir la arquitectura técnica requerida para la implementación del data warehouse; luego se identifican las fuentes de datos operacionales con las que se define un simple, completo y

²¹ El Servicio Nacional de Aprendizaje, SENA, es un establecimiento público del orden nacional colombiano. Adscrito al Ministerio del Trabajo de Colombia, ofrece formación gratuita a millones de colombianos que se benefician con programas técnicos, tecnológicos y complementarios.

consistente sistema de almacenamiento, en el cual los datos se agrupan para que el usuario final pueda entender y manipular el contexto del negocio.

- **Modelo de Ralph Kimball:** Se centra en identificar los procesos de negocio de la organización, desde los cuales se extraen los elementos principales y críticos para la definición de la estructura de las variables y dimensiones que conformaran los cubos de datos que al agruparlos componen el data warehouse. Kimball manifestó que un data warehouse no es más que la unión de los data marts, por lo que él defiende la metodología ascendente (bottom-up) al momento de diseñar un data warehouse.

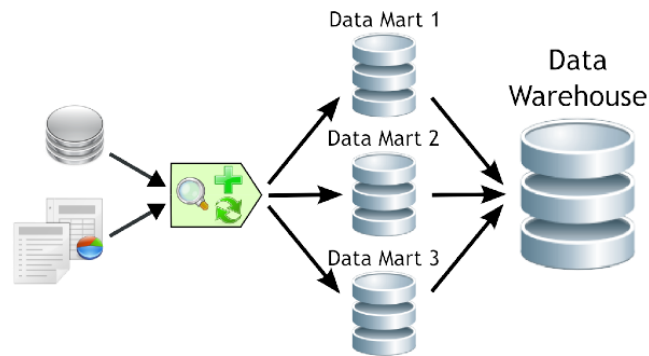


Ilustración 28. Arquitectura bottom-up de un data warehouse. Fuente: Bernabeu R., Dario (2009). Data marts. Recuperado de: <http://www.dataprix.com>

- **Modelo de William Bill Inmon:** basado en un modelo de datos corporativo desde un nivel de abstracción muy gerencial. Inmon defiende una metodología descendente (top-down), que define primero un data warehouse completo para luego crear los data marts dependiendo de las necesidades de información de la organización.

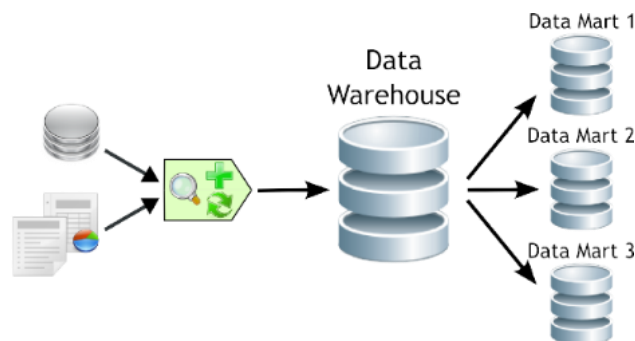


Ilustración 29. Arquitectura top-down de un data warehouse. Fuente: Bernabeu R., Dario (2009). Data marts [figura]. Recuperado de: <http://www.dataprix.com>

- **Modelo de Golfarelli Matteo, Maio Dario y Rizzi Stefano:** Conocido como multidimensional fact model (DFM), el cual propone una metodología semiautomática para lograr un esquema multidimensional partiendo de los modelos entidad relación (MER) de los sistemas transaccionales. Este esquema plantea una asociación de esquemas de hechos, en que participan las dimensiones, las jerarquías, los atributos y los hechos.
- **HEFESTO:** fue propuesta por el Ing. Bernabéu Darío Ricardo. Esta metodología comprende el proceso de construcción de un data warehouse desde un enfoque productivo, en el cual se identifican los elementos y requerimientos más importantes del modelo de negocio con el objetivo de detallar los componentes que conformaran el data warehouse. La construcción e implementación de un data warehouse se adapta muy bien a la mayoría de los procesos de un ciclo de vida de desarrollo de software.

Luego de definir las diferentes metodologías propuestas para el diseño de un data warehouse, se ha elegido la metodología HEFESTO, por estar enfocado a las necesidades de negocio e información de una empresa, por lo cual su estructura es de fácil adaptación a los cambios, respondiendo claramente al ¿Qué? y ¿Cómo? dentro de cada fase de la metodología.

CAPÍTULO IV

4. DESARROLLO Y APLICACIÓN DE LA SOLUCIÓN

Este capítulo describe las metodologías utilizadas para el diseño de un data warehouse y se detalla el proceso de la metodología seleccionada para la aplicación práctica en la empresa. El proceso incluye el análisis de requerimientos, determinación de indicadores, la construcción del data warehouse, proceso de ETL, el desarrollo de los cubos multidimensionales y la presentación de información. Es importante mencionar que en este capítulo se trabaja con las herramientas de Pentaho Community Edition (CE) para la ejecución de algunas tareas. A continuación, se presenta un resumen sobre datos importantes de la empresa.

4.1 Información de la empresa²²

4.1.1 Datos informativos

Nombre: Lácteos de Marco's

Dirección: Km 1 vía a la Primavera

Ciudad: Píllaro - Ecuador

Teléfono: 032874191/032875034

4.1.2 Situación actual

Lácteos de Marco's es una empresa familiar dedicada a la elaboración de productos lácteos, inicia su actividad en el año 1994 con un sistema de trabajo netamente casero y con producción exclusiva de yogurt, con el tiempo fue renovando hasta que en el año 2000 arrancan en la planta de producción actual con una infraestructura apropiada y con altos niveles de producción.

En el año 2003 se inicia la producción de leche pasteurizada y homogenizada, paralelamente se incorpora técnicos y profesionales acordes a las necesidades tanto en la parte de producción, administración, mantenimiento y ventas.

²² La información fue proporcionada por la empresa Lácteos de Marco's

Dentro de todas las actividades que realiza Lácteos de Marco's involucra muchos procesos en diferentes áreas como compras, inventarios, servicios, mantenimiento, producción, recepción, administración, transporte, ventas, etc. Las actividades dentro del proceso principal de producción se detallan a continuación:

- **Recepción de Leche:** la materia prima básica del negocio es la leche, debido a que tiene el poder de determinar la calidad de cualquiera de los productos finales que la empresa produce. Esta etapa, es determinante ya que es donde ocurren los controles de sanidad de la leche y un mal funcionamiento de ésta puede acarrear una multa o la clausura de Lácteos de Marco's.
- **Producción de yogurt:** este producto es uno de los más pedidos, debido a su alto consumo. Sin embargo, es producido en cantidades menores y con tiempos irregulares, es por esto que constantemente se realizan mejoras para brindar el mejor producto de calidad y conseguir mayores ganancias del mismo.
- **Producción de Queso:** es el producto de mayor consumo popular dentro del país y uno de los más solicitados por parte de los clientes. Por esto, se busca eliminar los desperdicios que existen dentro de esta área de producción para beneficio de la empresa y sus trabajadores.
- **Producción de leche Pasteurizada:** es un producto de primera necesidad de alto consumo y su proceso es muy diferente con respecto a la producción de los demás, por esto necesita un mayor control.

4.1.3 Misión

Fabricar y comercializar alimentos saludables, únicos y que satisfacen adecuadamente las necesidades de los clientes.

4.1.4 Visión

Brindar a las familias ecuatorianas una opción saludable de alimentación en base al consumo de productos lácteos de excelente calidad.

4.1.5 Objetivos

- Satisfacer las necesidades de nuestros clientes en base a la elaboración de productos lácteos de calidad.
- Generar plazas de trabajo que ayuden a mejorar el estilo de vida de las personas.
- Contar con procesos de producción amigables con el medio ambiente y utilizando los avances tecnológicos actuales.

4.1.6 Estrategias

- Analizar los requerimientos de los clientes para mejorar paulatinamente.
- Enfocarse en brindar productos de calidad a las familias ecuatorianas.
- Descubrir nuevos nichos de mercado de los productos lácteos.

4.1.7 Valores

- Responsabilidad con la satisfacción de nuestros clientes.
- Responsabilidad con el medio ambiente.
- Responsabilidad con la sociedad de la familia Lácteos de Marco's.

4.1.8 Políticas

La empresa Lácteos de Marco's, dedicada a procesar y comercializar productos lácteos, con unos clientes que cada día demandan más servicios y con mayores exigencias de Calidad, entiende que la Mejora Continua de la Calidad es un objetivo prioritario para lograr la satisfacción de los clientes y mejorar la competitividad.

4.1.9 Estructura de la organización

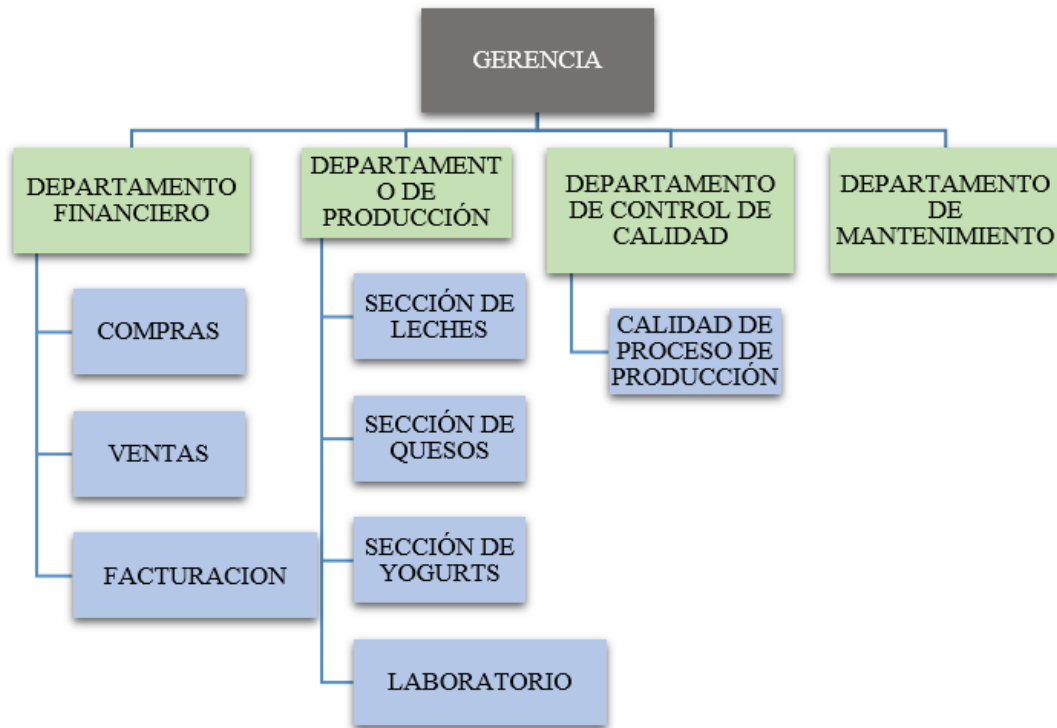


Ilustración 30. Organigrama de Negocio. Fuente: Lácteos de Marco's

4.1.10 Situación actual de TI

La empresa cuenta con un acercamiento básico con la tecnología. Dispone de un software que registra los productos almacenados que se dispone para la venta, al igual que un sistema con la información de sus proveedores y distribuidores para la compra de la materia prima y venta del producto terminado, esta información se la puede acceder desde las terminales de control tanto en el área de embotellado y envasado, empacado, bodega, despacho y desde el área de ventas.

El negocio principalmente se lo controla en papel y luego se lo respalda en digital al final del día, el control consiste en llevar un registro de que productos, la cantidad y el destino de los mismos. De esta forma controlan la salida de la mercadería y el ingreso de capital.

En cuanto a la parte de producción, los procesos se realizan a mano, la medición de la materia prima se la realiza en balanzas industriales, la receta para cada producto

es conocida por el técnico de preparación y es almacenada en una computadora en el área de preparación y respaldada en papel en la caja fuerte del gerente general.

4.1.10.1 Hardware

Tabla 6:

Hardware que dispone la empresa

EQUIPO	#	ÁREA
Computador desktop i3 500GB disco duro 4GB de RAM Monitor 17"	4	Gerencia Recepción Contabilidad Inventario
Computador desktop i3 1000GB disco duro 4GB de RAM Monitor 17"	4	Estaciones de acceso en áreas de la empresa

Fuente: Lácteos de Marco's

4.1.10.2 Software

La empresa Lácteos de Marco's tiene implementado un Sistema de Gestión Empresarial propio acorde a sus necesidades llamado "SIGEMP 3.0" el cual fue desarrollado exclusivamente para la empresa por el Ing. Gustavo Condoy.

Este software permite gestionar:

- Ventas de los productos en fábrica
- Las compras de materia prima
- Los clientes de la empresa
- Los proveedores de la empresa
- El inventario de productos

- El inventario de materia prima
- La contabilidad

Además, que permite controlar y generar reportes de:

- La producción diaria
- La recepción de leche
- Los costos de producción

4.1.11 Importancia del data warehouse para la empresa y su alcance

La solución de data warehouse propuesta está orientada al cumplimiento de las metas de la organización a través del diseño de un modelo que permita estructurar los datos de forma detallada y se pueda obtener información relevante para ayudar en la toma de decisiones. Esto es importante para poder lograr una ventaja competitiva y conocer los factores que influyen en el rendimiento de la organización.

Como se mencionó anteriormente la empresa cuenta con un software de gestión empresarial para la automatización de algunos procesos de administración; sin embargo, este no proporciona reportes con información minuciosa y organizada por diferentes indicadores disponible para que la alta gerencia pueda analizarla e interpretarla y tomar decisiones más acertadas.

El data warehouse aporta un gran valor a la empresa, y su alcance es el siguiente:

- Transforma los datos operativos obtenidos del software de gestión empresarial en información analítica.
- Permite a la alta gerencia contar con reportes personalizados y dinámicos con información estadística y gráficos representativos que brindaran soporte en la toma de decisiones en el departamento financiero de la empresa.
- Permite exportar los reportes a diferentes formatos provistos por la herramienta seleccionada.
- El costo y empleo de recursos para la elaboración de los reportes es mínimo.

4.2 Metodología HEFESTO

Ahora se detalla la metodología HEFESTO, que permite la construcción del data warehouse de una manera sencilla y ordenada, el cual sirve para proveer información necesaria para la generación de informes personalizados y dinámicos que ayuden a la toma de decisiones en el departamento financiero de la empresa.

En el libro “Hefesto”, el creador de esta metodología, Bernabéu (2010) afirma:

HEFESTO es una metodología propia, cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenadas de datos... La idea principal, es comprender cada paso que se realizara, para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué. (p.85)

La siguiente figura muestra los pasos a seguir para aplicar la metodología, los cuales se agrupan en 4 niveles: análisis de requerimientos, análisis de los OLTP, modelo lógico del data warehouse e integración de datos.



Ilustración 31. Metodología HEFESTO, pasos. Fuente: Bernabeu R., Dario (2010). Metodología HEFESTO, pasos. En HEFESTO (p. 87), Córdoba, Argentina, 2010

Como primera instancia se recolectan las necesidades de información de los usuarios para poder abstraer las preguntas claves a responder dentro del proceso de negocio de la empresa. Luego se procede a determinar los indicadores que resultan de las interrogantes y perspectivas de análisis que se utilizan para el diseño del enfoque conceptual del data

warehouse. Después se estudian los OLTP para determinar la manera en que se construirán los indicadores, es decir qué tipo de datos se necesitan, los campos y de que fuentes se conseguirán. Posteriormente, se pasa a la construcción del modelo lógico del data warehouse, en donde se define qué tipo de esquema se implementará para la construcción de las tablas de hechos y dimensiones con sus respectivas asociaciones. Para finalizar, mediante la utilización de técnicas de limpieza de datos, procesos ETL, se carga el data warehouse para poder utilizar su información en la generación de reportes.

4.2.1 Pasos y aplicación de la metodología

4.2.1.1 Análisis de requerimientos

Se inicia identificando los requerimientos de información que tiene de la alta gerencia a través de preguntas que reflejen los objetivos de la organización. Luego se analiza las preguntas con la finalidad de identificar cuáles las perspectivas de análisis e indicadores para la construcción del data warehouse. Por último, se muestra el modelo conceptual para visualizar el resultado de la primera etapa.

4.2.1.1.1 Identificar preguntas

El objetivo principal de este paso es identificar las necesidades de información relevante, para ayudar a la alta gerencia a tomar de decisiones de forma eficaz y eficiente. Para ello, se debe formular preguntas detalladas sobre el negocio, identificando posibles variables de análisis porque estas son las que permiten estudiar la información desde varias perspectivas. Una consideración a tomar en cuenta al realizar las preguntas, es que los datos deben estar soportados por algún OLTP, ya de otra forma, no se podrá elaborar el data warehouse.

Como se señaló anteriormente en el alcance de esta disertación, se va a considerar las necesidades de información del departamento financiero de la organización, por lo tanto, a continuación, se detalla el proceso que se vaya a cabo en este departamento:

El departamento financiero es supervisado por el gerente y propietario de la empresa, el Ing. Marco Vinicio Proaño Chiluiza; entre las funciones de este departamento está la de coordinar los canales de distribución de los productos fabricados. En los últimos años, mediante estrategias de venta la empresa ha apuntado al mercado interno del país, apoyándose en la distribución de sus productos en las principales ciudades del país, como Machala, Ambato, Quito, Píllaro, Guayaquil y Esmeraldas, donde son ampliamente reconocidos. El mercado de lácteos en Ecuador es un sector competitivo, son alrededor de seis empresas las que se pueden considerar grandes. Lácteos de Marco´s se encuentra entre las empresas medianas del país, a pesar de ello cuenta con importantes clientes a nivel nacional, el más importante es Tiendas Industrias Asociadas (TIA S.A.), la cual ofrece productos de consumo hogareño y personal a través de una amplia cobertura nacional.

Otro dato a considerar es que la empresa no cuenta con un plan ni un área de marketing dedicada al diseño y promoción de los productos, por lo que se trabaja con organizaciones externas para el diseño de envases, empaques, etiquetas y promociones que básicamente se van adaptando a los cambios en el mercado. La siguiente figura muestra la gestión actual de marketing a la cual obedece la empresa:



Ilustración 32. Gestión de marketing actual. Fuente: Lácteos de Marco´s

Como muestra la figura, al tratarse de nuevos clientes no existe un proceso como tal a seguir, simplemente se hace contacto directo con el cliente cuando lo requiere; con respecto a su cliente más grande, la empresa TIA S.A., tiene una gran importancia ya que las promociones y oferta de productos se dan de acuerdo a las necesidades del mismo, destacando que actualmente se maneja una marca propia para los productos que se ofrece a este cliente.

Se presentan las preguntas que se formularon con respecto a la información anteriormente mencionada sobre el proceso de venta:

- ¿Se cuenta con estadísticas de unidades vendidas de cada producto organizado por tiempo, ciudad y cliente?
- ¿Se cuenta con estadísticas de monto total de ventas de cada producto organizado por tiempo, ciudad y cliente?
- ¿Se cuenta con estadísticas de unidades compradas de cada materia prima organizado por tiempo y proveedor?
- ¿Se cuenta con estadísticas de monto total por compra de cada materia prima organizado por tiempo y proveedor?

En las preguntas formuladas se evidencia la importancia que se hizo a la dimensión Tiempo debido a que es un elemento fundamental en un data warehouse, ya que permite tener varias versiones de los datos con la finalidad de realizar un correcto análisis posterior. También se puede apreciar, que las preguntas van enfocadas a los objetivos y estrategias de la empresa, debido a que la información requerida es la que provee el ámbito para la toma de decisiones, que en este caso son las ventas, para así lograr analizar la información y obtener una ventaja competitiva y maximizar las ganancias.

4.2.1.1.2 Identificar indicadores y perspectivas de análisis

El próximo paso es la descomposición de las preguntas de negocio para encontrar los indicadores y las perspectivas de análisis que ayudan a la construcción del data warehouse.

- Proveedor

4.2.1.1.3 Modelo Conceptual

Se muestra el modelo conceptual del data warehouse a partir de los indicadores y perspectivas de análisis obtenidos; este modelo refleja el alcance del proyecto y un alto nivel de definición de datos, el cual permite ser explicado de manera sencilla al usuario. La propuesta de representación gráfica del modelo conceptual propuesto por el autor de esta metodología es la siguiente:

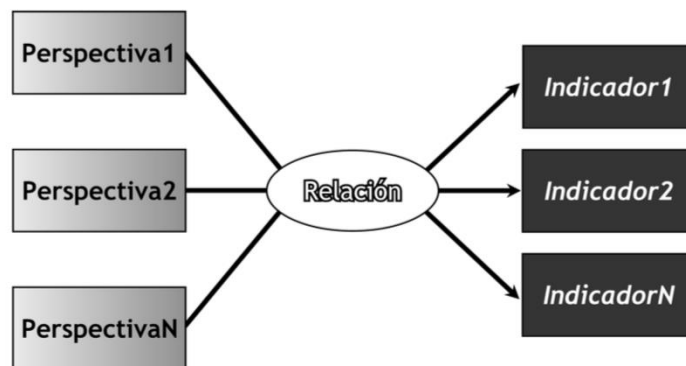


Ilustración 33. Modelo conceptual. Fuente: Bernabeu R., Dario (2010). Modelo conceptual [figura]. En HEFESTO (p. 92), Córdoba, Argentina, 2010.

Como muestra la figura, del lado izquierdo se colocan las perspectivas de análisis que son unidas a un óvalo central el cual representa la relación que existe entre ellas. La relación refleja el proceso seleccionado y se desprenden flechas que apuntan a los indicadores los cuales se ubican en la parte derecha del modelo.

A continuación, se presenta el modelo conceptual propuesto de ventas de productos y compra de materia prima para la construcción del data warehouse:

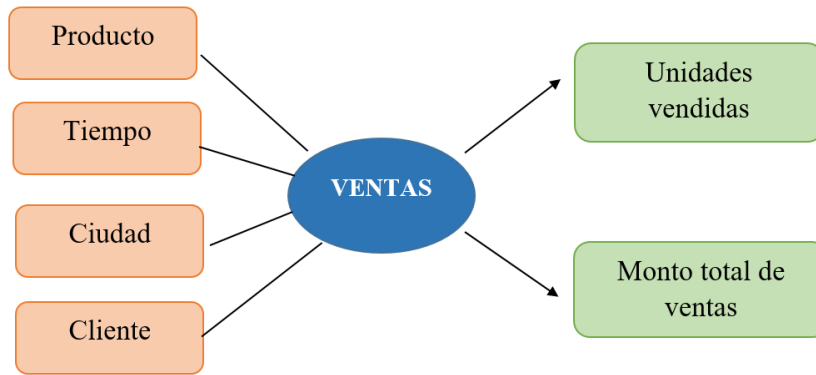


Ilustración 34. Modelo conceptual de ventas. Fuente: Christopher Torres y Adrián Carrión

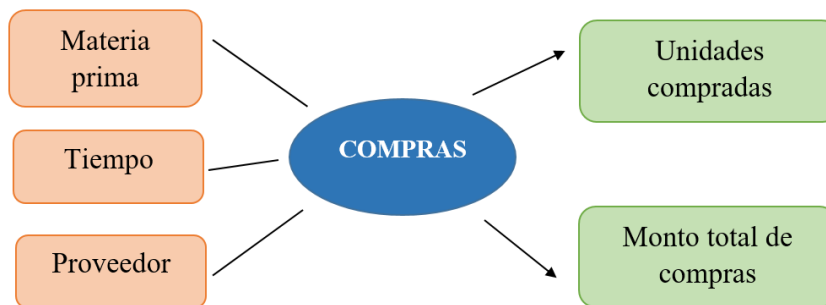


Ilustración 35. Modelo conceptual de compras. Fuente: Christopher Torres y Adrián Carrión

4.2.1.2 Análisis de los OLTP

4.2.1.2.1 Conformar los indicadores

Se explica a detalle cómo se calculan los indicadores, definiendo los siguientes pasos:

- Hecho/s que componen al indicador y la fórmula para calcularlo.
- La función de agrupamiento, ejemplo: SUM, COUNT, AVG, MIN, MAX

A continuación, se aplican los pasos antes mencionados a los indicadores:

II: Unidades vendidas

- Hechos: Unidades vendidas

- Función de agrupamiento: SUM

El indicador representa la sumatoria de las unidades que se han vendido de un producto determinado.

I2: Monto total de ventas

- Hechos: (Unidades vendidas) * (Precio de venta)
- Función de agrupamiento: SUM

El indicador representa la sumatoria de la cantidad total que se ha vendido de cada producto y se obtiene multiplicando las unidades vendidas por el precio respectivo.

I3: Unidades compradas

- Hechos: Unidades compradas
- Función de agrupamiento: SUM
- El indicador representa la sumatoria de las unidades que se han comprado de una materia prima determinada.

I4: Monto total de compras

- Hechos: (Unidades compradas) * (Precio de compra)
- Función de agrupamiento: SUM

El indicador representa la sumatoria de la cantidad total que se ha comprado de cada materia prima y se obtiene multiplicando las unidades compradas por el precio respectivo.

4.2.1.2.2 Establecer correspondencias

Se analizan las características de los OLTP disponibles que contengan la información necesaria para identificar las relaciones entre el modelo conceptual y las fuentes de datos.

La empresa cuenta con un sistema de gestión empresarial para el manejo de algunos procesos administrativos, este se enlaza a un servidor de base de datos donde se guarda toda la información. El proceso de venta de productos y compra de materia prima está representado por el modelo físico que se relaciona con el modelo conceptual en las siguientes figuras, para determinar de dónde se obtendrán los datos.

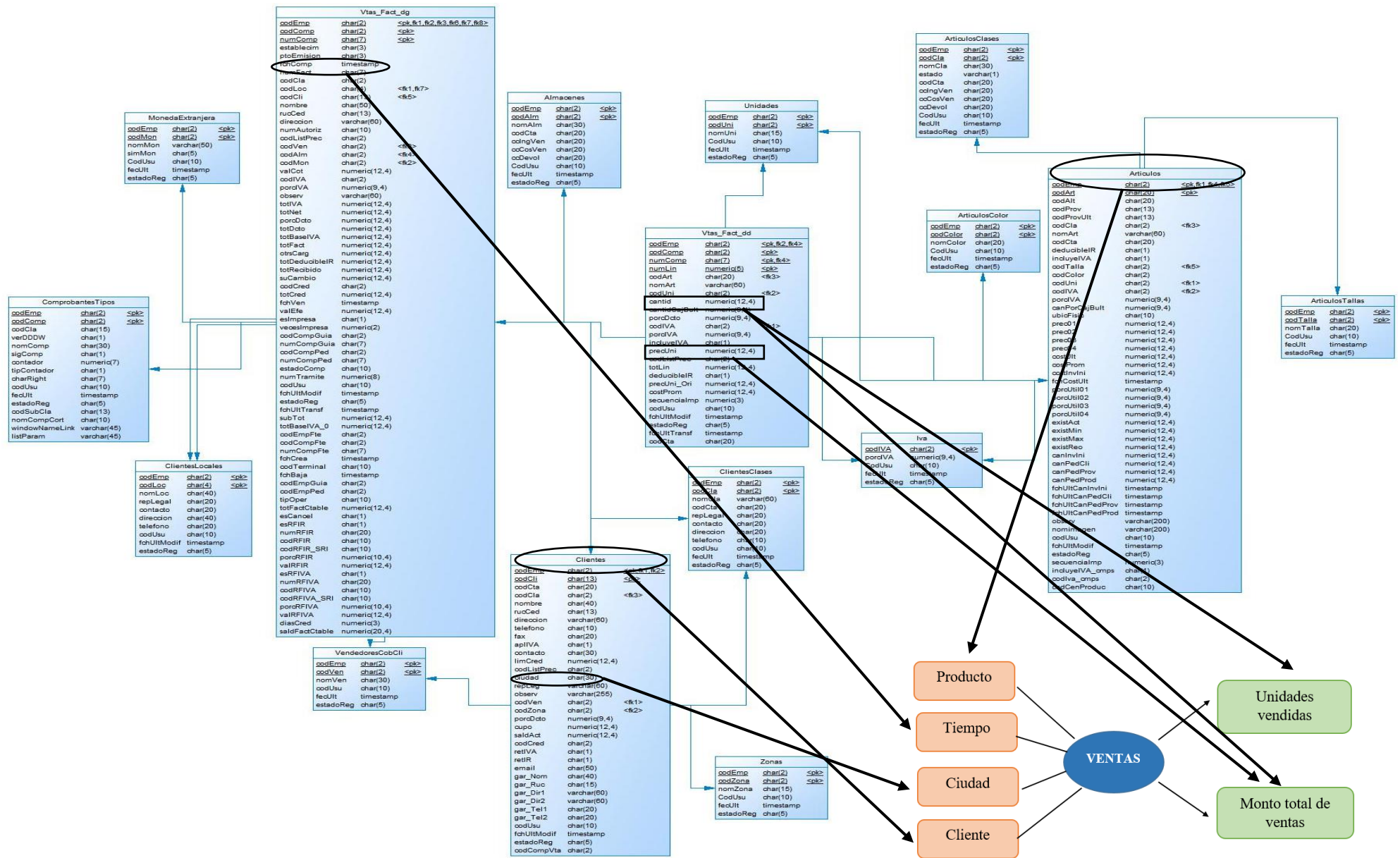


Ilustración 36. Modelo físico de la sección de venta de productos. Fuente: Lácteos de Marco's

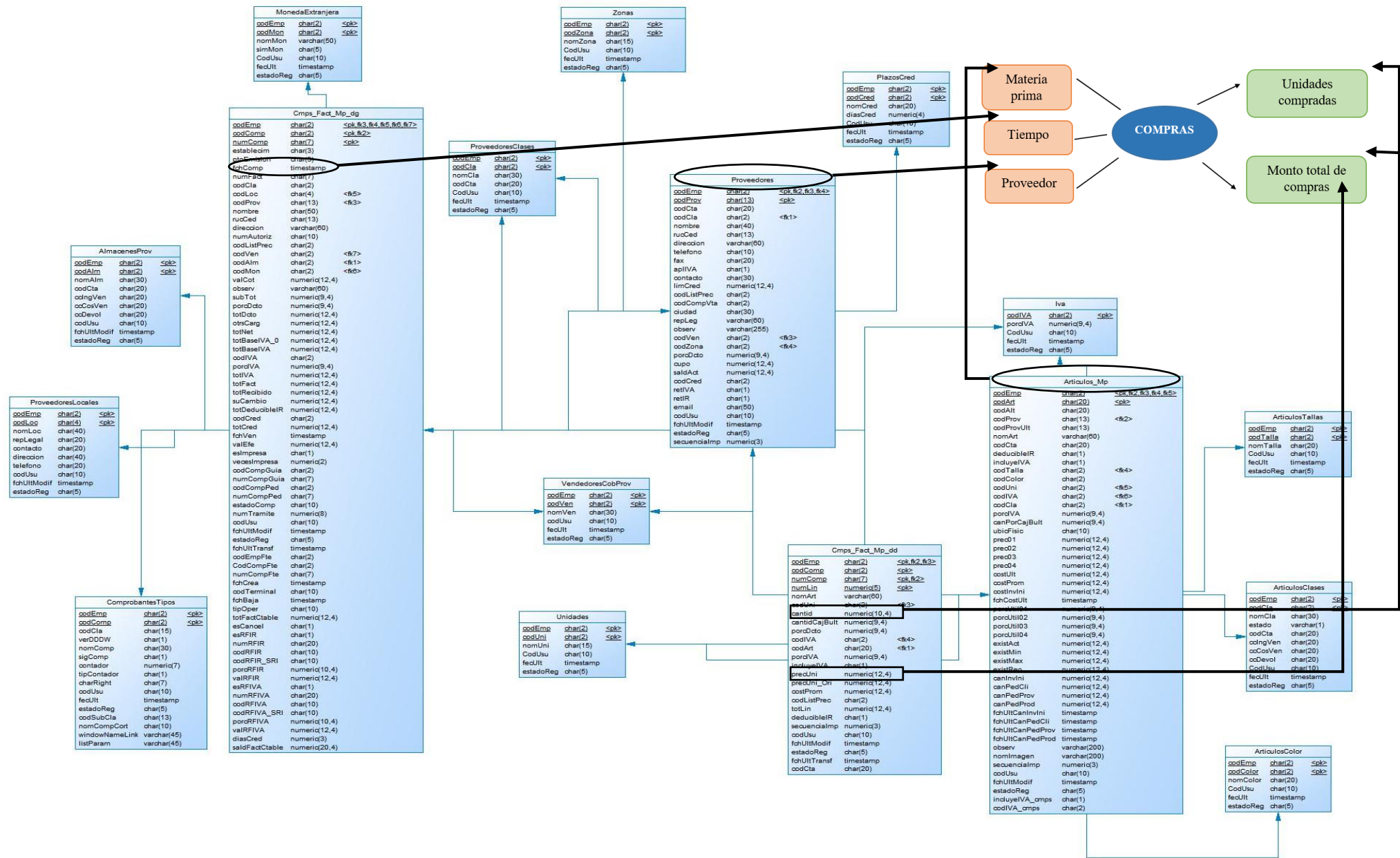


Ilustración 37. Modelo físico de la sección de compra de productos. Fuente: Lácteos de Marco´s

Las relaciones entre el modelo conceptual y el modelo físico de los procesos de ventas y compras que se identificaron son las siguientes:

Proceso de venta

- La tabla “Articulos” se relaciona con la perspectiva “Producto”.
- La tabla “Clientes” se relaciona con la perspectiva “Cliente”.
- El campo “fchComp” de la tabla “Vtas_Fact_dg” se relaciona con la perspectiva “Tiempo” debido a que es la fecha principal cuando se registra una venta.
- El campo “ciudad” de la tabla “Clientes” se relaciona con la perspectiva “Ciudad”.
- El campo “cantid” de la tabla “Vtas_Fact_dd” se relaciona con el indicador “Unidades vendidas”.
- El campo “cantid” multiplicado por el campo “precUni” de la tabla “Vtas_Fact_dd” se relaciona con el indicador “Monto total de ventas”.

Proceso de compra

- La tabla “Articulos_Mp” se relaciona con la perspectiva “Materia prima”.
- La tabla “Proveedores” se relaciona con la perspectiva “Proveedor”.
- El campo “fchComp” de la tabla “Cmps_Fact_Mp_dg” se relaciona con la perspectiva “Tiempo”.
- El campo “cantid” de la tabla “Cmps_Fact_Mp_dd” se relaciona con el indicador “Unidades compradas”.
- El campo “cantid” multiplicado por el campo “precUni” de la tabla “Cmps_Fact_Mp_dd” se relaciona con el indicador “Monto total de compras”.

4.2.1.2.3 Nivel de granularidad

Se seleccionan los campos que contienen las perspectivas de análisis, debido que a través de ellos se examina y filtra los indicadores. A continuación, se detalla los campos que contienen las perspectivas de acuerdo a las correspondencias establecidas con las tablas del modelo físico del paso anterior:

Tabla 7:

Campos de las perspectivas de análisis del proceso de venta

PERSPECTIVA DE ANÁLISIS	CAMPO	DESCRIPCIÓN
Producto	nomArt	Hace referencia al nombre de la materia prima, sacado de la tabla Articulos_Mp.
Tiempo	mes	Se determina la siguiente granularidad de acuerdo a las necesidades de información de la alta gerencia.
	trimestre	
	anio	La fecha se obtiene de la tabla Vtas_Fact_dg.
Ciudad	ciudad	Campo correspondiente al nombre de la ciudad correspondiente al Cliente.
Cliente	nombre	Hace referencia al nombre del cliente.

Fuente: SIGEMP 3.0, Lácteos de Marco's

Tabla 8:

Campos de las perspectivas de análisis del proceso de compra

PERSPECTIVA DE ANÁLISIS	CAMPO	DESCRIPCIÓN
Materia prima	nomArt	Hace referencia al nombre del producto sacado de la tabla Articulos.
Tiempo	mes	Se determina la siguiente granularidad de acuerdo a las necesidades de información de la alta gerencia. La fecha se obtiene de la tabla Cmps_Fact_Mp_dg.
	trimestre	
	anio	
Proveedor	nombre	Hace referencia al nombre del proveedor.

Fuente: SIGEMP 3.0, Lácteos de Marco's

4.2.1.2.4 Modelo conceptual ampliado

Para finalizar la segunda etapa de la metodología HEFESTO, se procede a graficar los resultados obtenidos en el modelo conceptual, colocando bajo cada perspectiva de análisis los campos correspondientes y la fórmula para obtener los indicadores. Las siguientes figuras, muestran el modelo conceptual ampliado para los procesos de venta y compra:

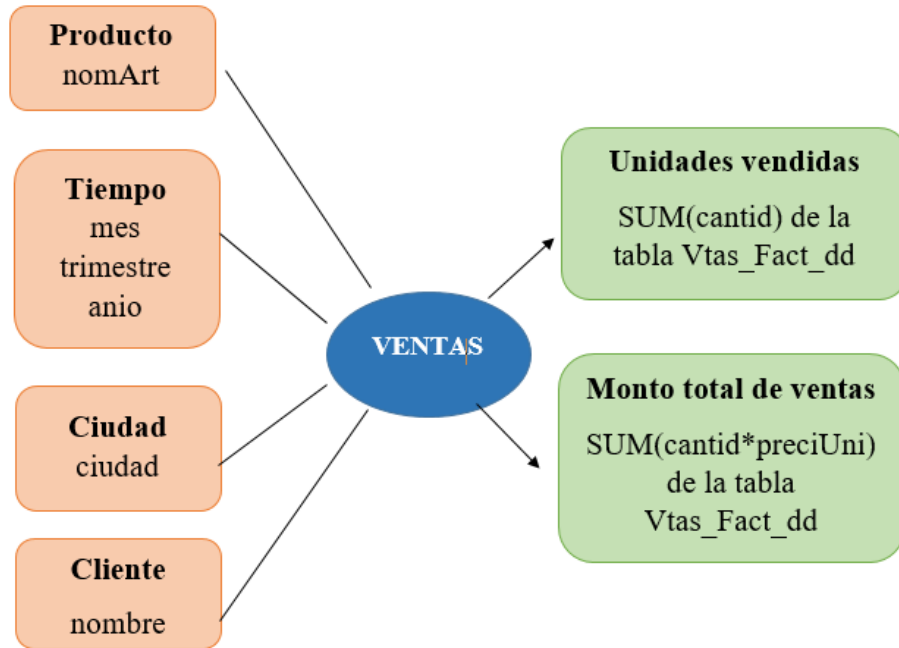


Ilustración 38. Modelo conceptual ampliado de ventas. Fuente: Christopher Torres y Adrián Carrión

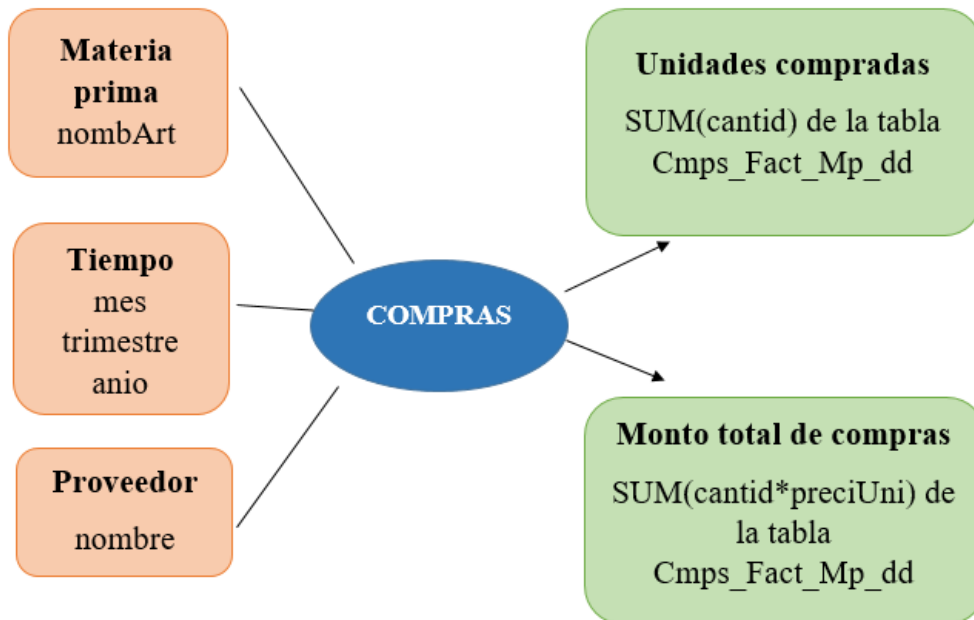


Ilustración 39. Modelo conceptual ampliado de compras. Fuente: Christopher Torres y Adrián Carrión

4.2.1.3 Modelo lógico del data warehouse

Se diseña la estructura del data warehouse partiendo del modelo conceptual. Para lo cual se debe definir el tipo de modelo lógico que se utilizara para tener en cuenta en la creación de las tablas de dimensiones y de hechos. Finalmente, se efectúan las uniones correspondientes entre las tablas.

4.2.1.3.1 Tipo de modelo lógico del data warehouse

Se indica el esquema a considerar para contener la estructura de almacenamiento de los datos; es importante definirlo de acuerdo a los requerimientos y necesidades de los usuarios finales, ya que esta decisión afecta directamente el diseño del modelo lógico del data warehouse.

El esquema utilizado es estrella, porque cumple con el alcance del diseño del data warehouse, ya que cada dimensión utiliza únicamente una tabla.

4.2.1.3.2 Tablas de dimensiones

Al aplicar el esquema estrella para la definición de las tablas de dimensiones, cada perspectiva de análisis del modelo conceptual constituye una tabla. Para ello se deben tomar en consideración los siguientes aspectos:

- Se debe elegir un nombre que representa a la tabla de dimensión.
- Cada tabla contendrá una clave principal que la identifique.
- Se pueden cambiar el nombre de los campos que no sean intuitivos.

A continuación, se presenta el diseño de las tablas de dimensiones:

Proceso de venta

- Perspectiva “Producto”
 - La tabla dimensión será nombrada “PRODUCTO”.
 - La clave principal con el nombre “idProducto”.
 - El campo “nombArt” se modifica con el nombre “producto”.

A continuación, se aprecia el resultado gráfico:

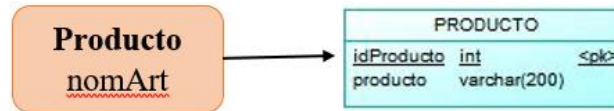


Ilustración 40. Tabla dimensión “PRODUCTO”. Fuente: Christopher Torres y Adrián Carrión

- Perspectiva “Tiempo”
 - La tabla dimensión será nombrada “FECHA”.
 - La clave principal con el nombre “idFecha”.

A continuación, se aprecia el resultado gráfico:

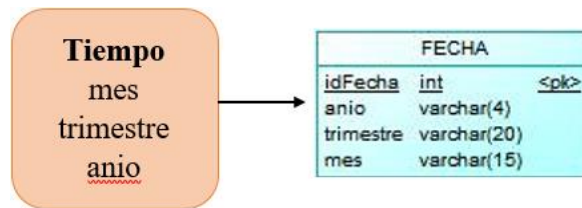


Ilustración 41. Tabla dimensión “FECHA”. Fuente: Christopher Torres y Adrián Carrión

- Perspectiva “Ciudad”
 - La tabla dimensión será nombrada “CIUDAD”.
 - La clave principal con el nombre “idCiudad”.

A continuación, se aprecia el resultado gráfico:

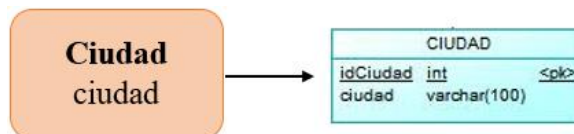


Ilustración 42. Tabla dimensión “CIUDAD”. Fuente: Christopher Torres y Adrián Carrión

- Perspectiva “Cliente”
 - La tabla dimensión será nombrada “CLIENTE”.
 - La clave principal con el nombre “idCliente”.

- El campo “nombre” se modifica con el nombre “cliente”.

A continuación, se aprecia el resultado gráfico:

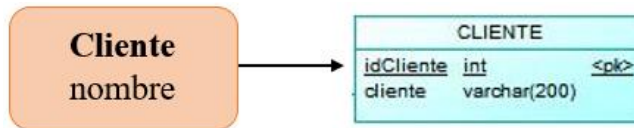


Ilustración 43. Tabla dimensión “CLIENTE”. Fuente: Christopher Torres y Adrián Carrión

Proceso de compra

- Perspectiva “Materia prima”
 - La tabla dimensión será nombrada “MATERIAPRIMA”.
 - La clave principal con el nombre “idMateriaPrima”.
 - El campo “nombArt” se modifica con el nombre “materiaPrima”.

A continuación, se aprecia el resultado gráfico:

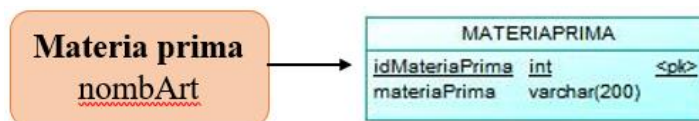


Ilustración 44. Tabla dimensión “MATERIAPRIMA”. Fuente: Christopher Torres y Adrián Carrión

- Perspectiva “Tiempo”
 - La tabla dimensión será nombrada “FECHA”.
 - La clave principal con el nombre “idFecha”.

A continuación, se aprecia el resultado gráfico:

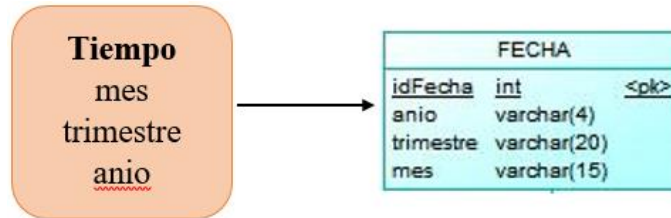


Ilustración 45. Tabla dimensión “FECHA”. Fuente: Christopher Torres y Adrián Carrión

- Perspectiva “Proveedor”
 - La tabla dimensión será nombrada “PROVEEDOR”.
 - La clave principal con el nombre “idProveedor”.
 - El campo “nombre” se modifica con el nombre “proveedor”.

A continuación, se aprecia el resultado gráfico:

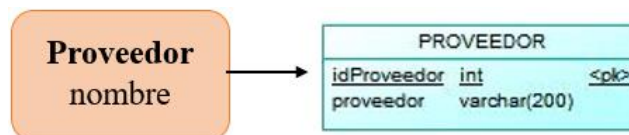


Ilustración 46. Tabla dimensión “PROVEEDOR”. Fuente: Christopher Torres y Adrián Carrión

4.2.1.3.3 Tablas de hechos

Para la construcción de las tablas de hechos se consideran los siguientes aspectos:

- Se debe asignar un nombre a la tabla de hechos que represente la información analizada.
- Cada tabla de hecho debe contener una clave primaria, esta se compone de la combinación de las claves primarias de cada tabla de dimensión con la que se relaciona.
- Cada tabla de hecho debe contener los campos que se propusieron en los indicadores del modelo conceptual y se les asigna un nombre.

A continuación, se presenta el diseño de las tablas de hechos:

Proceso venta

- La tabla de hechos tiene el nombre “VENTAS”.
- La clave primaria es la combinación de las claves: “idProducto”, “idFecha”, “idCiudad” y “idCliente”.
- Contiene dos campos correspondientes a los indicadores, los cuales tienen el nombre “cantidad” y “montoTotal”.

A continuación, se aprecia el resultado gráfico:

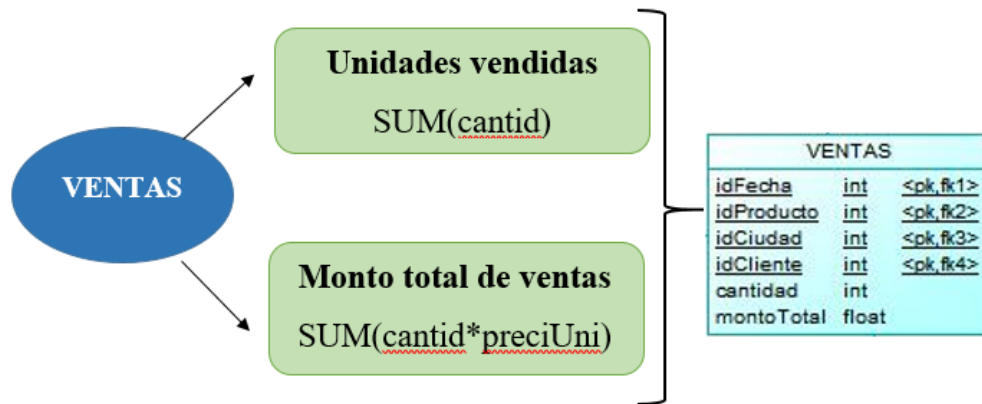


Ilustración 47. Tabla de hechos “VENTAS”. Fuente: Christopher Torres y Adrián Carrión

Proceso compra

- La tabla de hechos tiene el nombre “COMPRAS”.
- La clave primaria es la combinación de las claves: “idMateriaPrima”, “idFecha” y “idProveedor”.
- Contiene dos campos correspondientes a los indicadores, los cuales tienen el nombre “cantidad” y “montoTotal”.

A continuación, se aprecia el resultado gráfico:

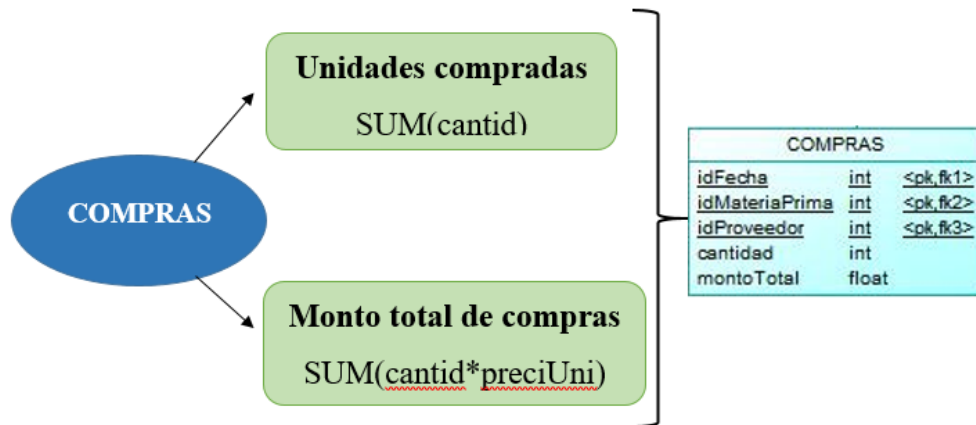


Ilustración 48. Tabla de hechos “COMPRAS”. Fuente: Christopher Torres y Adrián Carrión

4.2.1.3.4 Uniones

En este paso se presenta el modelo lógico del data warehouse, que compone en este caso la unión de dos data marts correspondientes a ventas y compras. En las siguientes figuras se muestra las uniones pertinentes de acuerdo a lo establecido en los pasos anteriores:

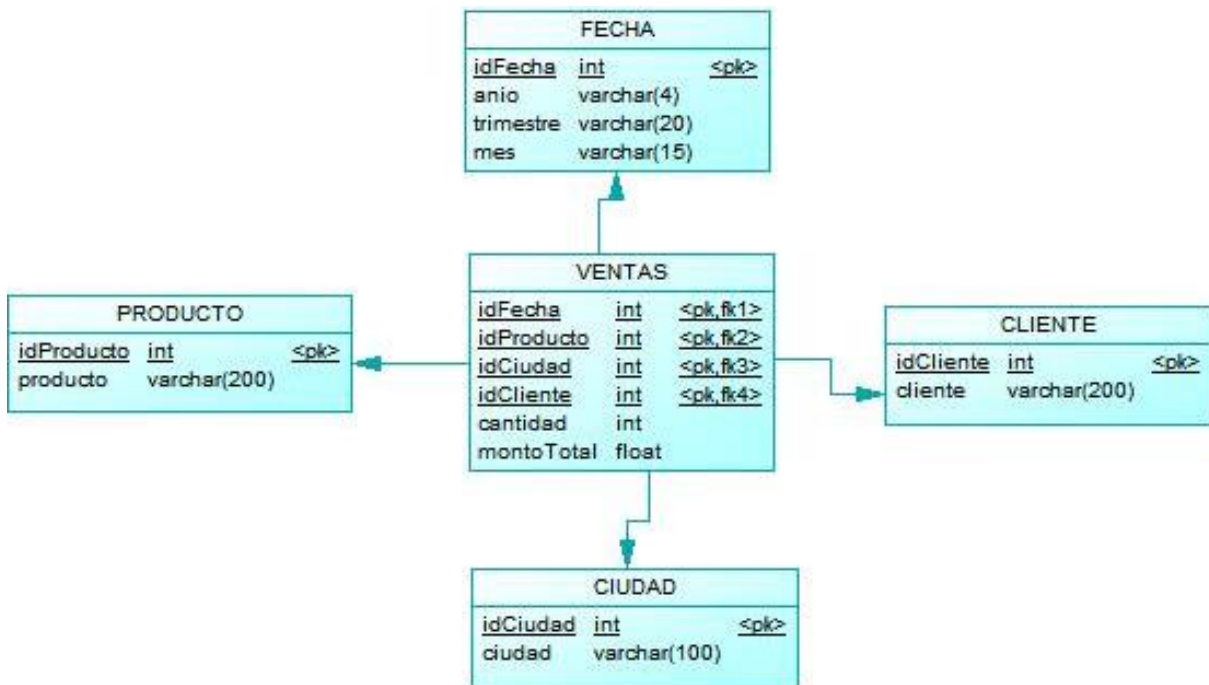


Ilustración 49. Uniones del proceso de venta. Fuente: Christopher Torres y Adrián Carrión

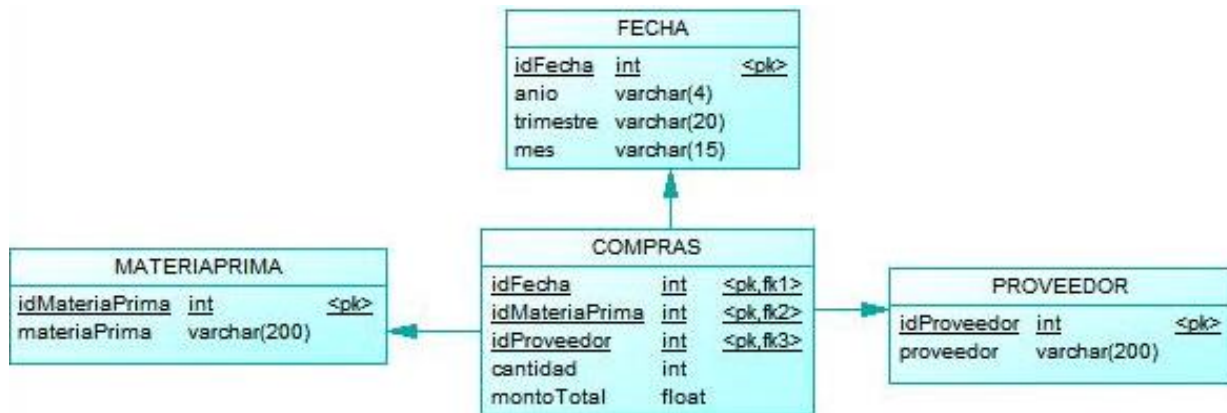


Ilustración 50. Uniones del proceso de compra. Fuente: Christopher Torres y Adrián Carrión

4.2.1.4 Integración de datos

En esta etapa se procede a probar el modelo lógico del data warehouse mediante la carga inicial de datos, utilizando técnicas de limpieza, procesos ETL, etc.

Para la integración de datos, se tomó la decisión de migrar de motor de base de datos que utiliza el sistema de gestión empresarial "SIGEMP 3.0", el cual almacena sus datos en Sybase Adaptive Server Anywhere versión 9.0.2.2451; se seleccionó MySQL versión 5.7.14 para almacenar la base de datos temporal, ya que este sistema gestor de base de datos relacional es compatible con todas las herramientas utilizadas de la Suite de Pentaho Community Edition (CE) para los posteriores pasos que se realiza. Para más información sobre la migración de datos, revisar el manual técnico²³. Los datos tomados para la aplicación de esta fase van de la fecha de enero del 2010 hasta marzo del 2016.

4.2.1.4.1 Carga inicial

En este paso se realiza un proceso de extracción de datos, transformación y carga en el data warehouse; para ello se ejecuta tareas que contienen cierta lógica para obtener los datos que son necesarios. La herramienta que se utiliza

²³ Ver anexo A

para esta tarea es Spoon, el cual es el diseñador gráfico de procesos ETL de Pentaho Data Integration, también conocido como Kettle.

Para la carga inicial de datos en Spoon, se selecciona una nueva transformación, luego se realiza dos conexiones a MySQL, una para la base de datos de la empresa y la otra para el data warehouse. Las siguientes imágenes muestran las conexiones:

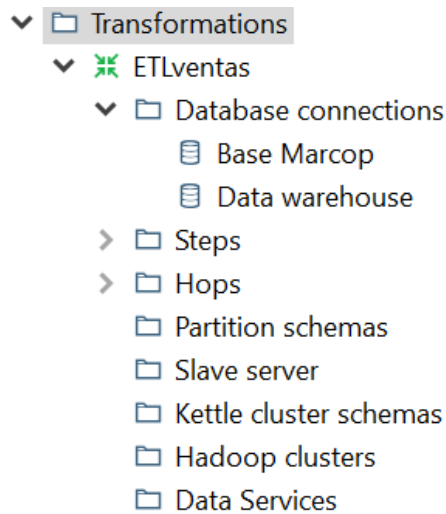


Ilustración 51. Vista general de la transformación. Fuente: Christopher Torres y Adrián Carrión

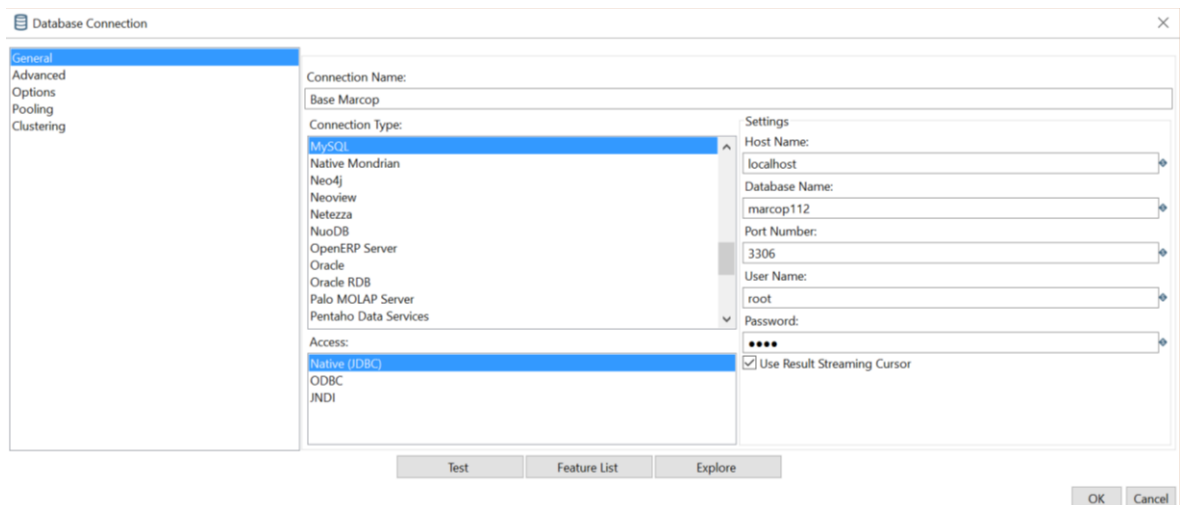


Ilustración 52. Conexión a la base de datos temporal de la empresa. Fuente: Christopher Torres y Adrián Carrión

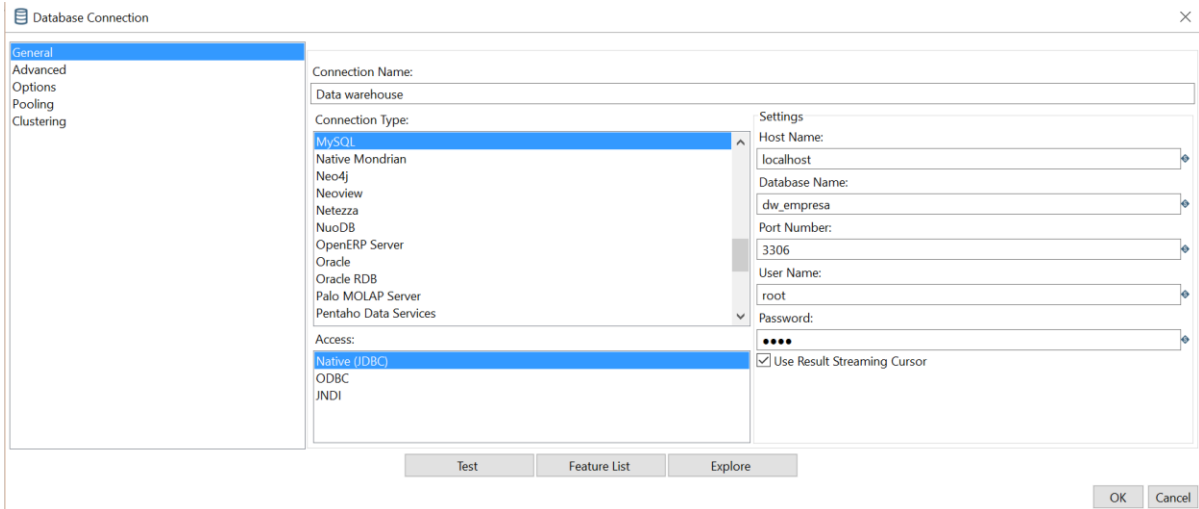


Ilustración 53. Conexión al data warehouse. Fuente: Christopher Torres y Adrián Carrión

La conexión a la base de datos temporal de la empresa y del data warehouse se realiza para cada transformación del data mart de ventas y compras.

Se empieza por cargar los datos de las dimensiones y luego las tablas de hechos, es importante tener en cuenta la correcta relación entre los elementos. Las siguientes imágenes presentan el proceso ETL para la carga inicial de datos del data mart de ventas:

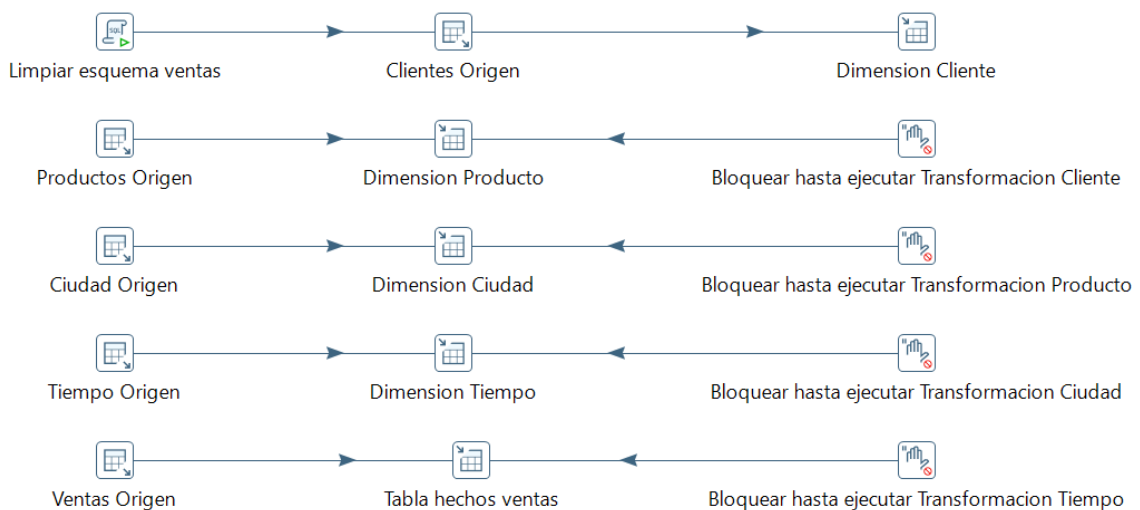


Ilustración 54. Carga inicial del data mart de ventas. Fuente: Christopher Torres y Adrián Carrión

Las tareas que lleva a cabo este proceso ETL para ventas son:

- Limpiar esquema ventas: ejecuta un proceso almacenado previamente creado en el data mart de ventas, que realiza un TRUNCATE de las tablas de dimensiones y la tabla de hechos ventas.
- Clientes Origen: ejecuta el siguiente script SQL para obtener los datos necesarios de clientes.

```
SELECT codCli AS IDCLIENTE,  
       nombre AS CLIENTE  
FROM clientes  
WHERE estadoReg='ACT'  
ORDER BY IDCLIENTE;
```

- Dimensión Cliente: se selecciona la tabla cliente del data warehouse y se relaciona los campos de la tabla origen con los de la tabla dimensión para la carga de datos.
- Cliente: ejecuta el siguiente script SQL para obtener los datos correspondientes de artículos.

```
SELECT codArt AS IdPRODUCTO,  
       nomArt AS PRODUCTO  
FROM articulos  
WHERE estadoReg='ACT'  
ORDER BY IDPRODUCTO;
```

- Bloquear hasta ejecutar Transformación Cliente: controla la carga de datos de la dimensión Producto.
- Dimensión Producto: se selecciona la tabla producto del data warehouse y se relaciona los campos de la tabla origen con los de la tabla dimensión para la carga de datos.
- Ciudad Origen: ejecuta el siguiente script SQL para obtener los datos correspondientes de las ciudades.

```

SELECT DISTINCT CONCAT(replace(SUBSTR(ciudad,1,1),
',',''),
replace(SUBSTR(ciudad,3,3),' ','')) AS IDCIUDAD,
ciudad
FROM clientes
WHERE estadoReg='ACT'
ORDER BY IDCIUDAD;

```

- Bloquear hasta ejecutar Transformación Producto: controla la carga de datos de la dimensión Ciudad.
- Dimensión Ciudad: se selecciona la tabla ciudad del data warehouse y se relaciona los campos de la tabla origen con los de la tabla dimensión para la carga de datos.
- Tiempo Origen: ejecuta el siguiente script SQL para obtener las fechas de las correspondientes ventas, con nivel de detalle de año, trimestre y mes.

```

SELECT DISTINCT
YEAR(fchComp) * 10000 + MONTH(fchComp) * 100 +
DAY(fchComp) AS CodigoFecha,
YAR(fchComp) AS ANIO,
CASE QUARTER(fchComp)
WHEN 1 THEN '1ER TRIMESTRE'
WHEN 2 THEN '2DO TRIMESTRE'
WHEN 3 THEN '3ER TRIMESTRE'
WHEN 4 THEN '4TO TRIMESTRE'
END AS TRIMESTRE,
CASE MONTH(fchComp)
WHEN 1 THEN 'ENERO'
WHEN 2 THEN 'FEBRERO'
WHEN 3 THEN 'MARZO'
WHEN 4 THEN 'ABRIL'
WHEN 5 THEN 'MAYO'
WHEN 6 THEN 'JUNIO'
WHEN 7 THEN 'JULIO'
WHEN 8 THEN 'AGOSTO'
WHEN 9 THEN 'SEPTIEMBRE'
WHEN 10 THEN 'OCTUBRE'
WHEN 11 THEN 'NOVIEMBRE'
WHEN 12 THEN 'DICIEMBRE'
END AS MES
FROM vtas_fact_dg
WHERE (fchComp IS NOT NULL) AND
(vtas_fact_dg.estadoReg='ACT')
ORDER BY CodigoFecha;

```

- Bloquear hasta ejecutar Transformación Ciudad: controla la carga de datos de la dimensión Tiempo.

- Dimensión Tiempo: se selecciona la tabla fecha del data warehouse y se relaciona los campos de la tabla origen con los de la tabla dimensión para la carga de datos.
- Ventas Origen: se ejecuta el siguiente script SQL para la obtención de los datos de las ventas.

```

SELECT YEAR(vtas_fact_dg.fchComp) * 10000 +
MONTH(vtas_fact_dg.fchComp) * 100 +
DAY(vtas_fact_dg.fchComp) AS IDFECHA,
articulos.codArt AS IDPRODUCTO,
clientes.codCli AS IDCLIENTE,
CONCAT(replace(SUBSTR(clientes.ciudad,1,1),
','),replace(SUBSTR(clientes.ciudad,3,3),',' ,')) as
IDCIUDAD,
truncate(vtas_fact_dd.cantid,0) as CANTIDAD,
ROUND((vtas_fact_dd.cantid*vtas_fact_dd.precUni),2) as
MONTOTOTAL
FROM vtas_fact_dg
inner join clientes on clientes.codCli =
vtas_fact_dg.codCli
inner join vtas_fact_dd on (vtas_fact_dd.codEmp =
vtas_fact_dg.codEmp) AND (vtas_fact_dd.codComp =
vtas_fact_dg.codComp) AND (vtas_fact_dd.numComp =
vtas_fact_dg.numComp)
inner join articulos on articulos.codArt =
vtas_fact_dd.codArt
where (vtas_fact_dg.fchComp IS NOT NULL) AND
(vtas_fact_dg.estadoReg='ACT') AND
(articulos.estadoReg='ACT') AND (clientes.estadoReg='ACT')
ORDER BY IDFECHA;

```

- Bloquear hasta ejecutar Transformación Tiempo: controla la carga de datos de la tabla de hechos ventas.
- Tabla de hechos ventas: se selecciona la tabla ventas del data warehouse y se relaciona los campos de la tabla origen con los de la tabla de hechos para la carga de datos.

A continuación, se muestra la imagen que representa la carga inicial de datos del data mart de compras:

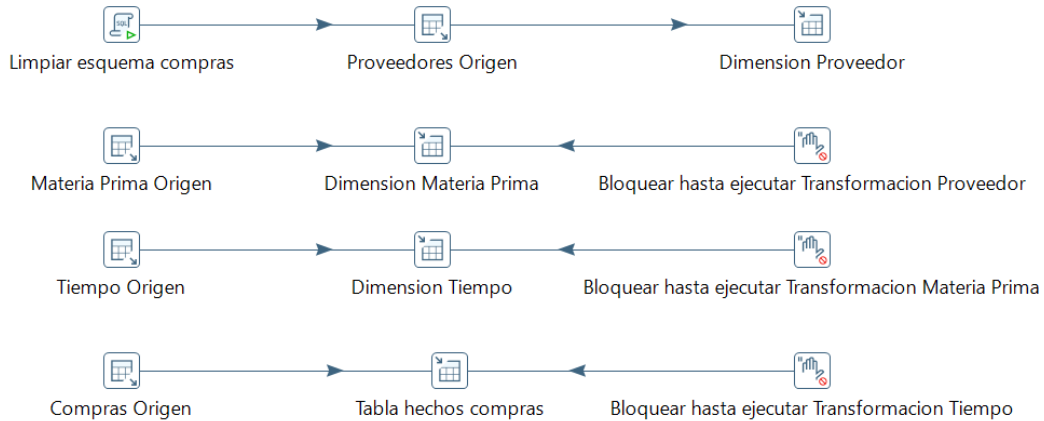


Ilustración 55. Carga inicial del data mart de compras. Fuente: Christopher Torres y Adrián Carrión

Las tareas que lleva a cabo este proceso ETL para compras son:

- Limpiar esquema compras: ejecuta un proceso almacenado previamente creado en el data mart de compras, que realiza un TRUNCATE de las tablas de dimensiones y la tabla de hechos compras.
- Proveedores Origen: ejecuta el siguiente script SQL para obtener los datos de los proveedores.

```

SELECT codProv as idProveedor,
       nombre as proveedor
FROM proveedores
WHERE (nombre IS NOT NULL)
AND (nombre != 'NO DEFINIDO')
AND (estadoReg='ACT')
ORDER BY idProveedor;
  
```

- Dimensión Proveedor: se selecciona la tabla proveedor del data warehouse y se relaciona los campos de la tabla origen con los de la tabla dimensión para la carga de datos.
- Material Prima Origen: ejecuta el siguiente script SQL para obtener los datos de los artículos de materia prima.

```

SELECT codArt as idMateriaPrima,
       concat(nomArt, ' ', nomUni) as materiaPrima
FROM articulos_mp
INNER JOIN unidades on unidades.codUni =
articulos_mp.codUni
WHERE (articulos_mp.nomArt IS NOT NULL)
AND (articulos_mp.estadoReg='ACT')
ORDER BY idMateriaPrima;

```

- Bloquear hasta ejecutar Transformación Proveedor: controla la carga de datos de la dimensión Materia Prima.
- Dimensión Materia Prima: se selecciona la tabla materiaprime del data warehouse y se relaciona los campos de la tabla origen con los de la tabla dimensión para la carga de datos.
- Tiempo Origen: ejecuta el siguiente script SQL para obtención de las fechas de compras de materia prima, detallando el año, trimestre y mes.

```

SELECT DISTINCT
YEAR(fchComp) * 10000 + MONTH(fchComp) * 100 + DAY(fchComp)
AS CodigoFecha,
YEAR(fchComp) AS Anio,
CASE QUARTER(fchComp)
WHEN 1 THEN '1ER TRIMESTRE'
WHEN 2 THEN '2DO TRIMESTRE'
WHEN 3 THEN '3ER TRIMESTRE'
WHEN 4 THEN '4TO TRIMESTRE'
END AS Trimestre,
CASE MONTH(fchComp)
WHEN 1 THEN 'ENERO'
WHEN 2 THEN 'FEBRERO'
WHEN 3 THEN 'MARZO'
WHEN 4 THEN 'ABRIL'
WHEN 5 THEN 'MAYO'
WHEN 6 THEN 'JUNIO'
WHEN 7 THEN 'JULIO'
WHEN 8 THEN 'AGOSTO'
WHEN 9 THEN 'SEPTIEMBRE'
WHEN 10 THEN 'OCTUBRE'
WHEN 11 THEN 'NOVIEMBRE'
WHEN 12 THEN 'DICIEMBRE'
END AS Mes
FROM cmpr_fact_mp_dg
WHERE (fchComp IS NOT NULL)
ORDER BY CodigoFecha;

```

- Bloquear hasta ejecutar Transformación Materia Prima: controla la carga de datos de la dimensión Tiempo.

- Dimensión Tiempo: se selecciona la tabla fecha del data warehouse y se relaciona los campos de la tabla origen con los de la tabla dimensión para la carga de datos.
- Compras Origen: se ejecuta el siguiente script SQL para obtener los datos de las compras de materia prima.

```

SELECT proveedores.codProv as IDPROVEEDOR,
articulos_mp.codArt as IDMATERIAPRIMA,
YEAR(cmps_fact_mp_dg.fchComp) * 10000 +
MONTH(cmps_fact_mp_dg.fchComp) * 100 +
DAY(cmps_fact_mp_dg.fchComp) AS IDFECHA,
truncate(cmps_fact_mp_dd.cantid,0) as CANTIDAD,
ROUND((cmps_fact_mp_dd.cantid*cmps_fact_mp_dd.precUni),2)
as MONTOTOTAL
FROM cmps_fact_mp_dg
inner join proveedores on proveedores.codProv =
cmps_fact_mp_dg.codProv
inner join cmps_fact_mp_dd on (cmps_fact_mp_dd.codEmp =
cmps_fact_mp_dg.codEmp) AND (cmps_fact_mp_dd.codComp =
cmps_fact_mp_dg.codComp) AND (cmps_fact_mp_dd.numComp =
cmps_fact_mp_dg.numComp)
inner join articulos_mp on articulos_mp.codArt =
cmps_fact_mp_dd.codArt
where (cmps_fact_mp_dg.fchComp IS NOT NULL) AND
(cmps_fact_mp_dg.estadoReg='ACT')AND
(proveedores.estadoReg='ACT') AND
(articulos_mp.estadoReg='ACT')
ORDER BY IDFECHA;

```

- Bloquear hasta ejecutar Transformación Tiempo: controla la carga de datos de la tabla de hechos compras.
- Tabla de hechos compras: se selecciona la tabla compras del data warehouse y se relaciona los campos de la tabla origen con los de la tabla de hechos para la carga de datos.

4.3 Creación de Cubos multidimensionales

El cubo multidimensional nos permite presentar los datos que se encuentran almacenados en filas y columnas, en una matriz de N dimensiones. La herramienta utilizada para la elaboración de los cubos en formato XML es Schema Workbench; dentro de los elementos más importantes para la construcción de los cubos multidimensionales tenemos:

- Atributos: se refiere a los criterios de análisis de las tablas de dimensiones.
- Jerarquías: representa el nivel de granularidad de las tablas de dimensiones.
- Indicadores: son las sumalizaciones que se realizan sobre las tablas de hechos.

En la siguiente figura se muestra el cubo multidimensional de ventas:

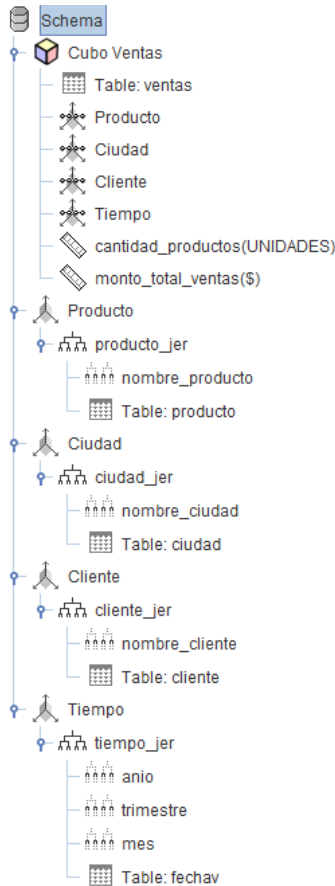


Ilustración 56. Cubo multidimensional de ventas. Fuente: Christopher Torres y Adrián Carrión

Para la elaboración del cubo multidimensional de ventas, se empezó agregando las dimensiones correspondientes a producto, cliente, ciudad y tiempo. Para cada dimensión se agregó una jerarquía que representa el nivel de detalle, en el caso de producto, cliente y ciudad su jerarquía es el atributo del nombre, para el caso del tiempo se toma en cuenta un nivel de detalle más específico, el año, trimestre y mes correspondiente a cada venta. Al final se agrega el cubo de ventas con los indicadores de cantidad_productos que se obtiene de la SUM(ventas.cantidad) y el monto _total de SUM(ventas.montotal), además

se relaciona las dimensiones con el cubo. La siguiente imagen representa el cubo multidimensional de compras:

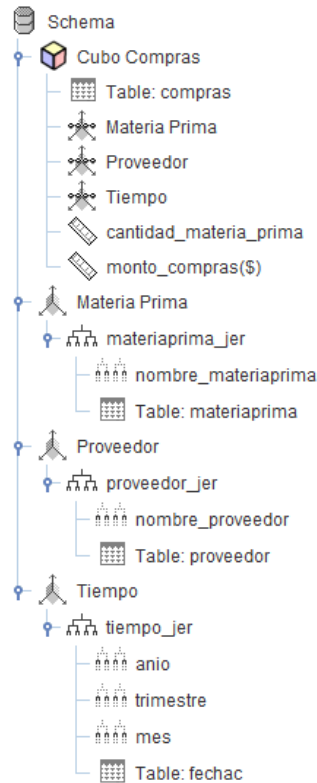


Ilustración 57. Cubo multidimensional de compras. Fuente: Christopher Torres y Adrián Carrión

4.4 Presentación de información y generación de reportes

Para la presentación de información, se utiliza Business Analytics Platform, el servidor de Pentaho para el análisis avanzado de datos a través del uso de múltiples herramientas que se pueden descargar e implementar conjuntamente. Para el proyecto se considera emplear el plug-in Saiku Analytics, ya que permite a los usuarios explorar fuentes de datos complejas de forma fácil, simplemente arrastrando los campos que necesite visualizar el usuario desde diferentes perspectivas. A continuación, se presenta una imagen de cómo se visualiza los datos mediante la utilización del plug-in de Saiku.

anio	nombre_cliente	nombre_producto	cantidad_productos(UNIDADES)	monto_total_ventas(\$)
2010	ACOSTA BRIGIDA	LECHE 1 LITRO	48	28.8
		DE LA A ENRIQUE	1	3.5
	YOGURT 1 GALON DE MORA	YOGURT 1 GALON DE FRESA	2	7
		YOGURT 1 GALON DE BOROJO	1	3.5
	YOGURT 2 LITROS DE MORA	YOGURT 2 LITROS DE FRESA	49	104.1
		YOGURT 2 LITROS DE DURAZNO	97	206.3
	YOGURT 2 LITROS DE BOROJO	YOGURT 2 LITROS DE DURAZNO	41	87.4
		YOGURT 2 LITROS DE BOROJO	18	38.2
	YOGURT 2 LITROS DE NATURAL	YOGURT 100 CC DE CEREAL VARIOS SABORES	6	12.9
		YOGURT PACA 20 BOLOS 100 CC VARIOS SABORES	120	36
ALEAGA RODRIGO	LECHE 1 LITRO	20	23	
	LECHE 1/2 LITRO	475	251.75	
ANDRADE TAPIA WILMER	QUESO MOZZARELLA 500 G	50	13.75	
	YOGURT 1 GALON DE MORA	20	58	
ARBOLEDA YADIRA	YOGURT 1 GALON DE MORA	6	20.4	
	YOGURT 1 GALON DE FRESA	8	27.2	
YOGURT 1 GALON DE DURAZNO	YOGURT 1 GALON DE DURAZNO	6	20.4	
	YOGURT 2 LITROS DE MORA	32	67.2	
YOGURT 2 LITROS DE MORA	YOGURT 2 LITROS DE FRESA	40	84	
	YOGURT 2 LITROS DE DURAZNO	28	58.8	
YOGURT 1 GALON DE MORA	YOGURT 1 GALON DE FRESA	18	60.05	
	YOGURT 1 GALON DE FRESA	26	86.45	
YOGURT 1 GALON DE DURAZNO	YOGURT 1 GALON DE DURAZNO	15	52.4	
	YOGURT 1 GALON DE MANGO	1	3.2	
YOGURT 1 GALON DE NATURAL	YOGURT 2 LITROS DE MORA	2	6.75	
	YOGURT 2 LITROS DE MORA	17	35.04	
YOGURT 2 LITROS DE FRESA	YOGURT 2 LITROS DE FRESA	26	53.18	
	YOGURT 2 LITROS DE DURAZNO	16	33.04	
YOGURT 2 LITROS DE MANGO	YOGURT 2 LITROS DE MANGO	5	9.84	
	YOGURT 2 LITROS DE GUANABANA	5	9.84	
YOGURT 2 LITROS DE BOROJO	YOGURT 2 LITROS DE BOROJO	5	9.99	

Ilustración 58. Cubo de ventas mediante Saiku. Fuente: Christopher Torres y Adrián Carrión

En la imagen anterior se puede visualizar la interfaz que ofrece el plug-in Saiku Analytics para la visualización de los cubos multidimensionales. En la parte superior izquierda, se elige el cubo que se desea presentar, conjuntamente se despliegan las dimensiones y medidas y se las arrastra a las casillas correspondientes a filas, columnas y medidas para que se ejecute la consulta. Además, el plug-in permite exportar la información a formato XLS, CSV y PDF, también permite visualizar la información gráficamente.

Para la generación de reportes se utiliza la herramienta Pentaho Report Designer, el cual permite personalizar los reportes mediante la utilización de parámetros para que el usuario pueda seleccionar los datos que necesite. Para más información de cómo generar un reporte ver el manual de usuario²⁴. A continuación, se presentan las imágenes con los respectivos reportes:

²⁴ Ver anexo B

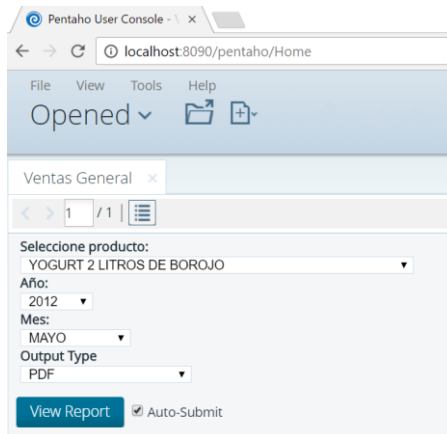


Ilustración 59. Parámetros para reporte general de ventas. Fuente: Christopher Torres y Adrián Carrión

septiembre 14, 2016 @ 10:56

LACTEOS DE MARCO'S

Reporte de ventas general

Producto: YOGURT 2 LITROS DE BOROJO **Cantidad vendida:** 132
Año: 2012 **Ingresos(S):** 289,26
Mes: MAYO

Cliente	Ciudad	#	\$
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	4	8,8
TORRES LUIS	MACHALA	1	2,09
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2
TIENDAS INDUSTRIALES ASOCIADAS TIA S.A.	GUAYAQUIL	1	2,2

septiembre 14, 2016 @ 10:56

Ilustración 60. Reporte general de ventas. Fuente: Christopher Torres y Adrián Carrión

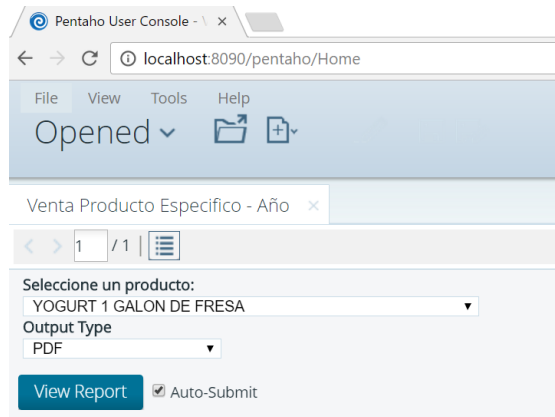


Ilustración 61. Parámetros para reporte de venta de un producto específico por año. Fuente: Christopher Torres y Adrián Carrión

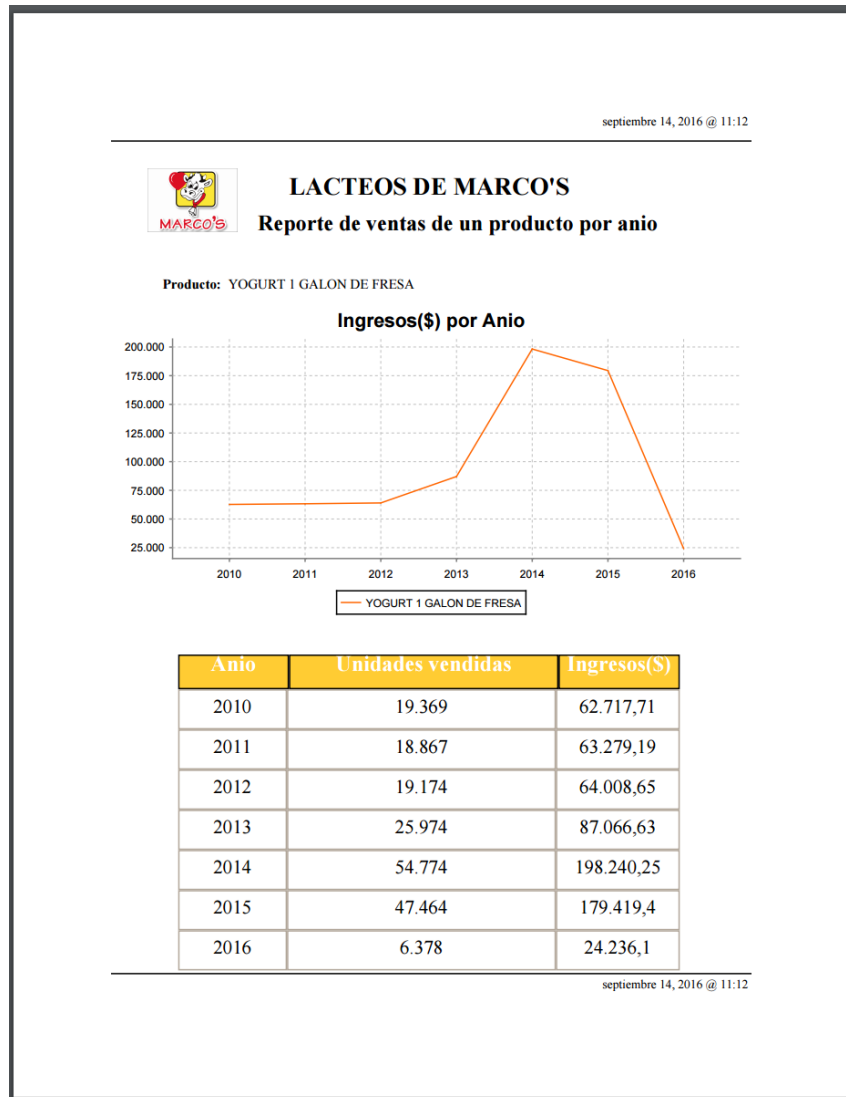


Ilustración 62. Reporte de venta de un producto específico por año. Fuente: Christopher Torres y Adrián Carrión

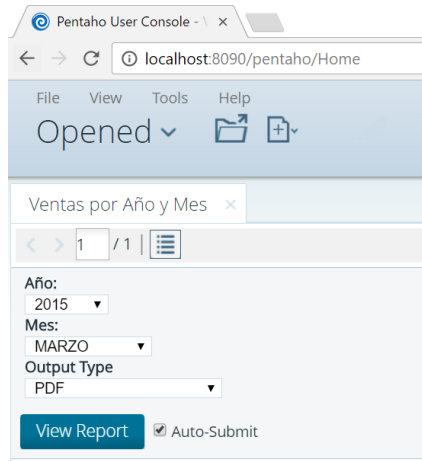


Ilustración 63. Parámetros para reporte de venta de todos los productos por año y mes. Fuente: Christopher Torres y Adrián Carrión

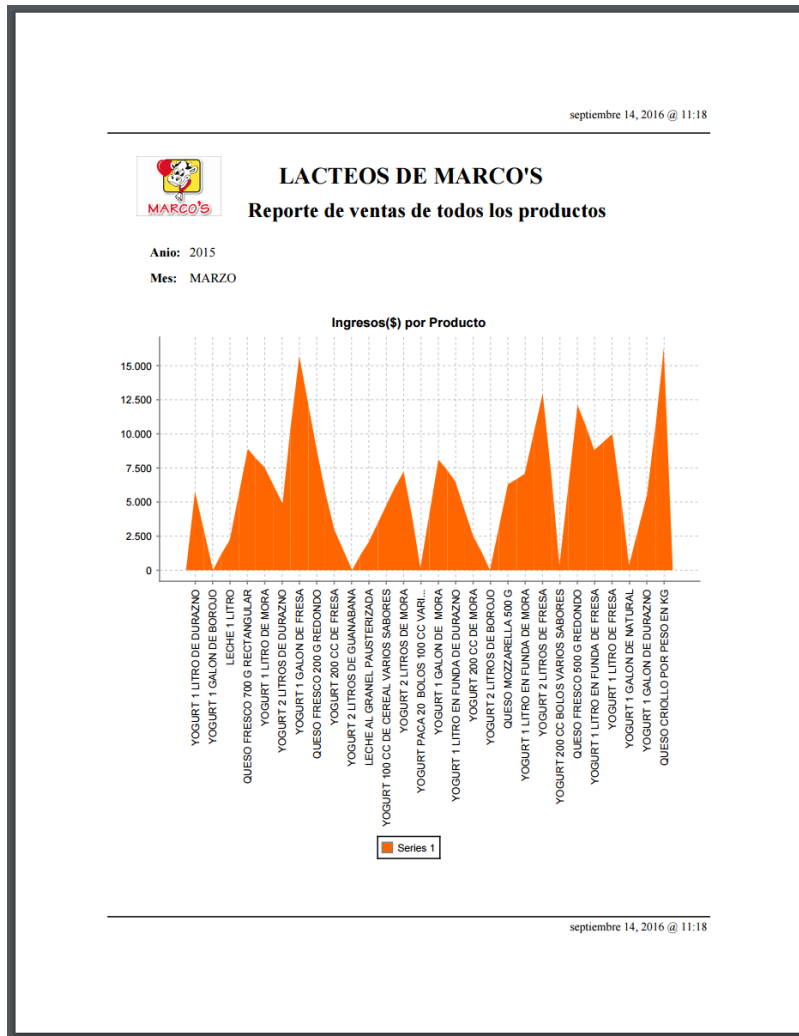


Ilustración 64. Reporte de ventas de todos los productos por año y mes. Fuente: Christopher Torres y Adrián Carrión



LACTEOS DE MARCO'S

Reporte de ventas de todos los productos

Anio: 2015

Mes: MARZO

Producto	Uni. vendidas	Ingresos(\$)
YOGURT 1 LITRO DE DURAZNO	4.431	5.760,2
YOGURT 1 GALON DE BOROJO	1	3,4
LECHE 1 LITRO	3.380	2.230,4
QUESO FRESCO 700 G RECTANGULAR	2.927	8.909,1
YOGURT 1 LITRO DE MORA	5.805	7.546,5
YOGURT 2 LITROS DE DURAZNO	1.964	4.888,64
YOGURT 1 GALON DE FRESA	4.139	15.704,43
QUESO FRESCO 200 G REDONDO	9.040	8.588
YOGURT 200 CC DE FRESA	8.401	3.023,36
YOGURT 2 LITROS DE GUANABANA	4	8,4
LECHE AL GRANEL PAUSTERIZADA	3.480	2.088
YOGURT 100 CC DE CEREAL VARIOS SABORES	13.488	4.855,68
YOGURT 2 LITROS DE MORA	2.902	7.223,44

Ilustración 65. Tabla reporte de ventas de todos los productos por año y mes. Fuente: Christopher Torres y Adrián Carrión

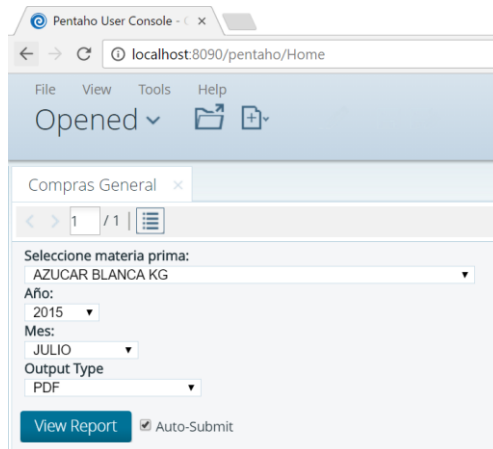



Ilustración 66. Tabla reporte de ventas de todos los productos por año y mes. Fuente: Christopher Torres y Adrián Carrión

septiembre 14, 2016 @ 11:26



LACTEOS DE MARCO'S

Reporte de compras general

Materia Prima: AZUCAR BLANCA KG **Cantidad comprada:** 15.000
Año: 2015 **Egresos(\$):** 11.240
Mes: JULIO

Año	Mes	Proveedor	#	S
2015	JULIO	SALCEDO FABIAN	2.500	1.875
2015	JULIO	SALCEDO FABIAN	2.500	1.885
2015	JULIO	SALCEDO FABIAN	5.000	3.750
2015	JULIO	SALCEDO FABIAN	5.000	3.730

septiembre 14, 2016 @ 11:26

Ilustración 67. Reporte general de compras. Fuente: Christopher Torres y Adrián Carrión

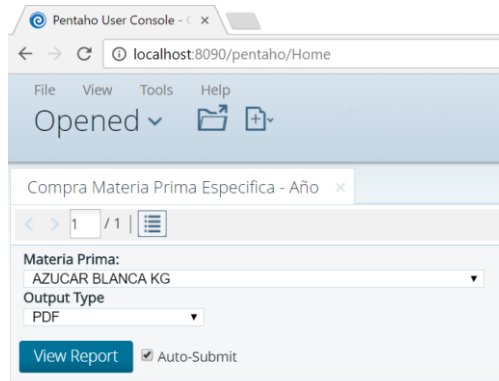


Ilustración 68. Parámetros para reporte de compra de materia prima específica por año. Fuente: Christopher Torres y Adrián Carrión

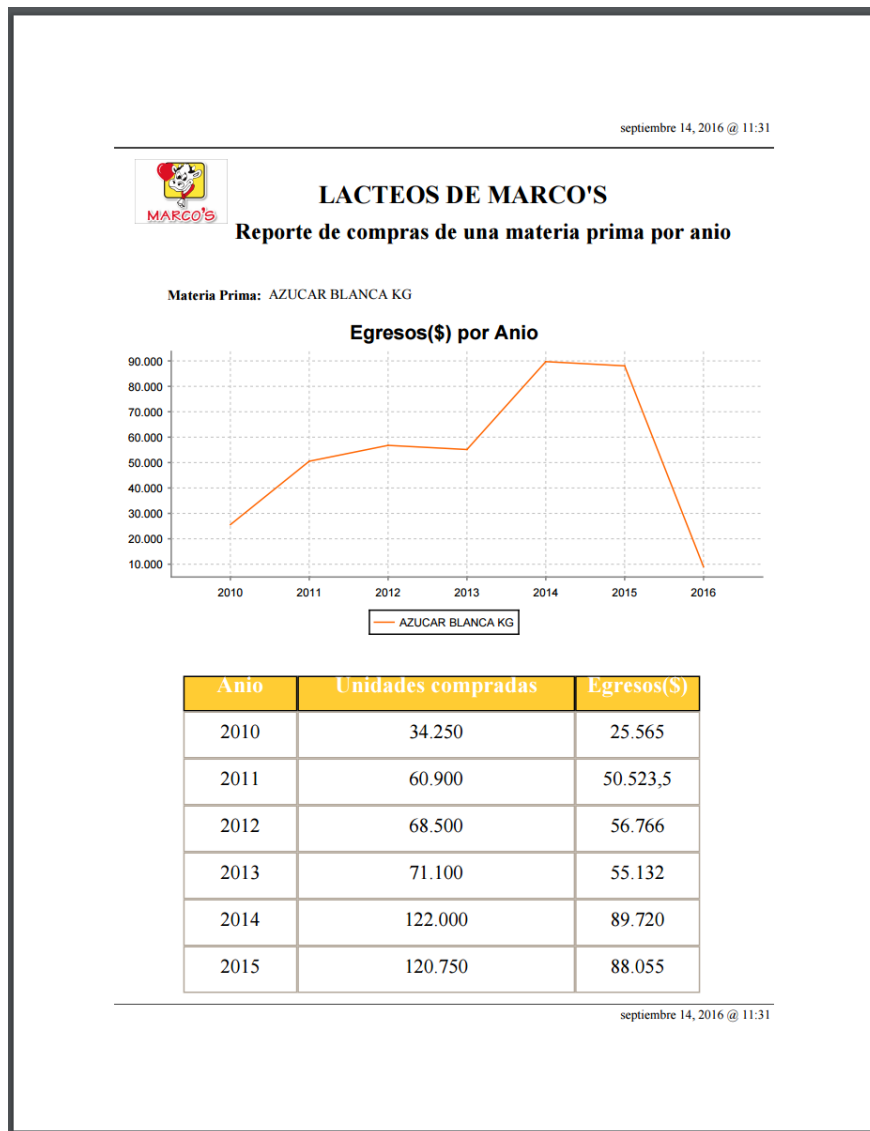


Ilustración 69. Reporte de compra de materia prima específica por año. Fuente: Christopher Torres y Adrián Carrión

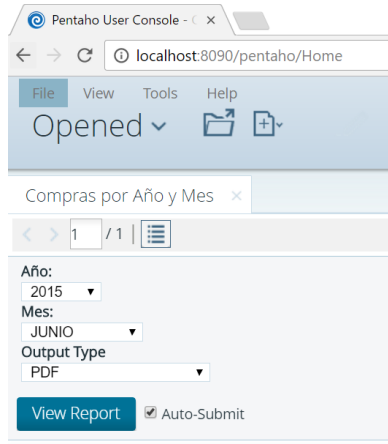


Ilustración 70. Parámetros para reporte de compra de toda la materia prima por año y mes. Fuente: Christopher Torres y Adrián

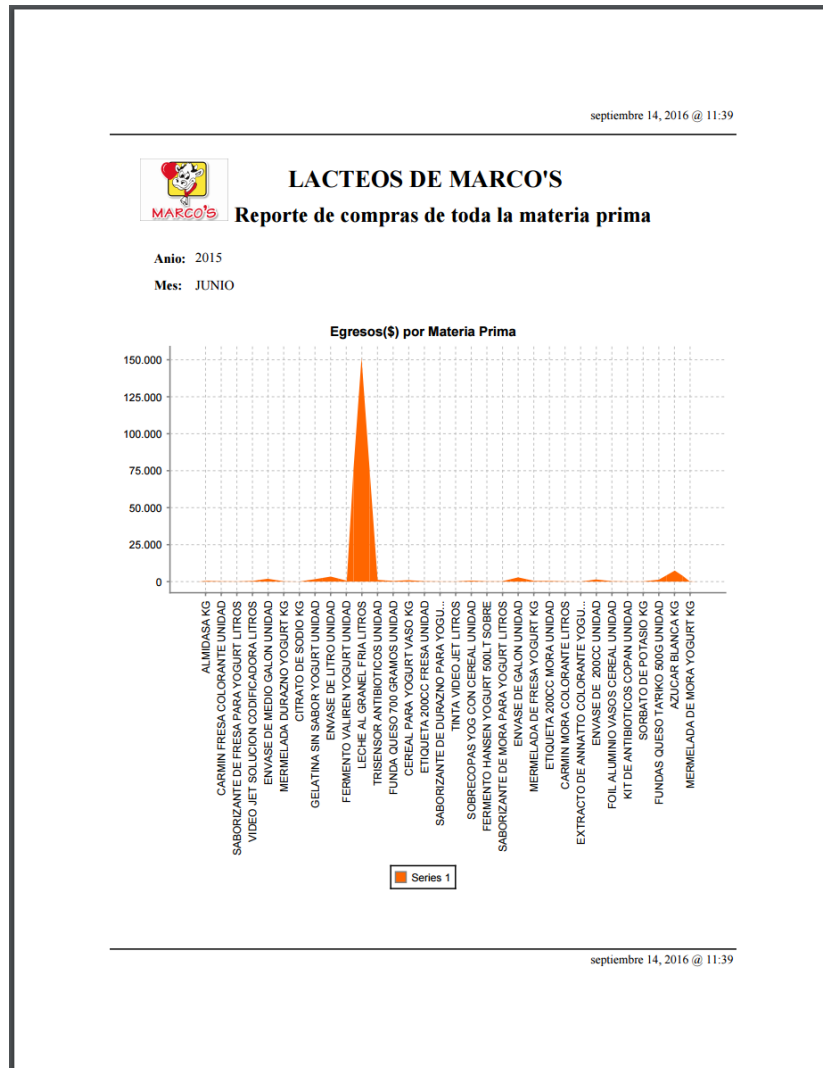


Ilustración 71. Reporte de compra de toda la materia prima por año y mes. Fuente: Christopher Torres y Adrián Carrión



LACTEOS DE MARCO'S

Reporte de compras de toda la materia prima

Anio: 2015

Mes: JUNIO

Materia Prima	Uni. compradas	Egresos(\$)
ALMIDASA KG	500	480
CARMIN FRESA COLORANTE UNIDAD	16	256
SABORIZANTE DE FRESA PARA YOGURT LITROS	8	164
VIDEO JET SOLUCION CODIFICADORA LITROS	9	466,55
ENVASE DE MEDIO GALON UNIDAD	8.000	1.920
MERMELADA DURAZNO YOGURT KG	120	224,39
CITRATO DE SODIO KG	50	67,5
GELATINA SIN SABOR YOGURT UNIDAD	200	1.600
ENVASE DE LITRO UNIDAD	18.000	3.420
FERMENTO VALIREN YOGURT UNIDAD	100	625
LECHE AL GRANEL FRIA LITROS	354.271	151.405,16
TRISENSOR ANTIBIOTICOS UNIDAD	288	1.170
FUNDA QUESO 700 GRAMOS UNIDAD	7.400	444,73
CEREAL PARA YOGURT VASO KG	280	957,01
ETIQUETA 200CC FRESA UNIDAD	10.000	290

Ilustración 72. Tabla reporte de compra de toda la materia prima por año y mes. Fuente: Christopher Torres y Adrián Carrión

(Bernabeu, 2010, p.97)

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

El objetivo de este capítulo es presentar los resultados del proyecto desde un enfoque general. A continuación, se detallan las conclusiones y recomendaciones:

5.1 Conclusiones

1. La Inteligencia de Negocios (BI) reside en la capacidad de tomar decisiones para generar valor, aprovechando las oportunidades, siendo eficientes con los recursos y desarrollando capacidades internas de la empresa mediante la generación de conocimiento.
2. Los sistemas basados en data warehouse como soporte para la toma de decisiones facilitan la obtención de información según las necesidades de la empresa, sin embargo, son los altos directivos de la empresa los que realmente analizan la información obtenida y llegan a una conclusión.
3. Pentaho alcanzando un puntaje en el cuadro ponderativo con un promedio de 3.5/4, es la herramienta para Inteligencia de Negocios (BI) que cumple la mayoría de capacidades, destacando la compatibilidad con una gran variedad de fuentes de datos y la administración de los datos, que hace referencia a la gestión de metadatos, procesos ETL y preparación de datos mediante la generación de cubos multidimensionales, lo que la convierte en una alternativa factible y conveniente para llevar a cabo soluciones de Inteligencia de Negocios.
4. La aplicación de soluciones de Inteligencia de Negocios (BI) para empresas PYMES es una oportunidad de negocio, a través de la oferta de soluciones parciales y de bajo costo, en el sentido de que solo involucra la implementación y soporte mas no las licencias del producto.
5. La utilización de la metodología HEFESTO para el diseño lógico del data warehouse representa un proceso eficiente en cuanto a tiempo y recursos, principalmente porque está enfocada en conocer el negocio para el desarrollo de una solución acorde al nivel de madurez de la empresa, basándose en las

necesidades del usuario e involucrándolo en cada etapa del proceso, gracias a su estructura adaptable.

6. La Suite de Pentaho Community Edition (CE) provee herramientas indispensables para la implementación de un data warehouse como solución de Inteligencia de Negocios (BI). Es importante recalcar que el tiempo de carga inicial de datos en la herramienta Data Integration son altos y dependen de la cantidad de datos y el diseño del proceso ETL.
7. La calidad de los datos para la carga del data warehouse es un elemento clave en el éxito de la implementación de una solución de Inteligencia de Negocios (BI), porque influye directamente en el resultado de las consultas y generación de reportes.
8. La publicación de reportes personalizados en la plataforma BI Server de Pentaho permite un manejo intuitivo y sencillo para el usuario, el cual puede generar reportes a partir del ingreso de parámetros con el fin de analizar la información de acuerdo a las necesidades del negocio.

5.2 Recomendaciones

1. Instruirse bien en el uso de las herramientas para la implementación de una solución de Inteligencia de Negocios (BI) antes de llevar a producción, porque se puede conocer muy bien el proceso, sin embargo, es importante aprovechar al máximo el potencial y características que ofrecen las herramientas.
2. Durante el desarrollo de la solución de Inteligencia de Negocios (BI) se invirtió el tiempo mayormente en el análisis de los datos; para reducir considerablemente el tiempo se recomienda que participe activamente una persona de la empresa a la que se aplica el caso de estudio, que conozca sobre el giro del negocio y tenga experiencia en los procesos que se realiza dentro de la empresa.
3. Para la implementación de la solución de Inteligencia de Negocios (BI) en la empresa Lácteos de Marco's, se recomienda que se adquiriera equipos que como mínimo tenga 8 GB de RAM y un buen procesador de cuarta generación, para que ofrezca un rendimiento óptimo de la plataforma BI Server de Pentaho y las

herramientas involucradas en la solución de Inteligencia de Negocios (BI). Se recomienda además que la empresa migre su base de datos a una versión más actual para posteriores soluciones que puedan beneficiarla.

4. Para trabajos futuros de tesis, se recomienda a los estudiantes profundizar en el estudio y aplicación de técnicas de data mining al data warehouse para potenciar la aplicación de una solución de Inteligencia de Negocios (BI) a una empresa.
5. Con el objetivo de incentivar e involucrar al estudiante dentro del campo de Inteligencia de Negocios (BI), se recomienda a la Pontificia Universidad Católica del Ecuador agregar a la malla curricular una materia dedicada al estudio de este campo, que incluya las nuevas tecnologías sobre explotación de los datos.

BIBLIOGRAFÍA

- Whitten, J., Bentley, L., & Dittman, K. (2004). *Systems Analysis and Design Methods* (6th ed.). Pensilvania: Pennsylvania State University. Recuperado de <http://www.books.google.com.ec>
- Schuster Rosa, M. *La era digital, el impacto en los negocios y en los recursos humanos* (Maestría en Dirección de Empresas). Universidad del CEMA. Recuperado de <http://www.books.google.com.ec>
- Curto, D. J., & Curto, D. J. (2010). *Introducción al Business Intelligence*. Barcelona, ES: Editorial UOC. Recuperado de <http://www.ebrary.com>
- Nader, J. (2004). *Sistema de Apoyo Gerencial Universitario* (Magister en Ingeniería de Software). Instituto Tecnológico de Buenos Aires. Recuperado de <http://www.books.google.com.ec>
- Date, C. (2001). *Introducción a los sistemas de bases de datos* (7th ed.). México: José Luis Vázquez. Recuperado de <http://www.books.google.com.ec>
- Luna, M. E., & García, M. F. (2006). *El repositorio de metadatos en un data warehouse*. Santiago de Chile, CL: Red Revista de Facultad de Ingeniería. Recuperado de <http://www.ebrary.com>
- Trujillo, J. C. (2013). *Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional*. Alicante, ES: ECU. Recuperado de <http://www.ebrary.com>
- Mendez del Rio, L. (2006). *Más allá del business intelligence: 16 experiencias de éxito* (6th ed.). Gestion 2000. Recuperado de <http://www.books.google.com.ec>
- Edison Medina La Plata. (2012). *Business Intelligence: Una guía práctica*. (Universidad Peruana de Ciencias Aplicadas - UPC, Ed.) (1ra. ed., pp. 9–109). Lima. Recuperado de <http://hdl.handle.net/10757/338680>
- Amat Salas, O. & Campa, F. (2013). *Manual Del Controller*. Bresca. Recuperado de <http://www.books.google.com.ec>
- Bernabeu, Darío. (2010). *Data Warehousing y metodología Hefesto*. DataPrix. Recuperado de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto>

- Gartner. (2016). *Magic Quadrant for Business Intelligence and Analytics Platforms*. Recuperado <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>
- GNU. (2016). *¿Qué es el software libre?*. Recuperado el 18 de mayo de 2016 de <http://www.gnu.org/philosophy/free-sw.es.html>
- InnovateSolutions. (2016). *BIRT Review*. Recuperado de <http://www.innoventionsolutions.com/birt-review.html>
- InnovateSolutions. (2016). *Jaspersoft Studio Review*. Recuperado de <http://www.innoventionsolutions.com/jasper-review.html>
- InnovateSolutions. (2016). *Pentaho Reports Review*. Recuperado de <http://www.innoventionsolutions.com/pentaho-review.html>
- Pentaho. (2016). *About us*. Recuperado el 18 de mayo de 2016 de <http://www.pentaho.com/about>

ANEXOS

Anexo A

MANUAL TÉCNICO

En este manual se describe los requerimientos del sistema y la instalación de cada herramienta utilizada de la Suite de Pentaho Community Edition (CE) y el gestor de base de datos MySQL. Además, se describe el proceso de migración de la base de datos de la empresa de SyBase SQL Anywhere v. 9.0.2.2451 a MySQL v. 5.7.14

A.1. MySQL

A.1.1. Requisitos previos

Para instalar y ejecutar MySQL Workbench 5.7 en Windows, se necesita tener algunas bibliotecas instaladas, las cuales se describen a continuación:

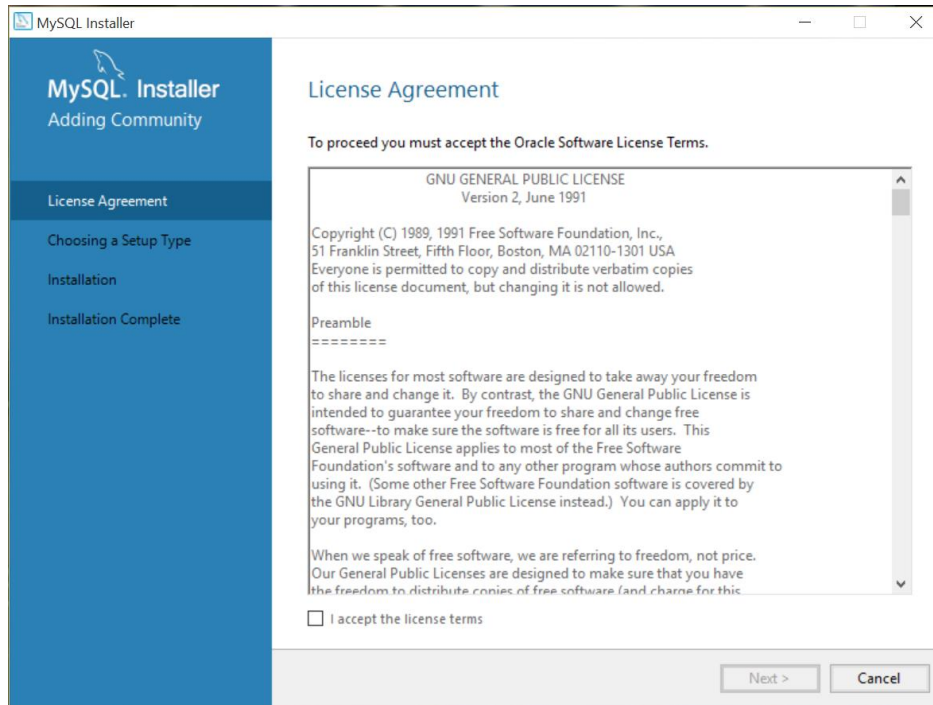
- Microsoft .NET Framework 4 Client Profile, el cual puede descargar del siguiente link:
<https://www.microsoft.com/es-es/download/details.aspx?id=24872>
- Microsoft Visual C ++ 2010 Redistributable Package (x86), el cual puede descargar del siguiente link:
<https://www.microsoft.com/es-ES/download/details.aspx?id=5555>

A.1.2. Instalación y configuración

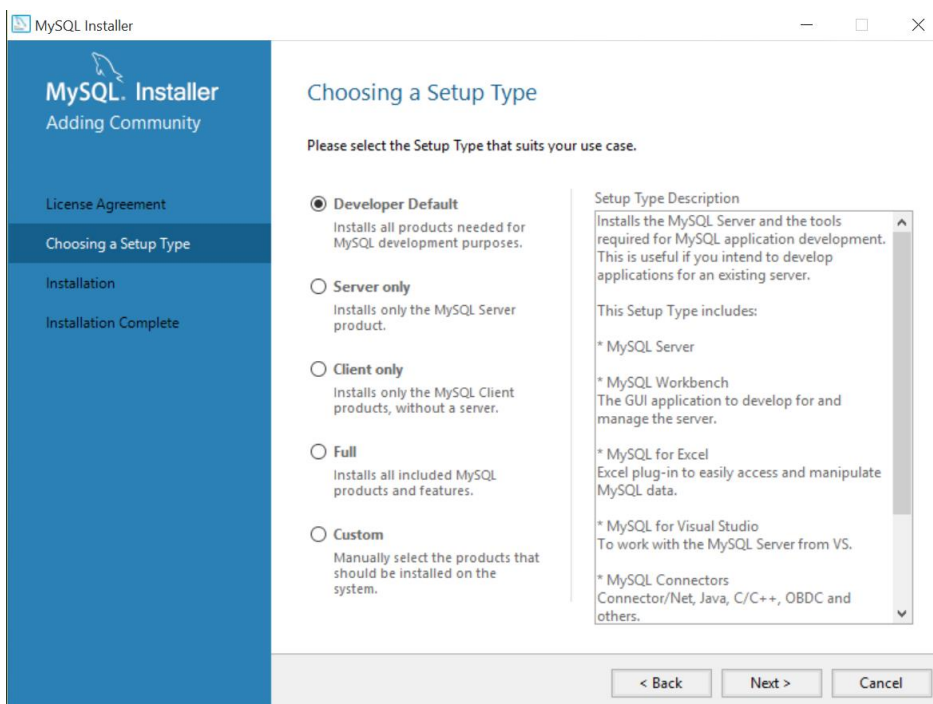
MySQL Workbench es una herramienta visual de diseño de bases de datos. MySQL Workbench está disponible para Windows, Linux y Mac OS X. A continuación, se describe el procedimiento para su instalación:

Paso 1: Descargar el fichero ejecutable de la última versión liberada de la página oficial de MySQL del siguiente link: <https://dev.mysql.com/downloads/workbench/>

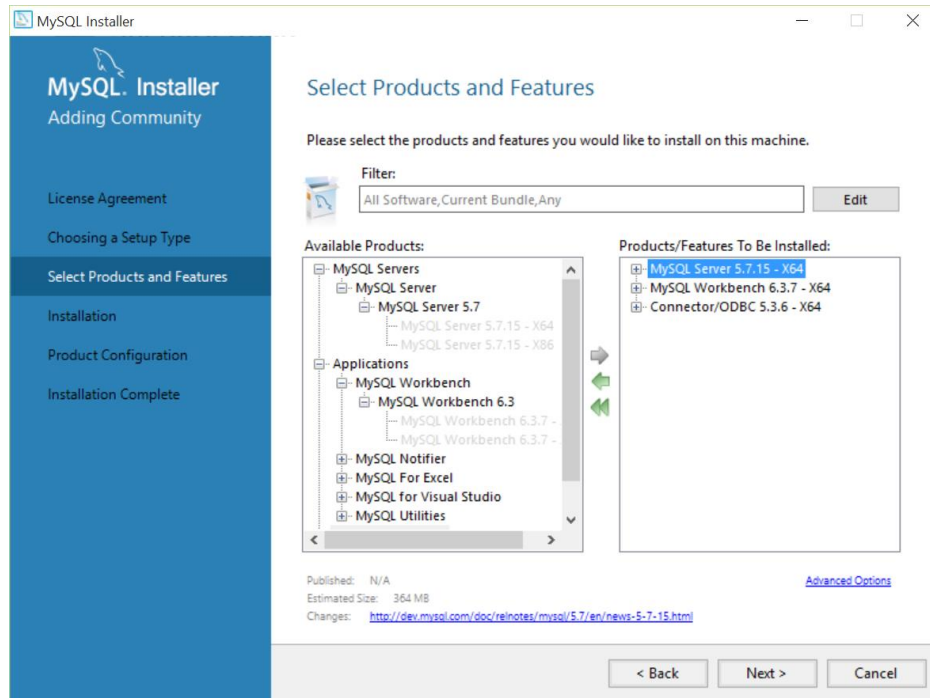
Paso 2: Buscar la carpeta donde se descargó el archivo y dar doble clic para comenzar el proceso de instalación, el cual muestra la pantalla de inicio. Aceptar términos de licencia y dar clic en *Next*.



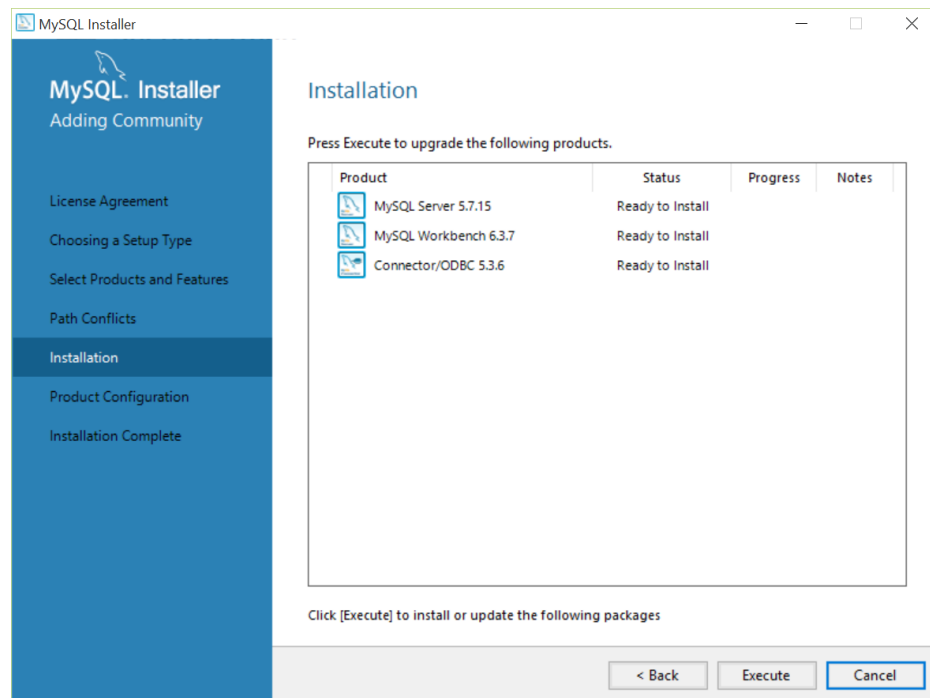
Paso 3: Seleccionar el tipo de instalación *Custom*.



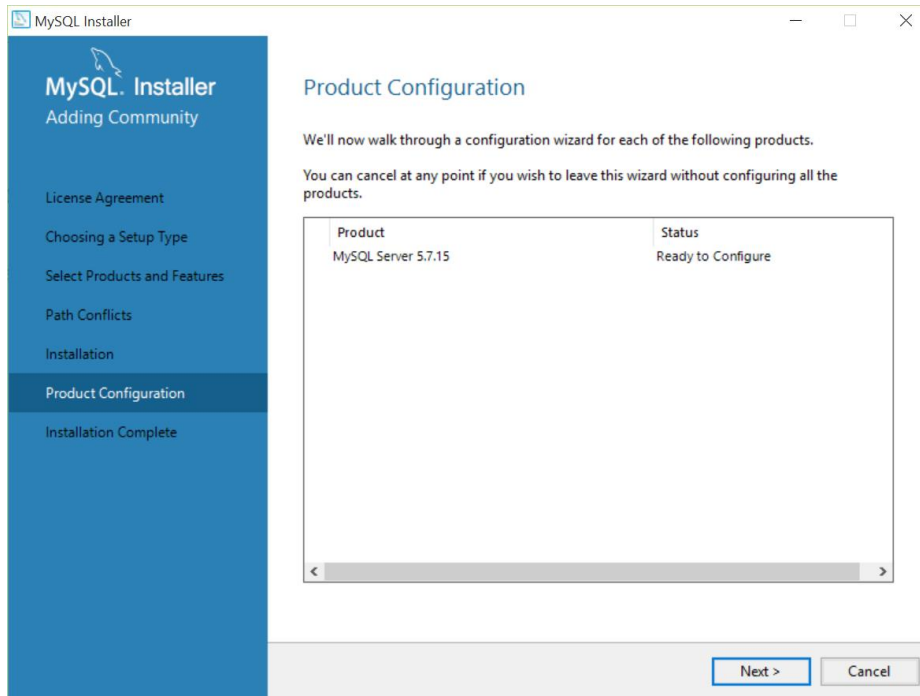
Paso 4: Seleccionar *MySQL Server*, *MySQL Workbench*, y *Connector/ODBC*. Tener en cuenta el tipo de sistema operativo 32 o 64 bits.



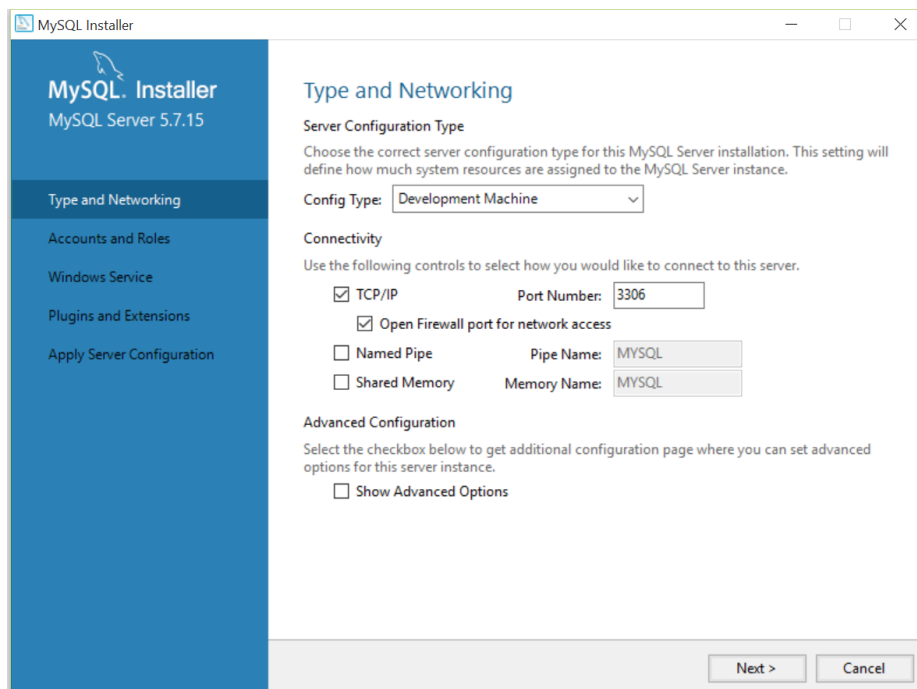
Paso 5: Dar clic en *Execute* para comenzar la instalación de los productos. Una vez que se complete la instalación dar clic en *Next*.



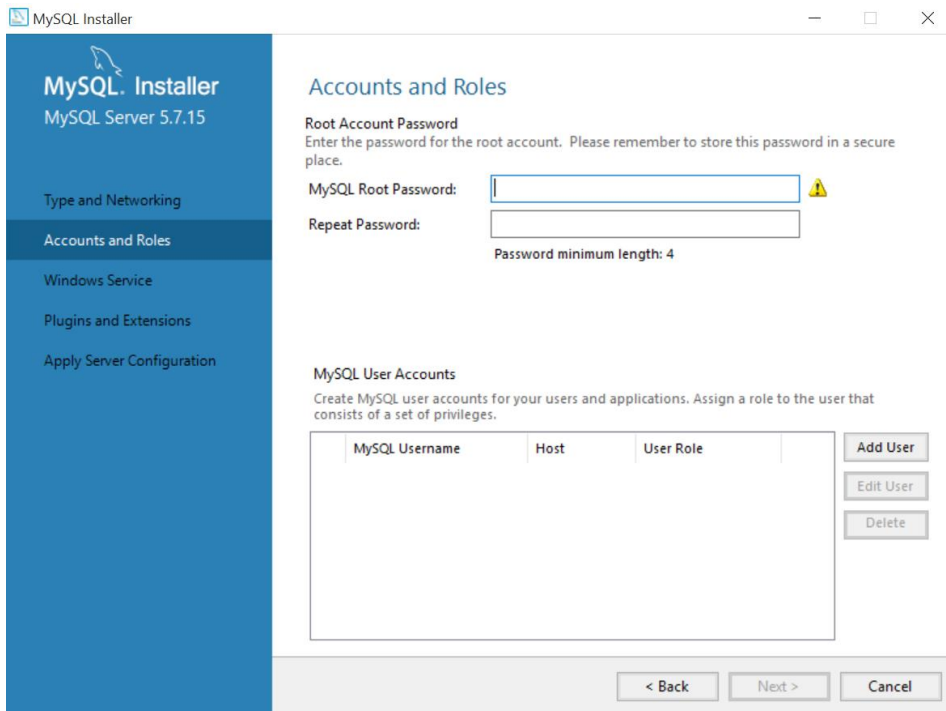
Paso 6: Dar clic en *Next* para empezar la configuración del servidor MySQL.



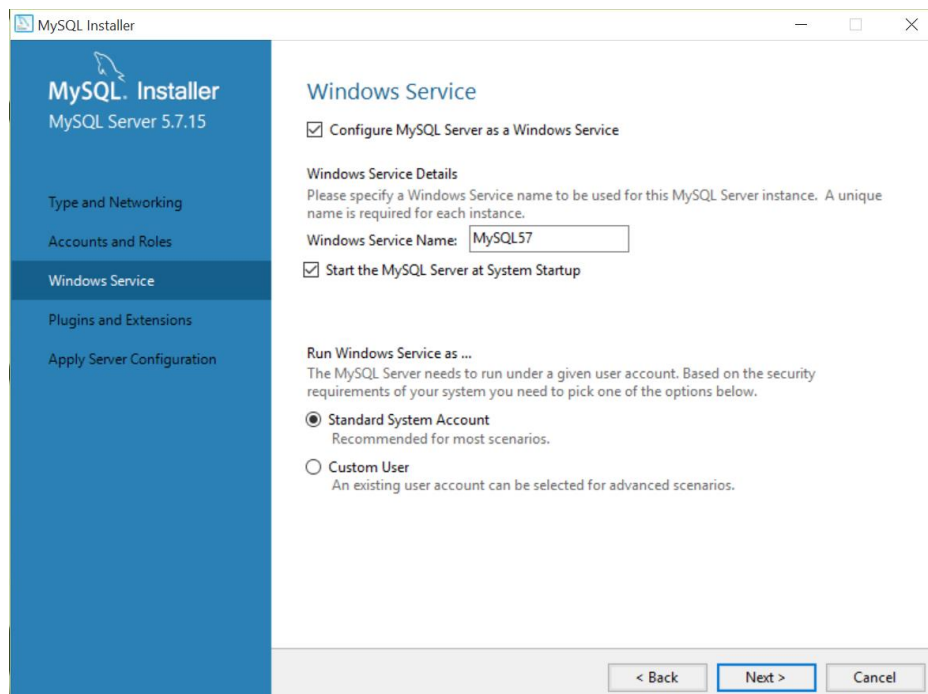
Paso 7: Se procede a configurar el servidor dependiendo de las necesidades del usuario. Para el caso del proyecto se deja por defecto todo y dar clic en *Next*.



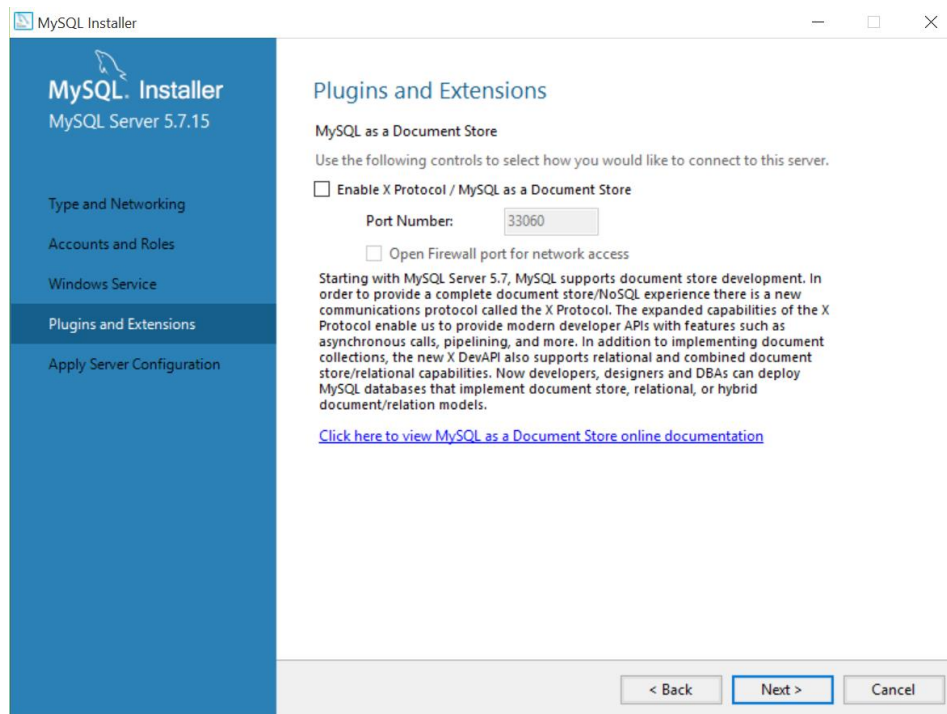
Paso 8: Se debe escribir una contraseña para el servidor y dar clic en *Next*.



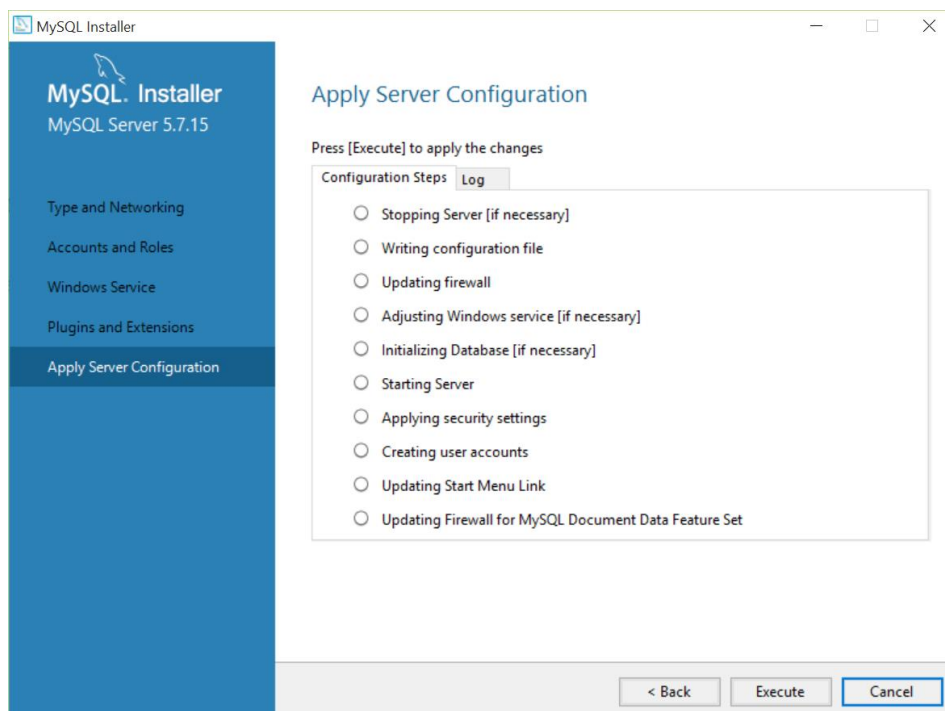
Paso 9: Decidir si MySQL Server se configura como un servicio de Windows, además de iniciar MySQL cuando el Sistema Operativo se inicie. Para el caso del proyecto se deja la configuración que se observa en la imagen.



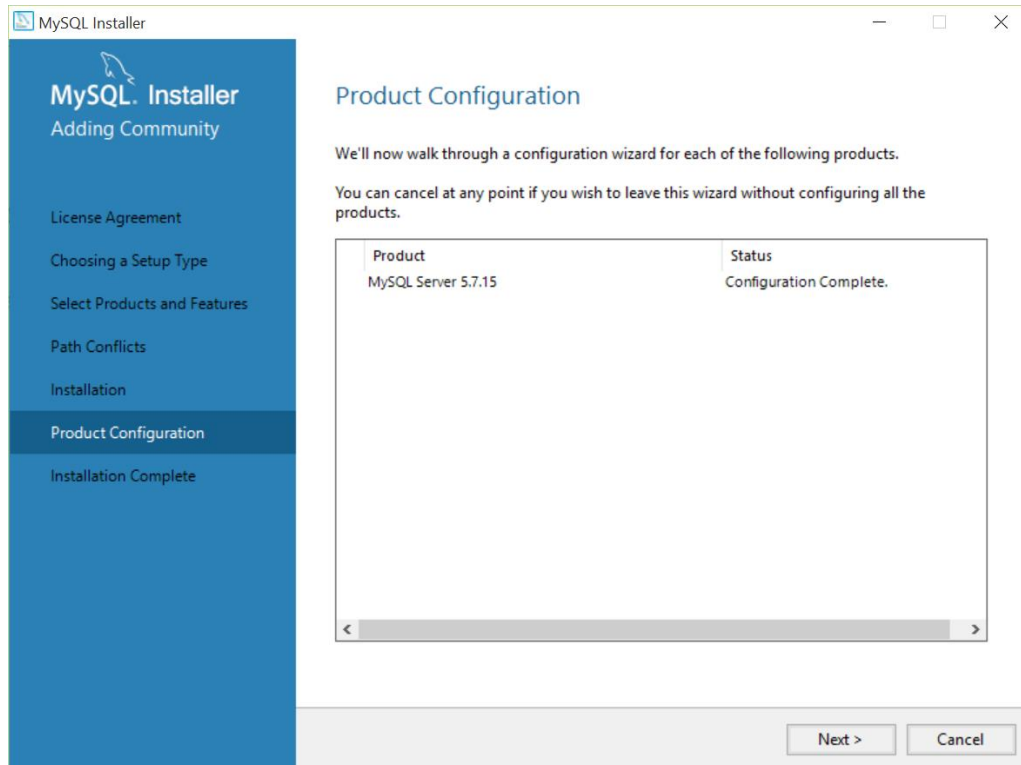
Paso 10: Decidir si MySQL como un almacén de documentos. Para el caso del proyecto dejar la configuración por defecto que se muestra en la imagen.



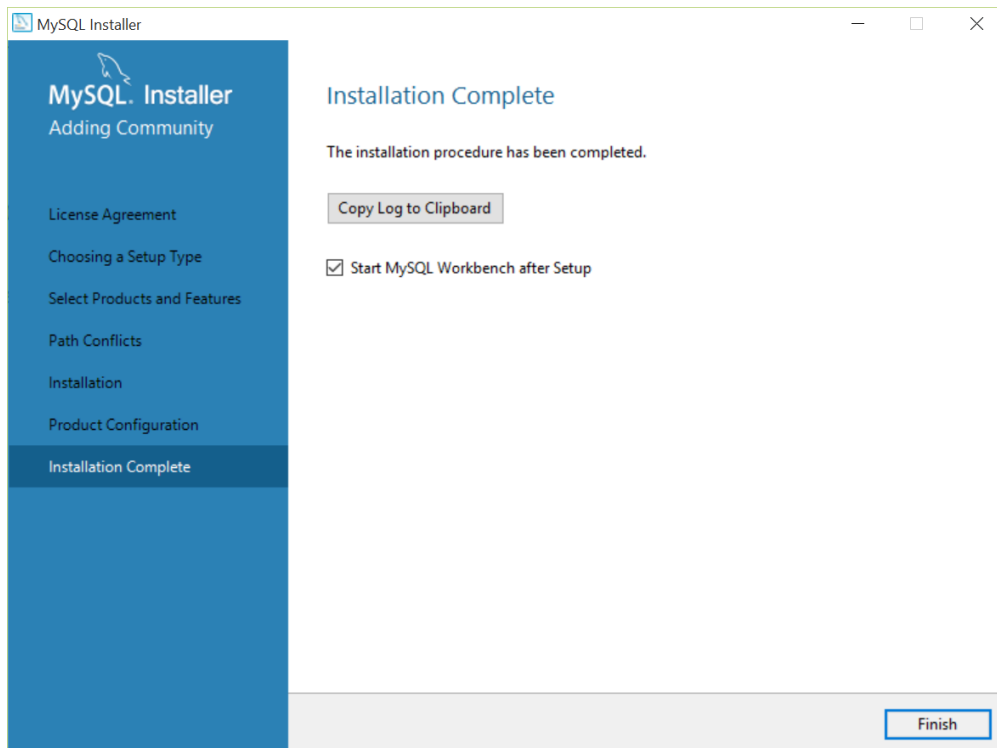
Paso 11: Dar clic en *Execute* para aplicar la configuración del servidor MySQL. Una vez finalizado la configuración dar clic en *Finish*.



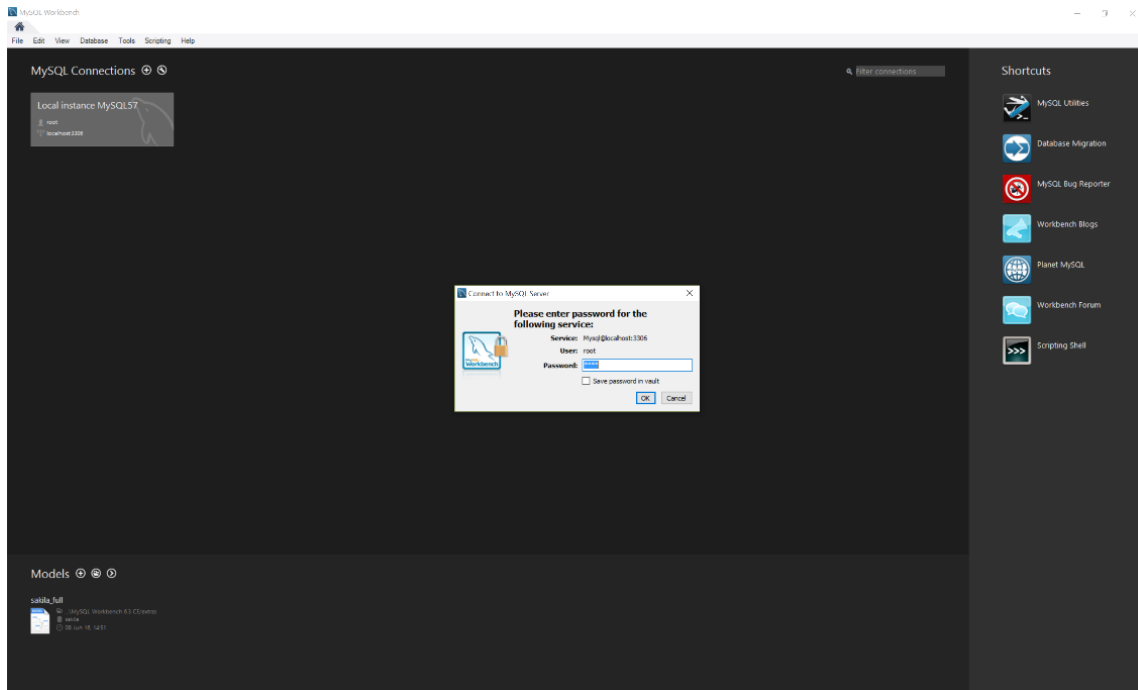
Paso 12: Dar clic en *Next* para completar la configuración.



Paso 13: Dar clic en *Finish* para terminar con la instalación de MySQL.



Finalmente, al ejecutar MySQL Worbench se muestra una interfaz como la que se presenta en la siguiente imagen, donde se debe hacer clic en la instancia que se creó y configuro e ingresar la contraseña para acceder.



A.2.3. Migración de SyBase SQL Anywhere v. 9.0.2.2451 a MySQL v. 5.7.14

Para realizar la migración de la base de datos para el proyecto, se sigue el siguiente esquema:

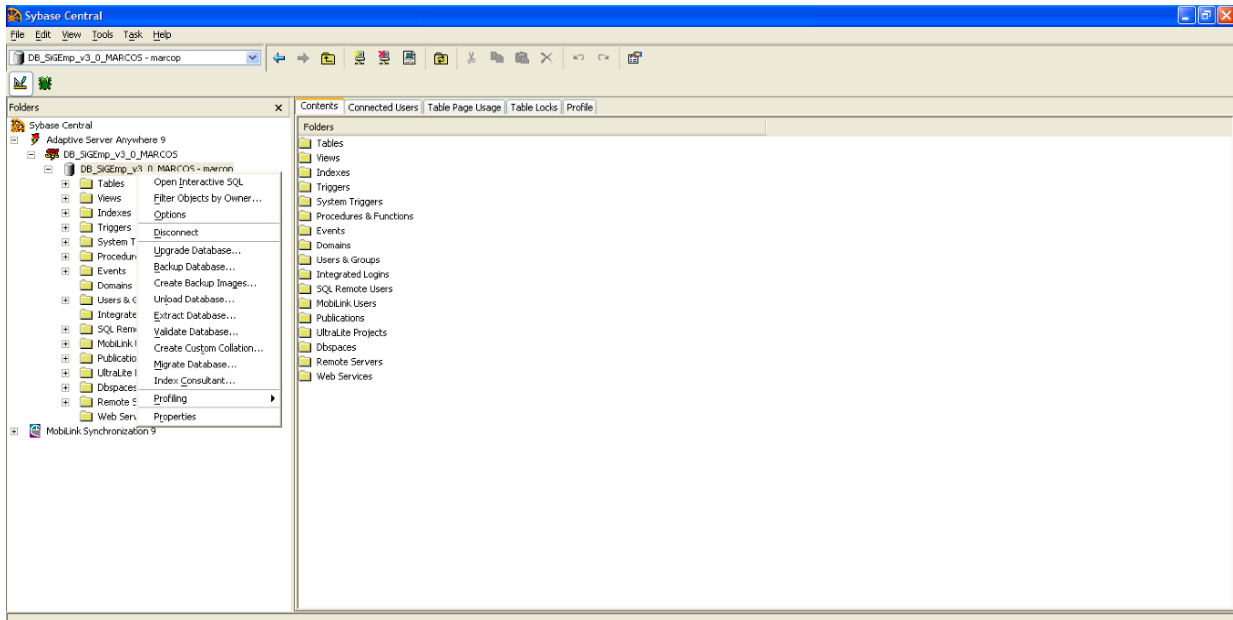
Como se muestra en la imagen se tuvo que migrar de SyBase Sql Anywhere v. 9.0.2.2451 a



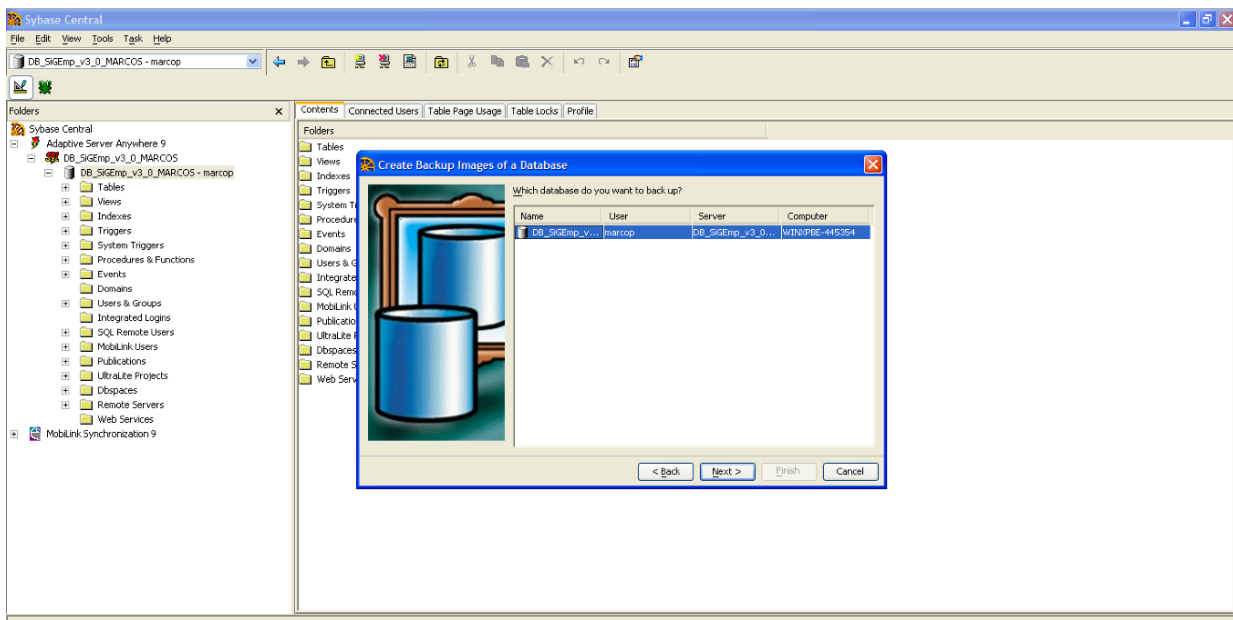
SyBase Sql Anywhere Developer Edition v. 16, debido a que al momento de utilizar la herramienta de migración de MySQL Worbench, este es compatible con versiones de la 12 para adelante de SyBase Sql Anywhere.

Lo primero que se hizo fue sacar un respaldo de imagen de la base de datos de SyBase SQL Anywhere v. 9.0.2.2451 que maneja la empresa, como se describe a continuación:

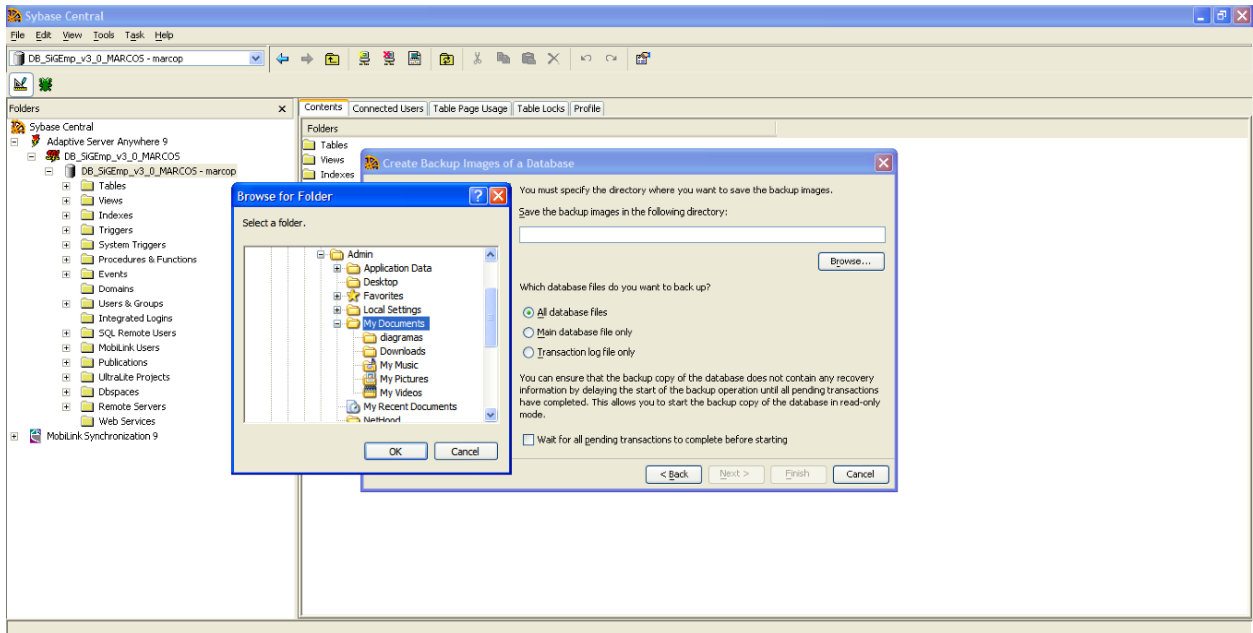
En la siguiente imagen se muestra la herramienta visual para el manejo de SQL Anywhere 9, llamado Sybase Central. Una vez conectado a la base de datos, se procede hacer clic derecho sobre la base y seleccionar la opción *Create Backup Image*.



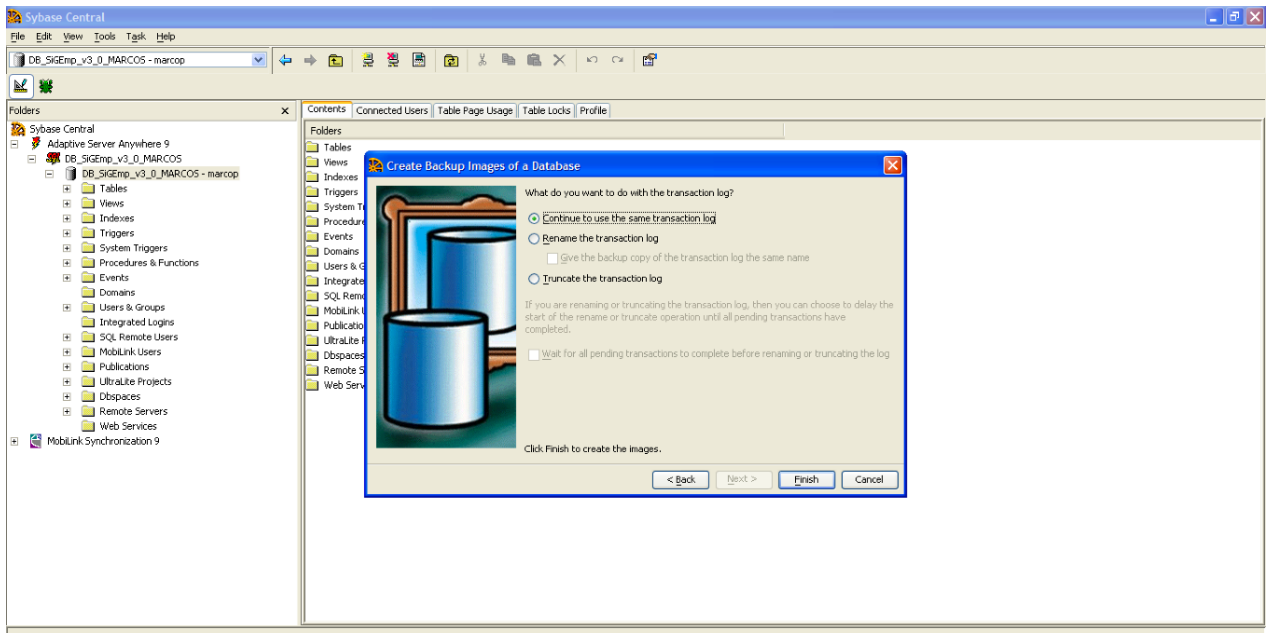
Se selecciona la base de datos que se quiere respaldar y se da clic en *Next*.



Se selecciona el directorio donde se desea guardar el respaldo de la base de datos y se da clic en *Next*.

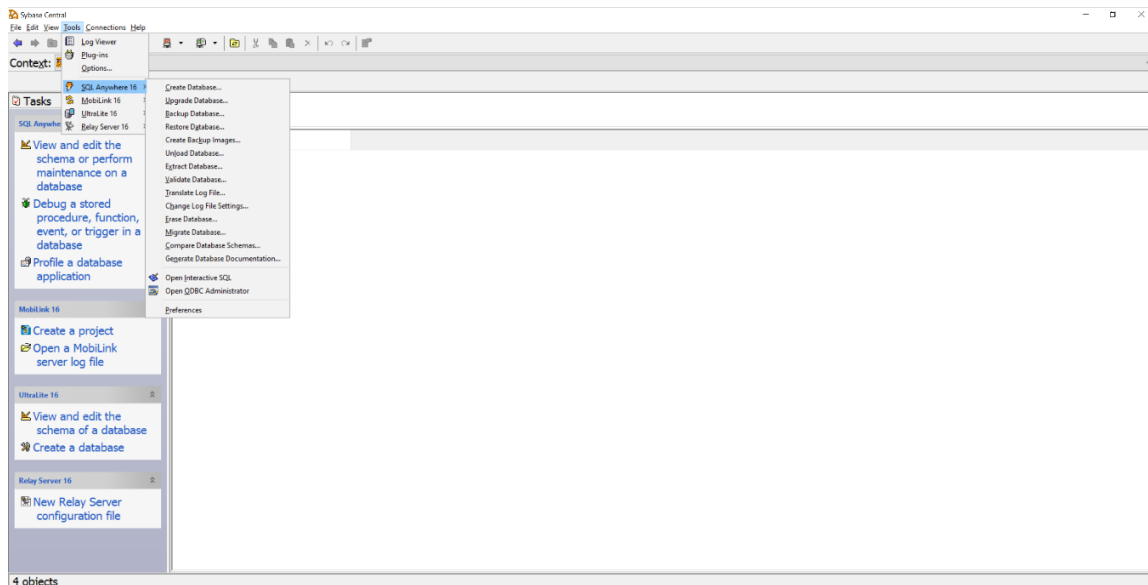


Finalmente se elige la opción *Finish* para terminar de respaldar la base de datos.

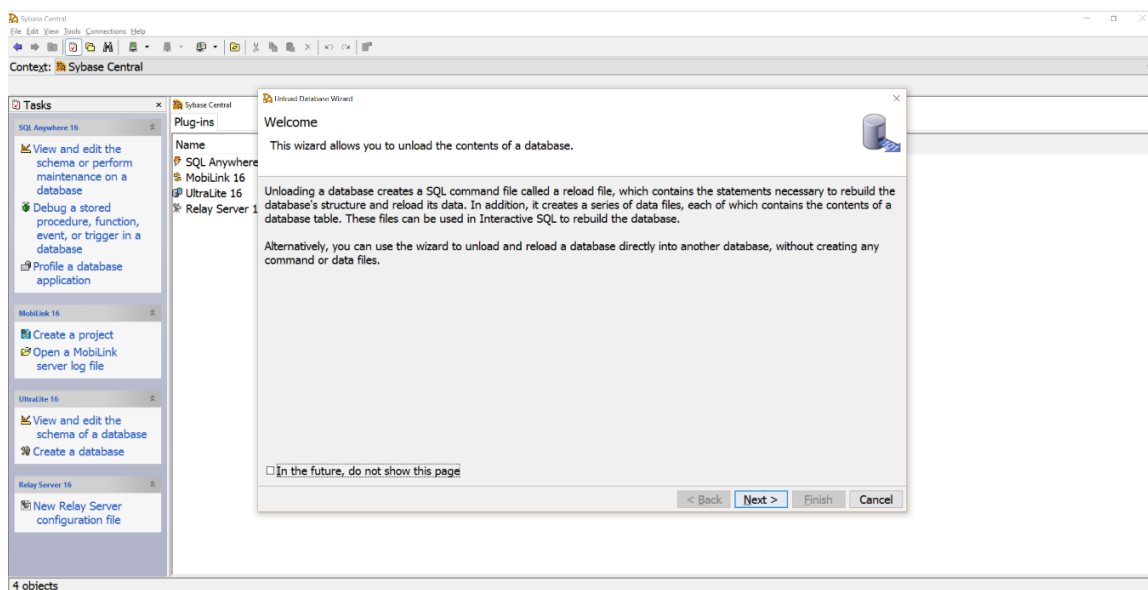


Para la siguiente parte del proceso de migración se necesita tener instalado SyBase SQL Anywhere Developer Edition v. 16, para más información sobre su instalación puede consultar en el siguiente link: <http://scn.sap.com/docs/DOC-50289>.

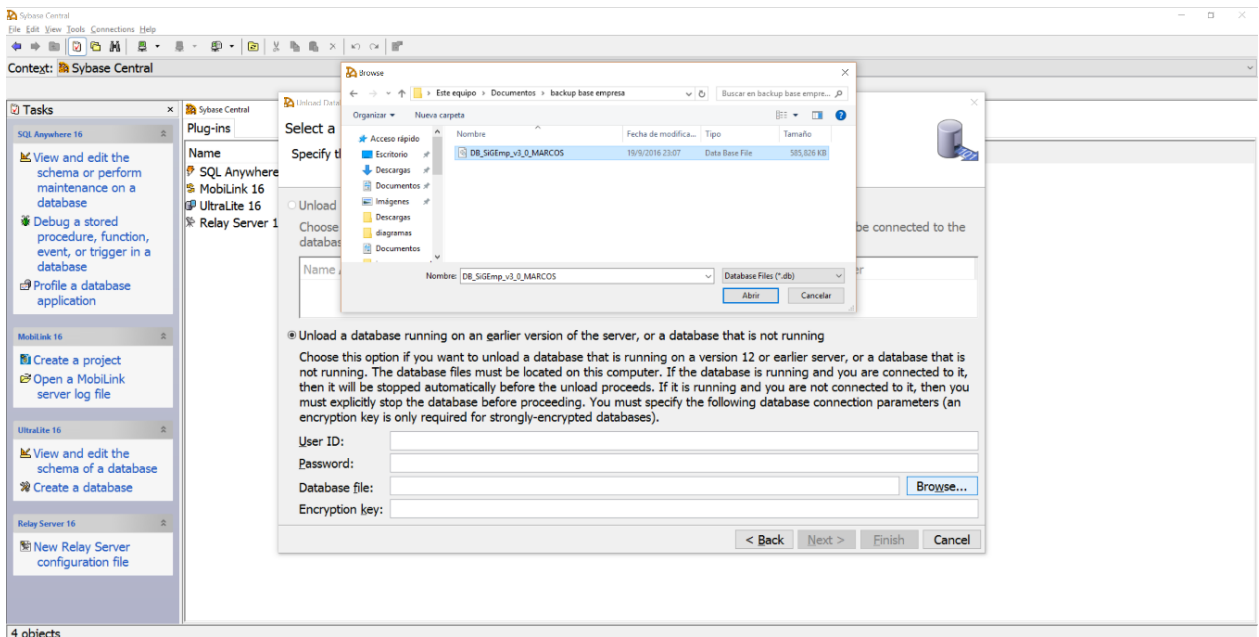
Para migrar la base de datos, se debe abrir el programa SyBase Central e ir al menú Tools y seleccionar la opción *Unload Database*.



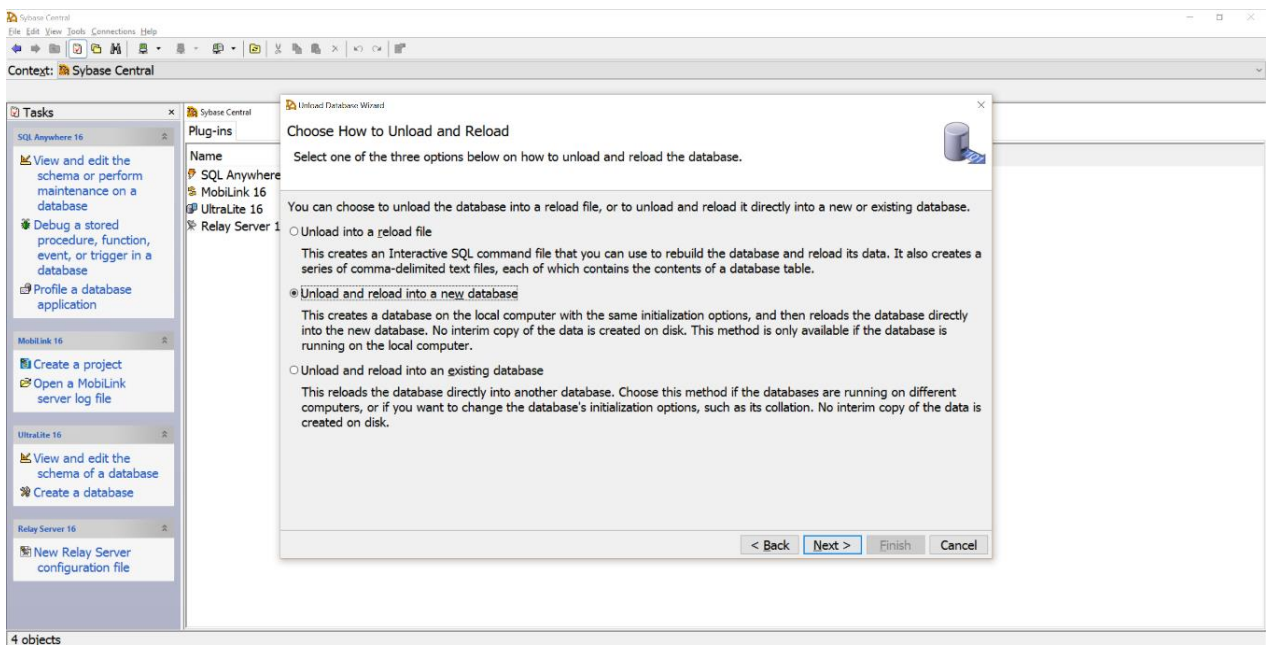
Aparece la pantalla de inicio de la migración y se da clic en *Next*.



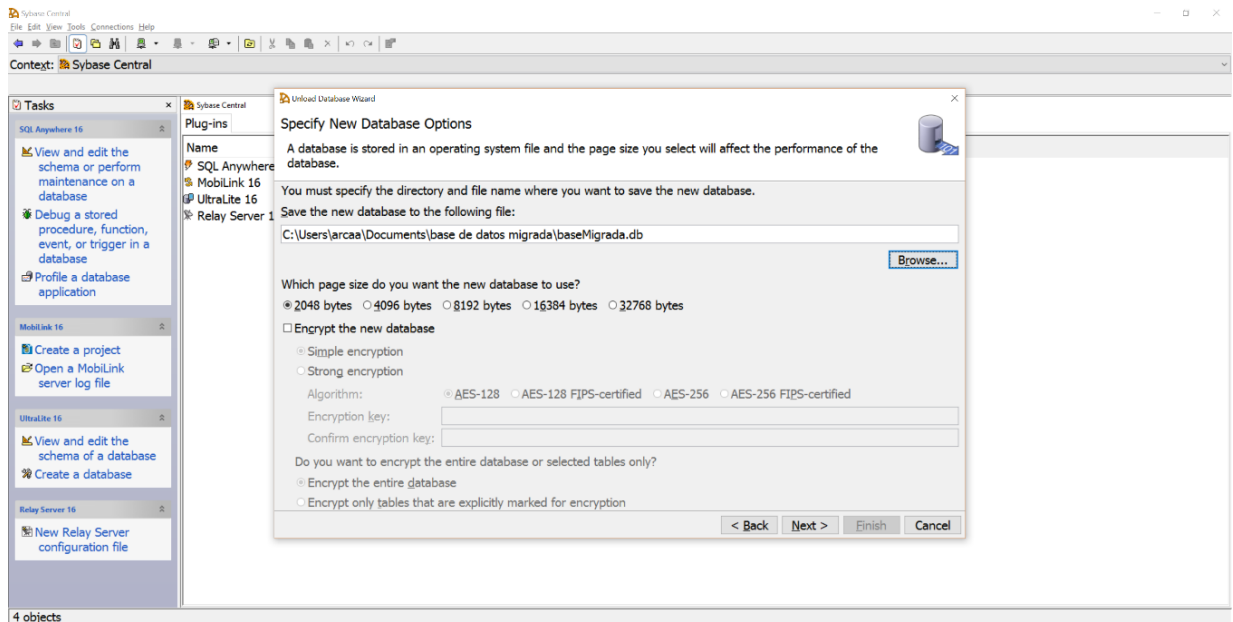
Se debe buscar el respaldo de imagen de la base de datos y llenar los datos de *User ID* y *Password* con los respectivos y dar clic en *Next*.



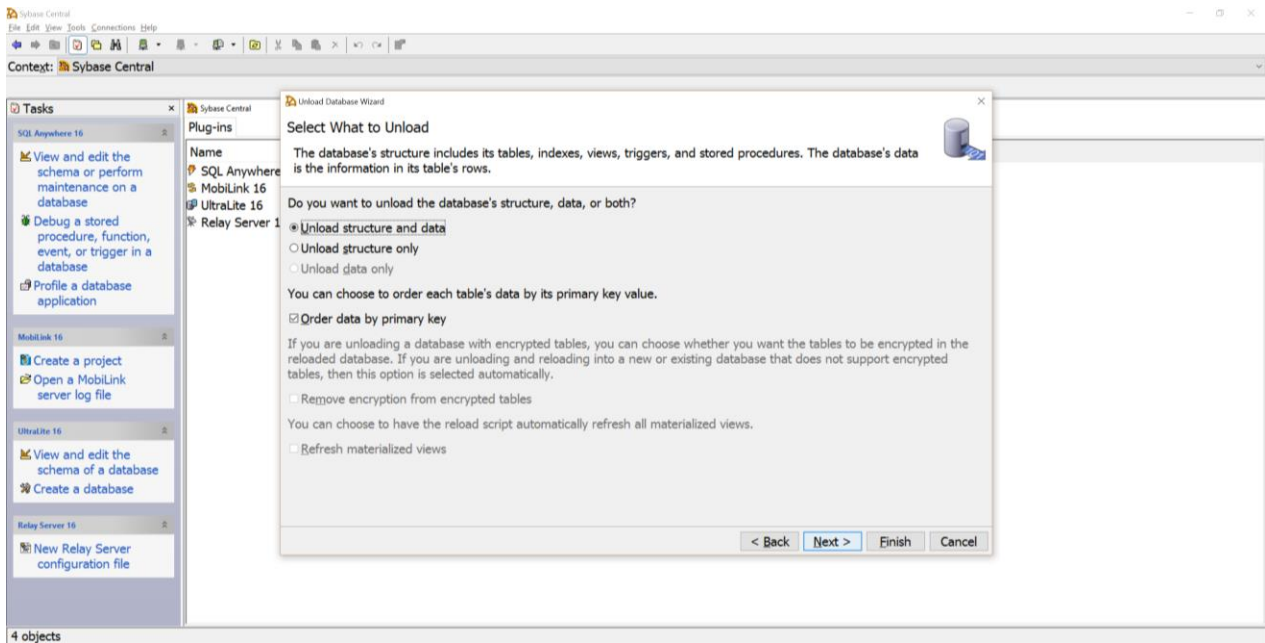
Se selecciona la opción *Unload and reload a new database*, para que se cree una nueva base de datos con el esquema y datos de la base que se migra.



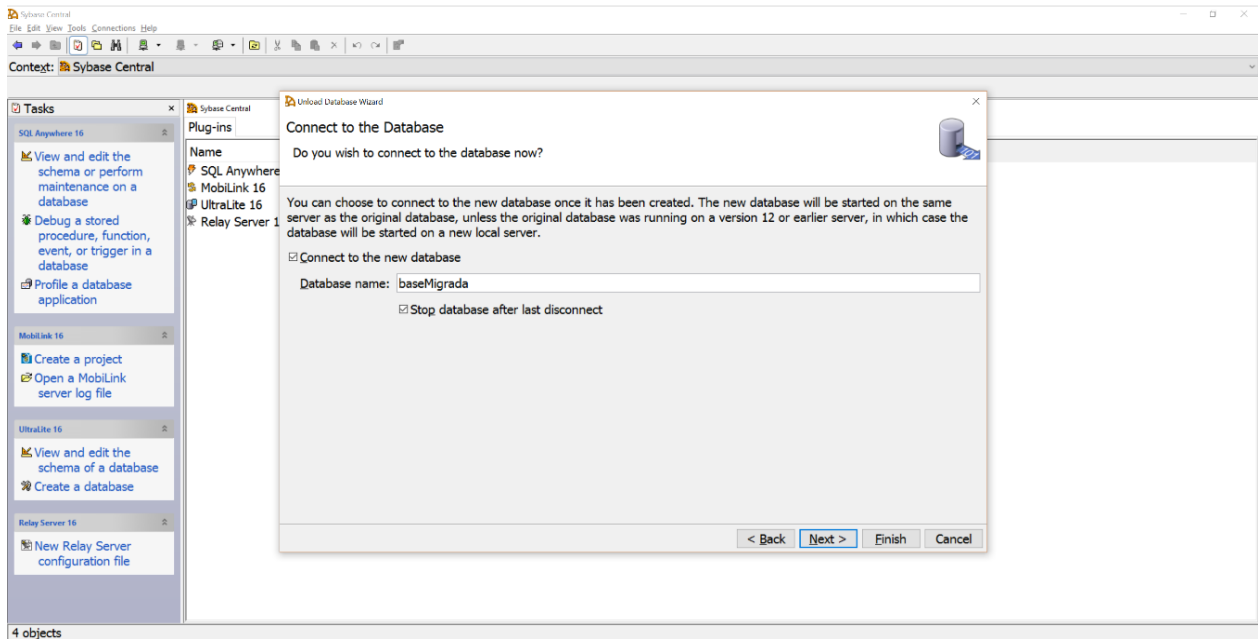
Buscar el directorio donde se guardará la nueva base de datos y configurar según las especificaciones de la base de datos que se migra.



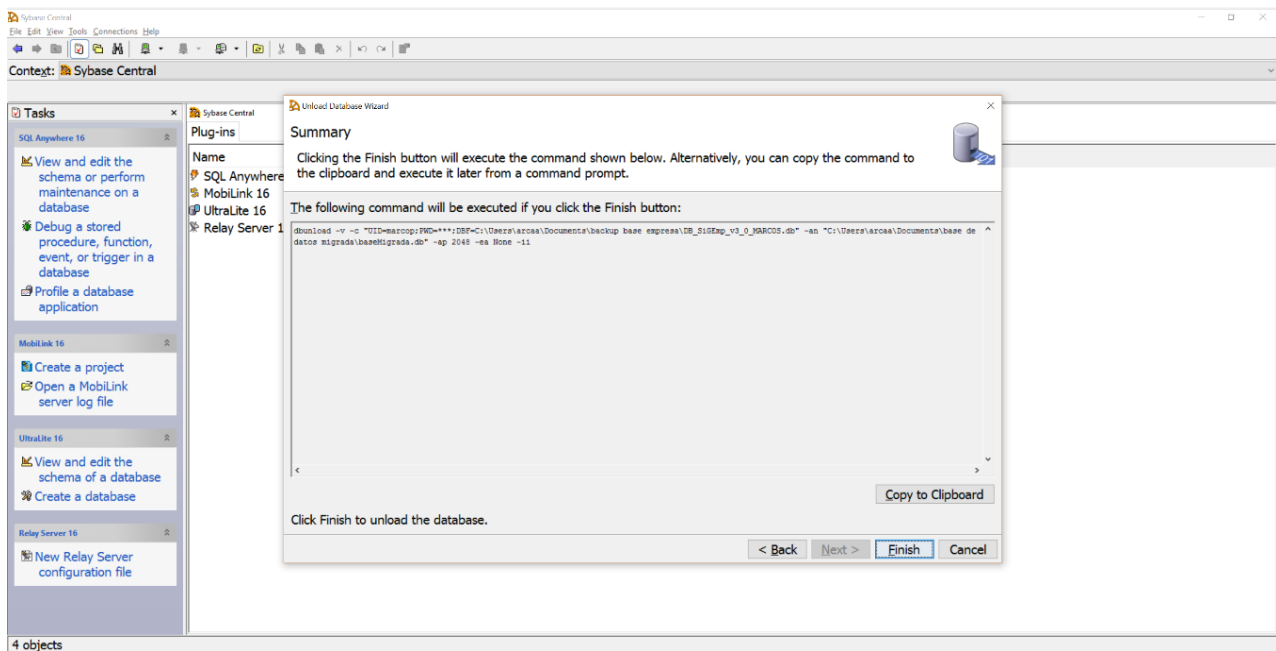
Seleccionar la opción *Unload structure and data* para migrar el esquema y datos.



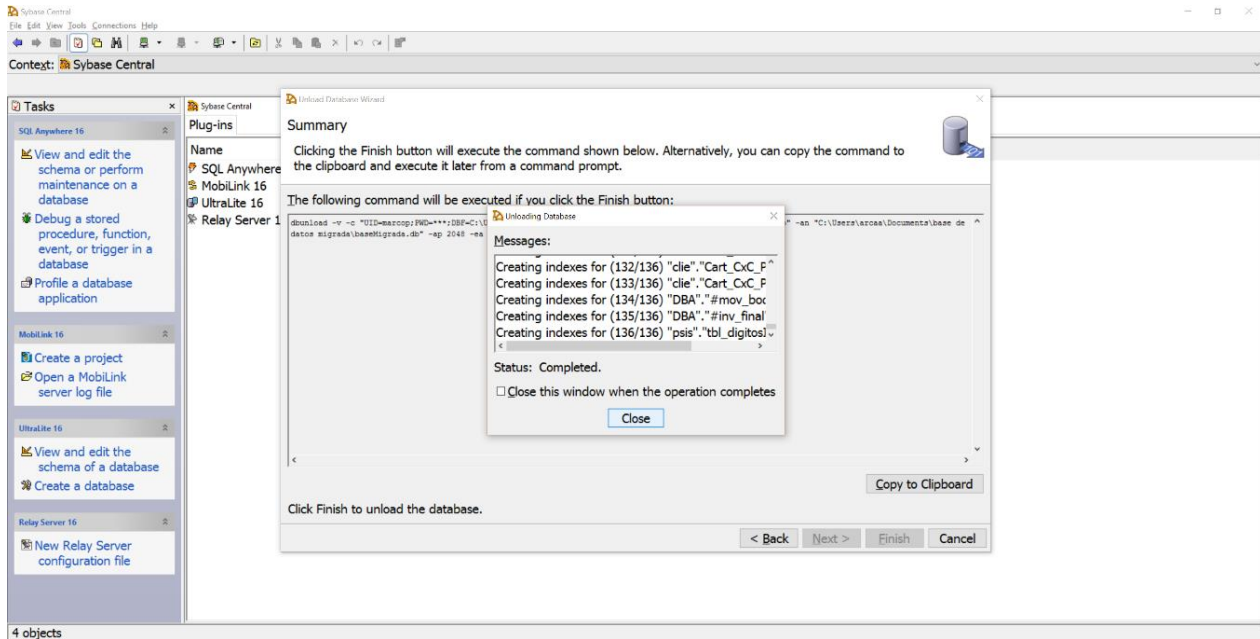
Decidir si quiere conectar a la base de datos en ese momento y dar clic en *Next*.



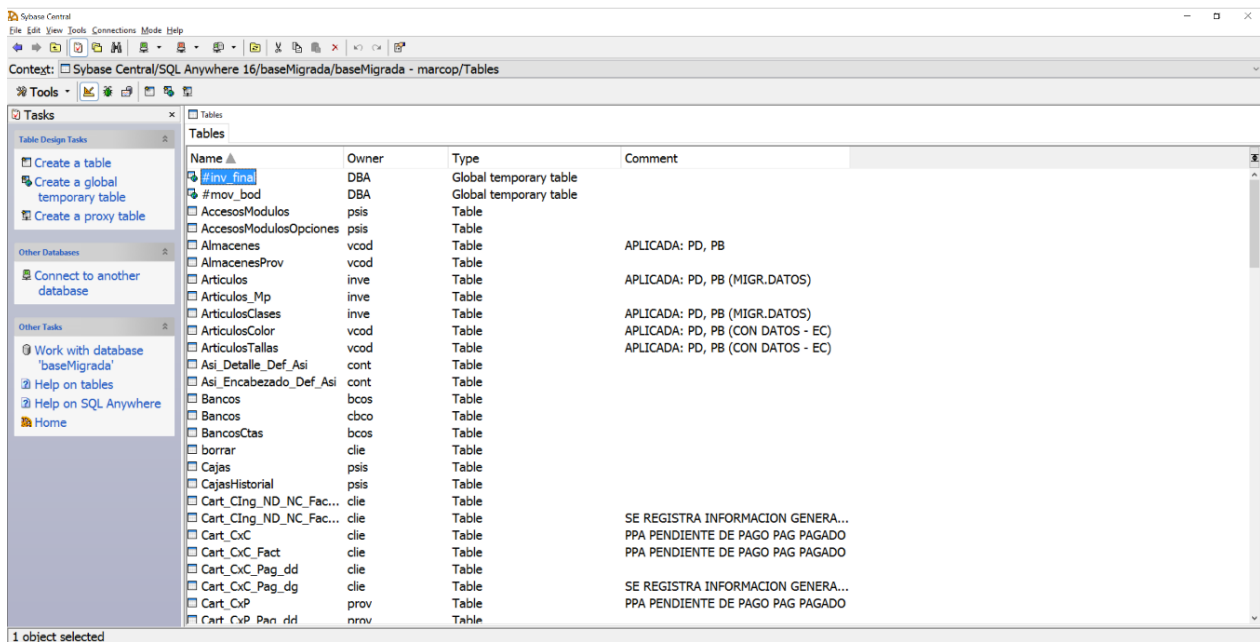
A continuación, se presenta los comandos que se ejecutaran para la migración de la base de datos y dar clic en *Finish*.



Finalmente, empieza a migrar la base de datos y presenta un mensaje de completado cuando termina el proceso.

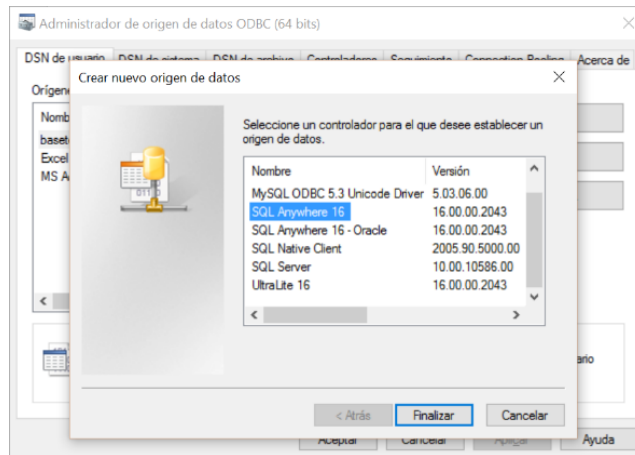


En la siguiente imagen se muestra la base de datos de la empresa migrada a SyBase SQL Anywhere v. 16.

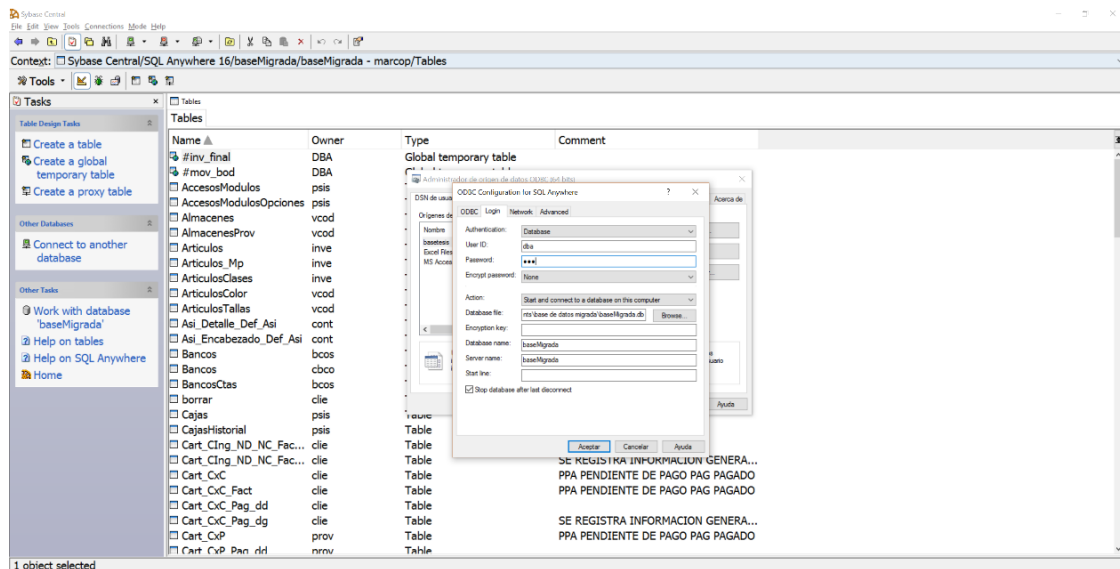


Para continuar con la migración de la base de datos a MySQL, debemos de configurar una conexión ODBC a nuestra base de datos migrada, para ello realizar el siguiente procedimiento:

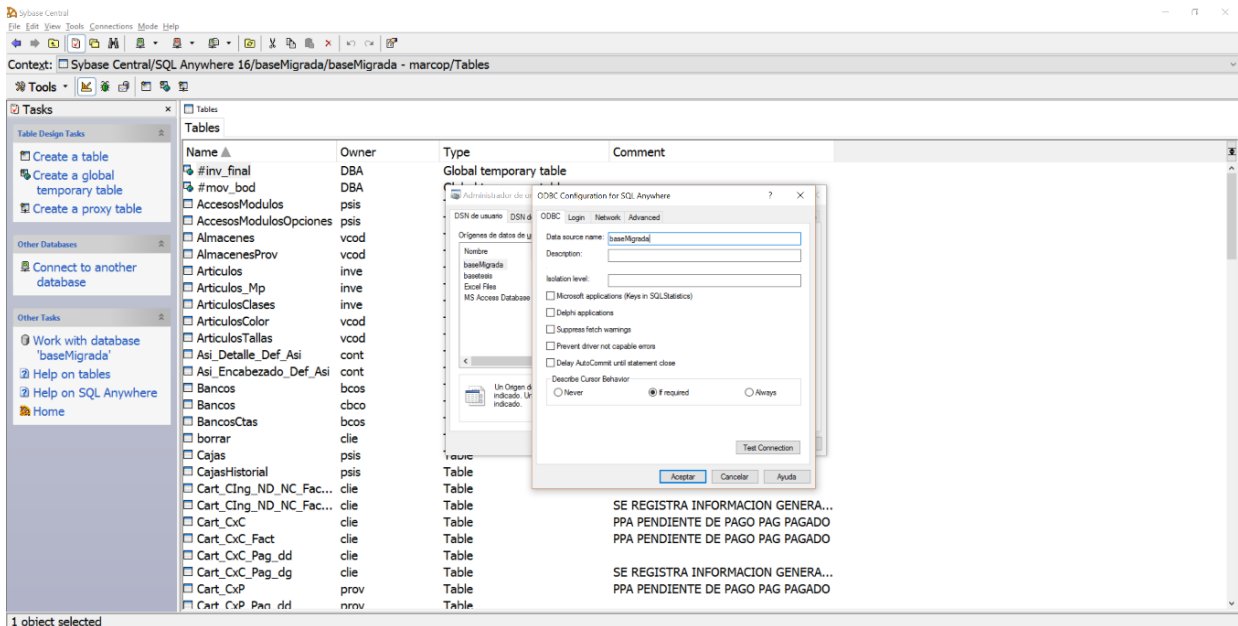
Abrir el programa Orígenes de datos ODBC, seleccionar la opción de Agregar en la pestaña de Orígenes de datos de usuario, luego elegir SQL Anywhere 16 y dar clic en *Finalizar*.



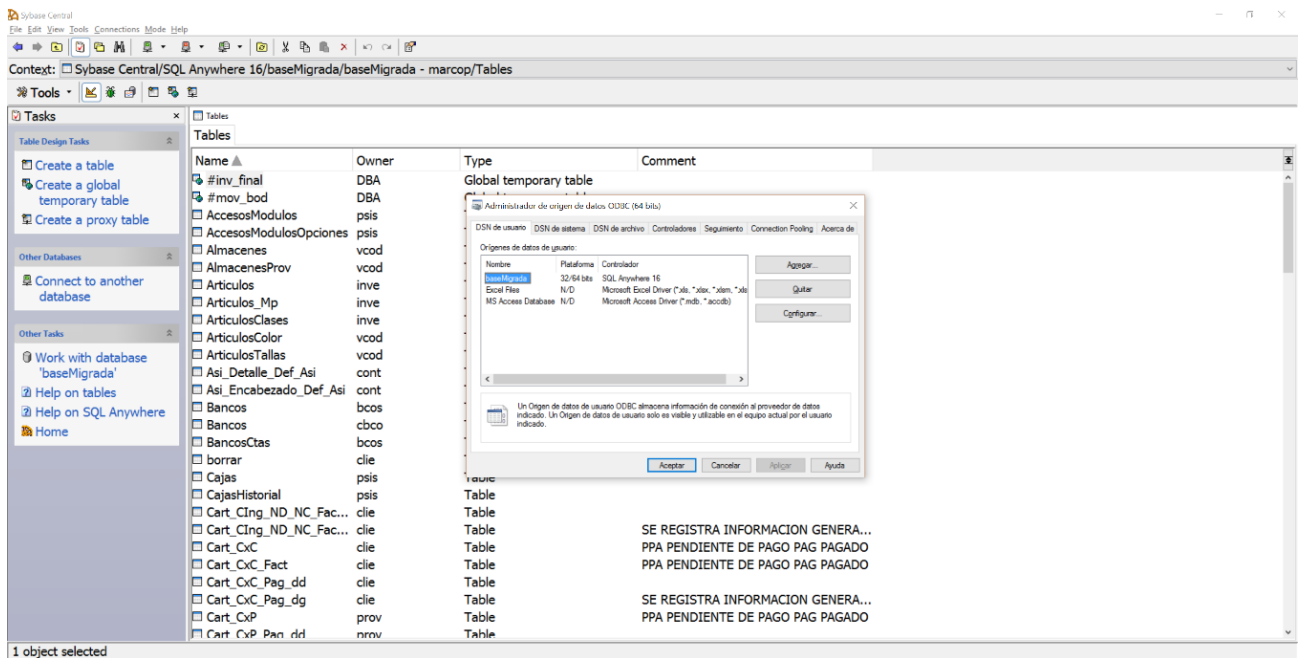
Dirigirse a la pestaña Login y llenamos los datos de *User ID*: dba y *Password*: sql. En la opción de *Action*, seleccionar *Start and connect to a database on this computer*, luego buscar el archivo de la base de datos que se migro y colocar el mismo nombre de la base de datos en la casilla de *Database name* y *Server name* y dar clic en *Aceptar*.



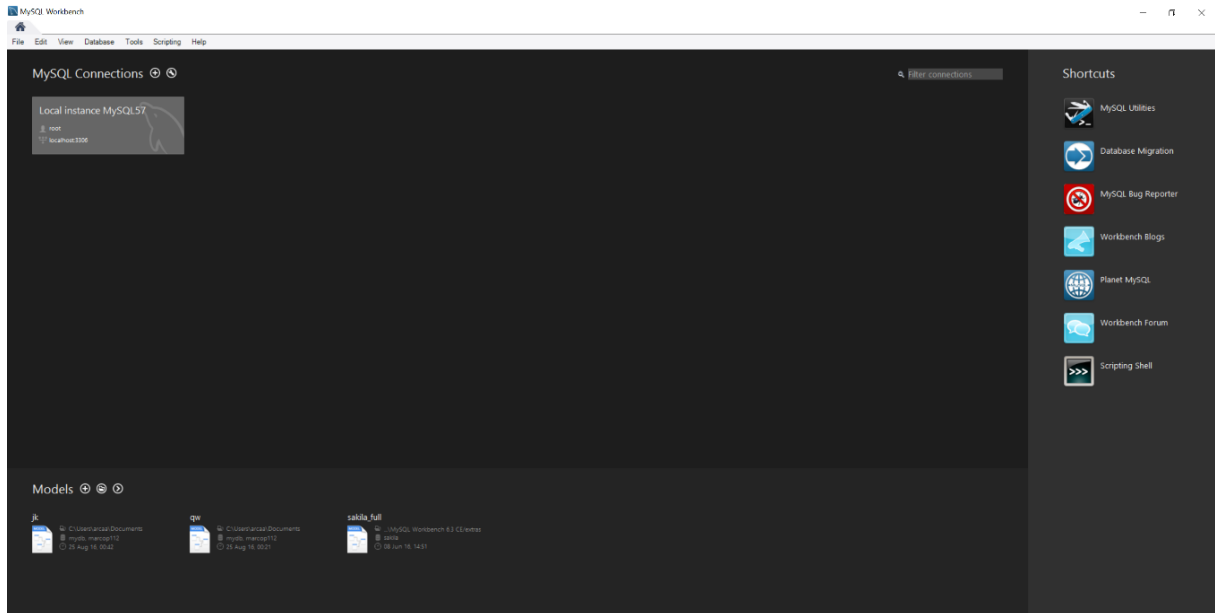
Dirigirse a la pestaña *ODBC* y colocar un nombre a la conexión en la casilla *Data source name* y dar clic en *Aceptar*.



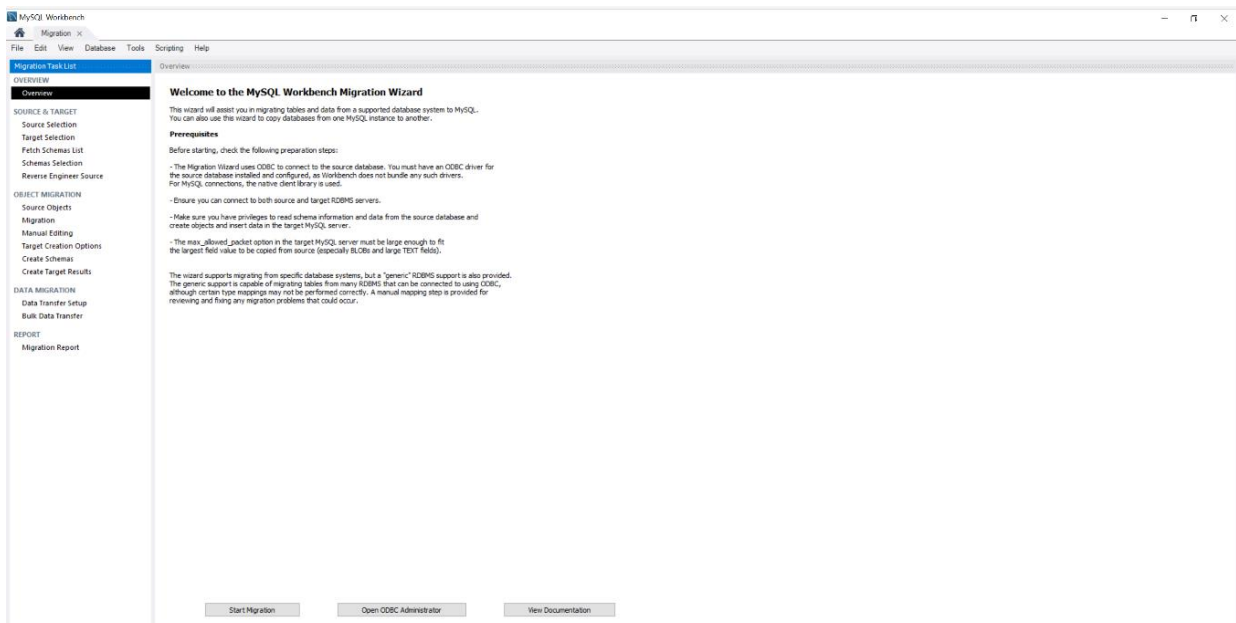
Finalmente, dar clic en *Aceptar* para finalizar la configuración de una nueva conexión ODBC.



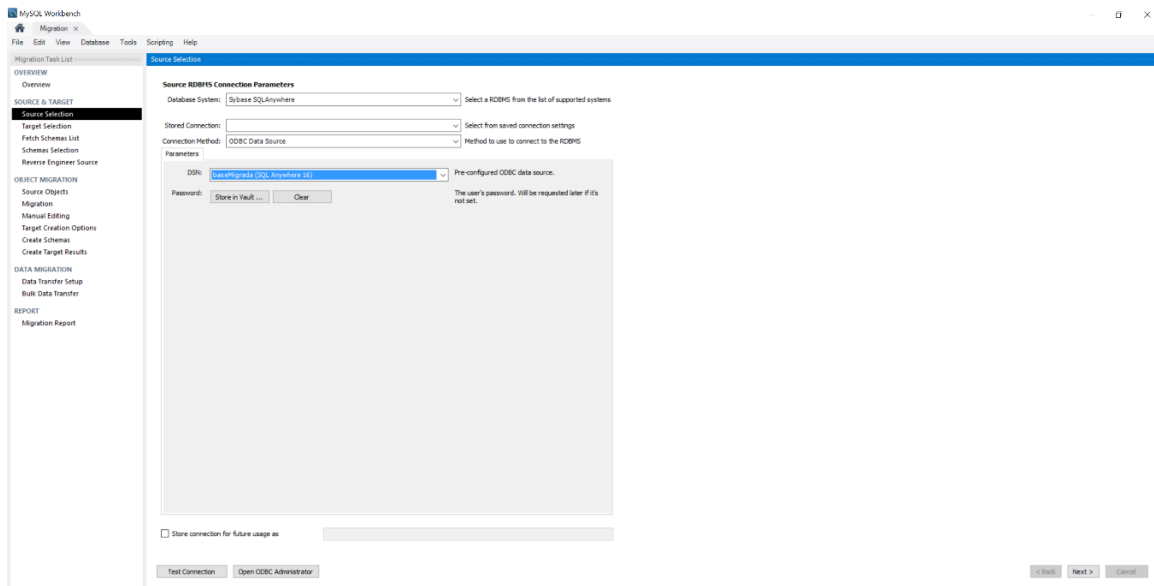
Para la última parte de la migración de la base de datos a MySQL, se utiliza la herramienta de migración que ofrece MySQL Workbench. Para iniciar el proceso, se debe seleccionar la opción Database Migration que se encuentra en la parte derecha de la pantalla de inicio del programa.



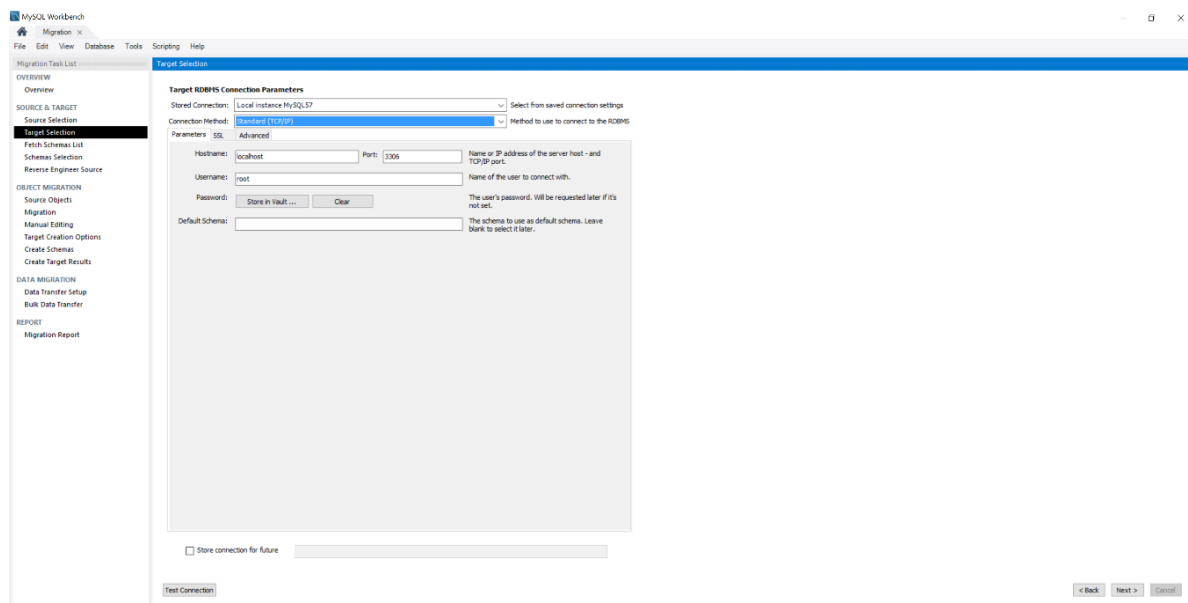
Aparece la pantalla de bienvenida de la herramienta de migración, dar clic en *Start Migration*.



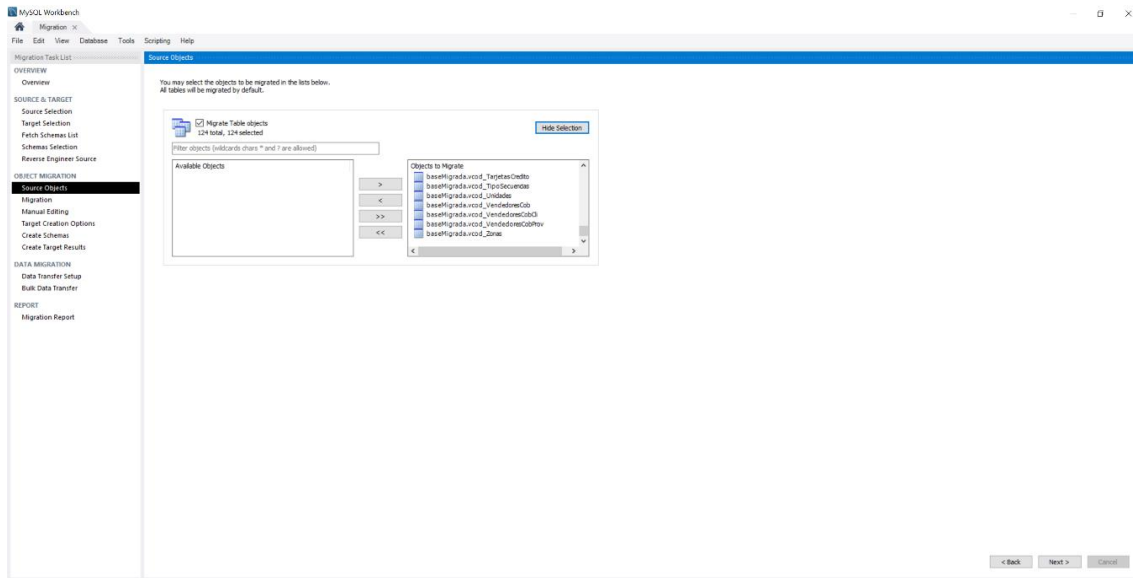
Aparece la pantalla para configurar el tipo de base de datos y la el método de conexión, para el caso del proyecto seleccionar en *Database System* la opción *Sybase SQLAnywhere* y en *Connection Method* seleccionar el nombre con el que se guardo la conexión ODBC, que en esta caso es *baseMigrada (SQL Anywhere 16)* y dar clic en *Next*.



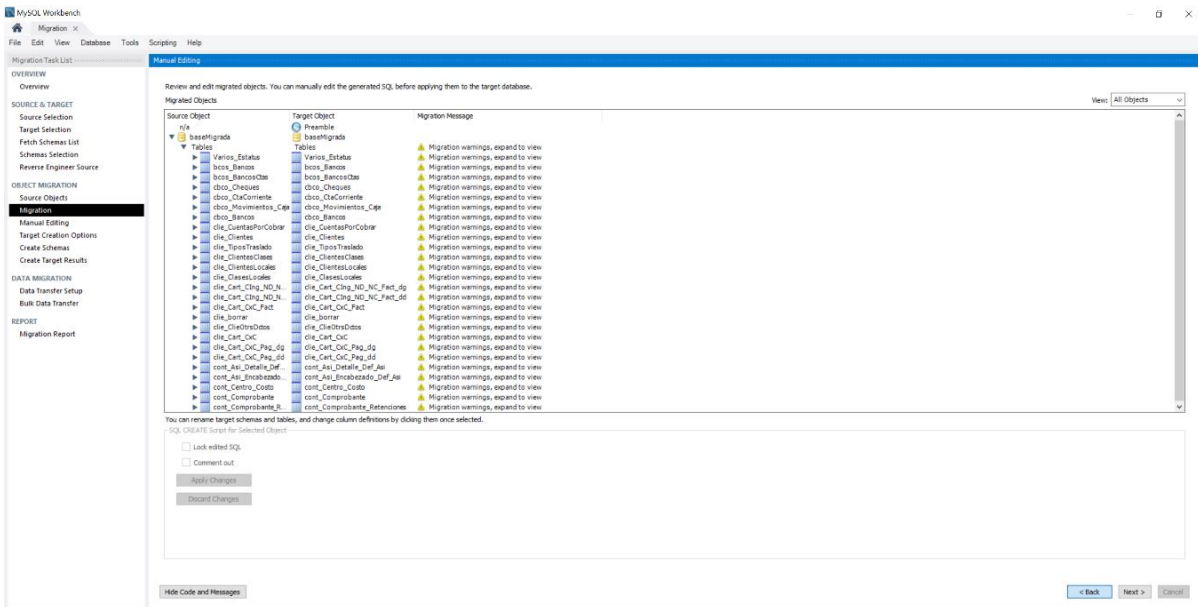
A continuación, se debe configurar la conexión de la instancia en donde se migrará la base de datos, para el caso del proyecto dejar las opciones por defecto y dar clic en *Next*.



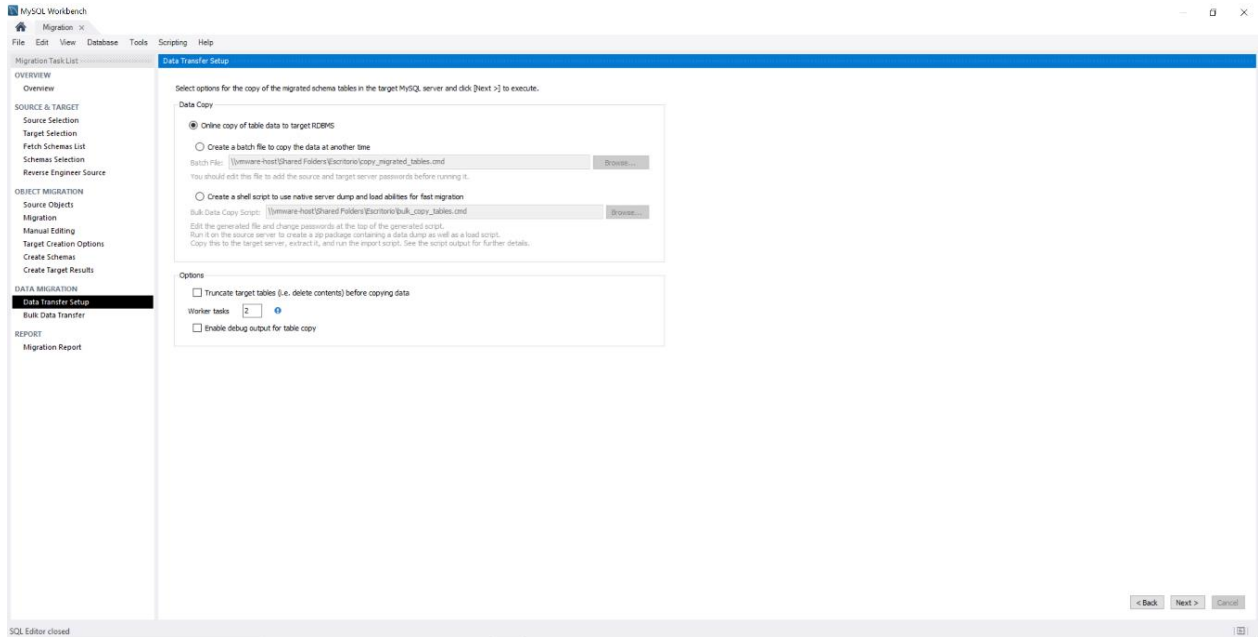
A continuación, se presenta la pantalla con las tablas que se migran, permitiendo seleccionar las que desea el usuario. Dar clic en *Next*.



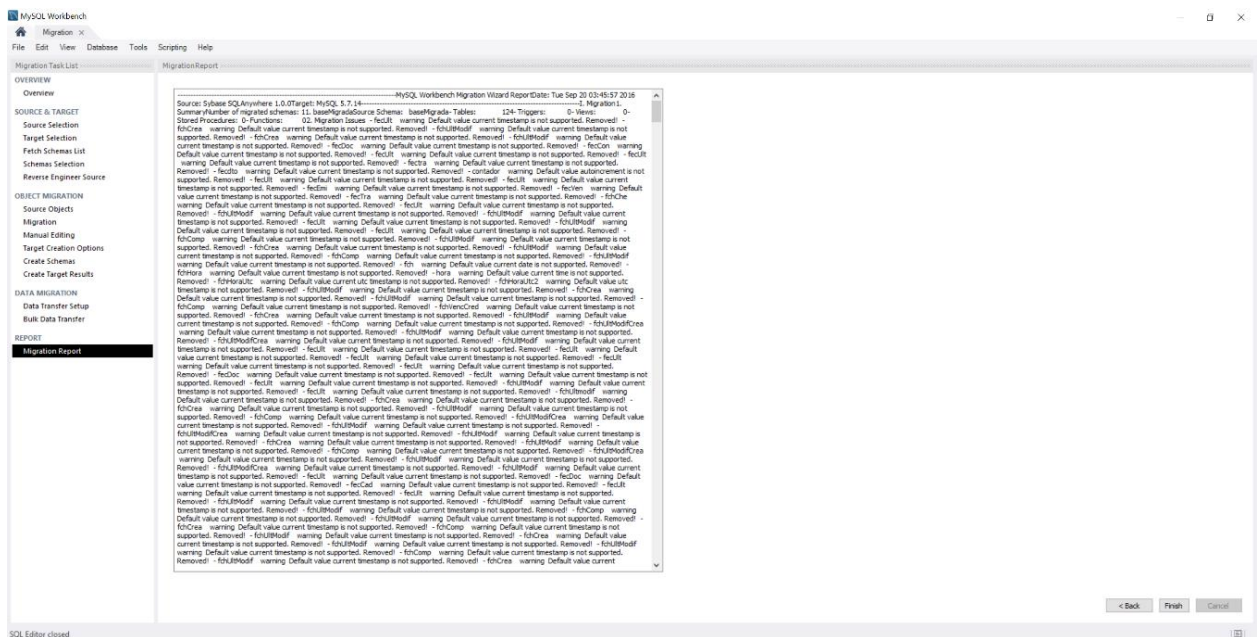
La pantalla siguiente permite corregir errores comunes que pueden presentar ciertos tipos de datos respecto de una u otro gestor de base de datos, para el caso del proyecto se corrige el dato de tipo *date time* y el manejo de un dato por defecto para ingresar el usuario logeado. Dar clic en *Next*.



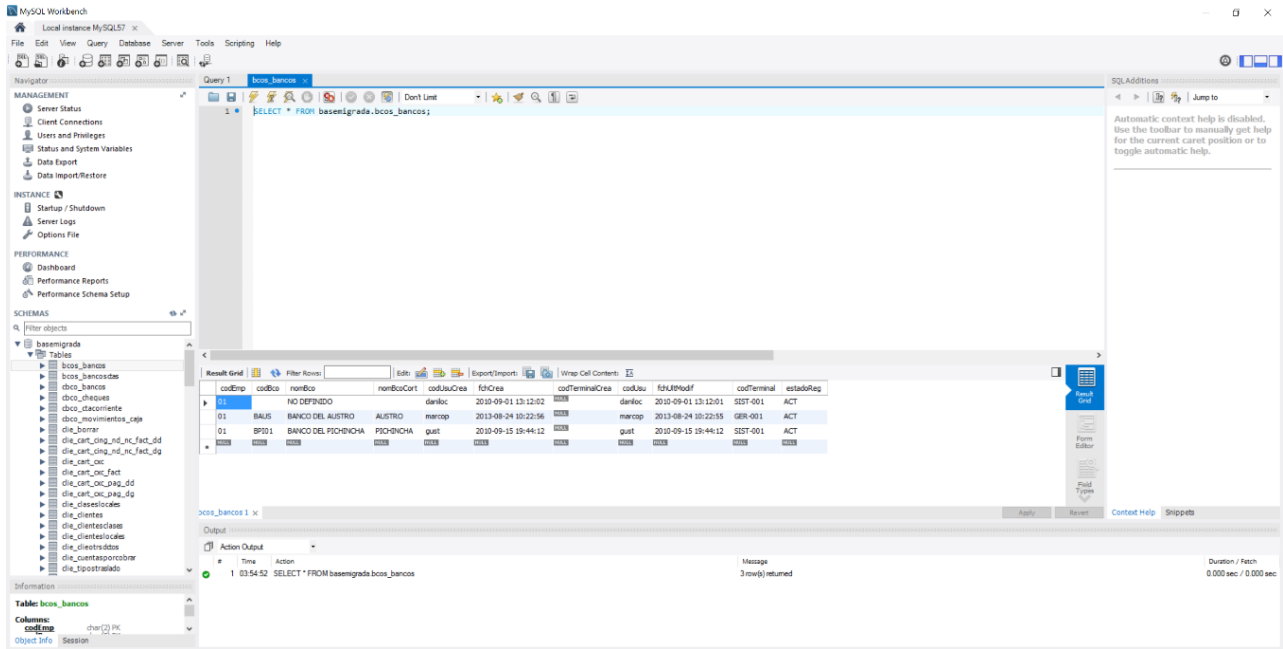
Una vez que se crean las tablas, se procede a migrar los datos. Dar clic en *Next*.



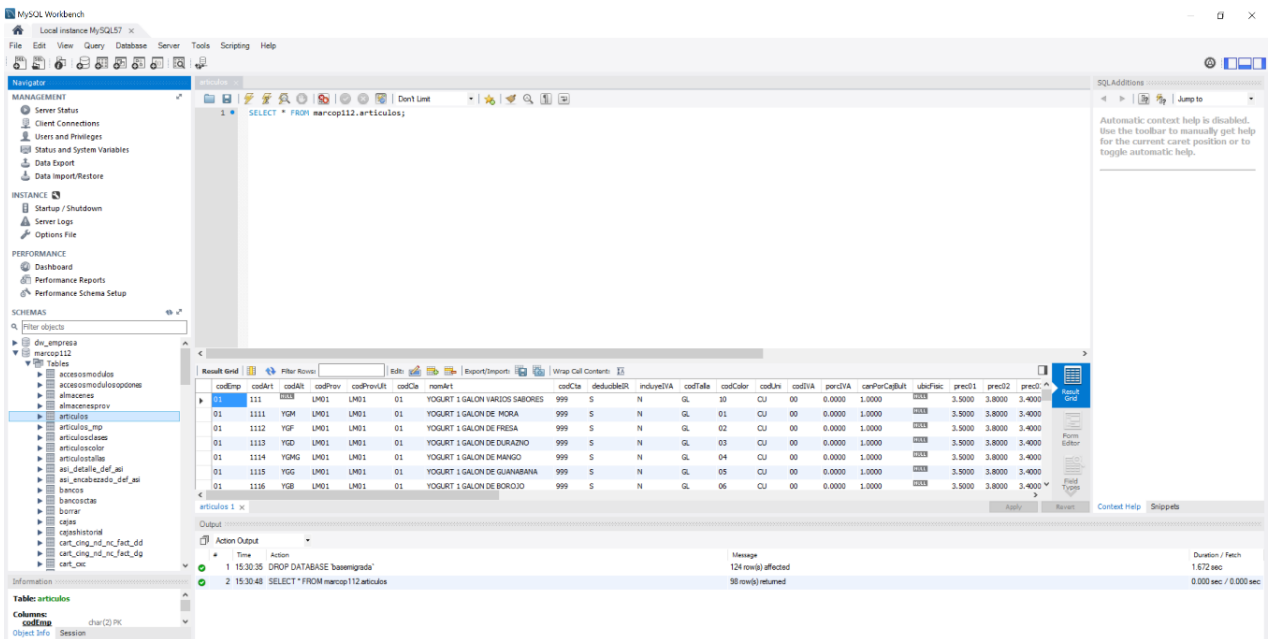
Finalmente, se presenta un informe de todo el proceso de migración de la base de datos, para terminar, dar clic en *Finish* e ingresar a la instancia de MySQL para visualizar la base de datos y comprobar que todo este correcto.



Al ingresar a la instancia de MySQL, se puede comprobar que las tablas y los datos fueron migrados correctamente.



Para fines practicos del proyecto, se modifico el archivo script que genero la herramienta de migracion de MySQL, como se muestra en la siguiente imagen.



A.2. Pentaho Community Edition (CE)

A.2.1. Requerimientos del sistema

Los requerimientos que se detallan a continuación se obtuvieron de la página oficial de Pentaho Community Edition (CE). Para las herramientas Pentaho Data Integration, Pentaho Schema Workbench y Pentaho Report Designer se debe cumplir los siguientes requerimientos:

Hardware - 32 o 64 bits	Sistema Operativo – 32 o 64 bits
<p>Procesador:</p> <ul style="list-style-type: none"> • Apple Macintosh Dual-Core, solo 64 bits • Intel EM64T o AMD64 Dual-Core <p>RAM: 2 GB RAM para las herramientas de diseño, Pentaho Data Integration requiere 2 GB dedicado.</p> <p>Espacio de disco: 2 GB libres después de la instalación.</p> <p>Tamaño mínimo de pantalla: 1280 x 960</p>	<ul style="list-style-type: none"> • Apple Macintosh OS 10.8 & 10.9, solo 64 bit. • Microsoft Windows 7 & 8 & 10 • Ubuntu Desktop 12.x & 14.x • iPad 2, 3, & Air (iOS 6.x & 7.x)

En la siguiente tabla se muestran los requerimientos del sistema para la herramienta Pentaho Business Analytics Platform:

Hardware - 64 bits	Sistema Operativo – 64 bits
<p>Procesador:</p> <ul style="list-style-type: none"> • Apple Macintosh Pro Quad-Core o Macintosh Mini Quad-Core • Intel EM64T or AMD64 Dual-Core <p>RAM: 8 GB con 4 GB dedicado a servidores Pentaho.</p> <p>Espacio de disco: 20 GB libres después de la instalación.</p>	<ul style="list-style-type: none"> • Apple Macintosh OS X Server 10.8 & 10.9 • CentOS Linux 5 & 6 • Microsoft Windows 2008 Server R2, 2012 Server R2 • Red Hat Enterprise Linux 5 & 6 • Solaris 10 • Ubuntu Server 12.x & 14.x

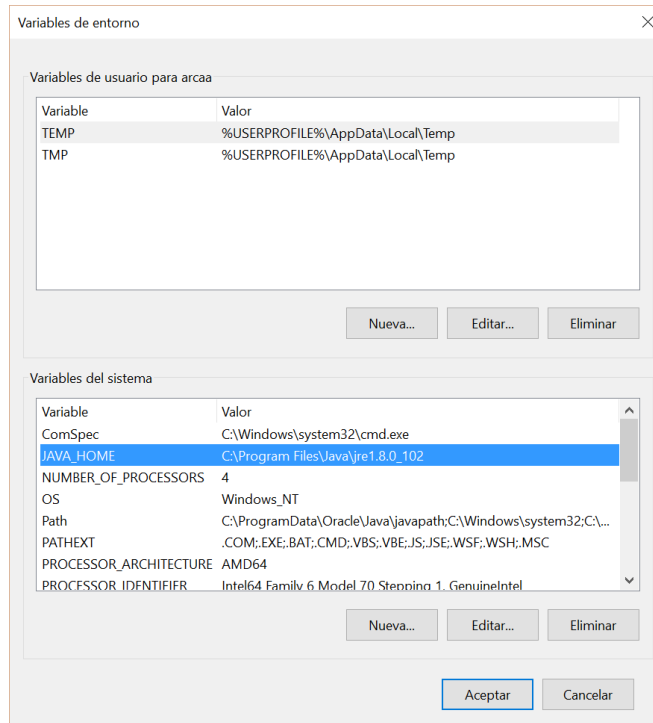
Además, es importante mencionar que todos los productos de Pentaho están desarrollados en Java, por lo tanto es necesario tener instalado el Java Development Kit (JDK), el cual lo puede descargar del siguiente link:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Para más información sobre su instalación, visitar el siguiente link oficial de Oracle:

<http://docs.oracle.com/javase/7/docs/webnotes/install/windows/jdk-installation-windows.html>

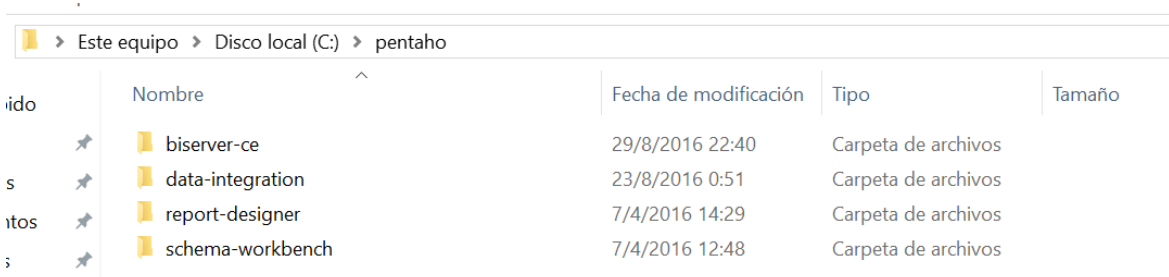
Una vez instalado el JDK, se procede a agregar la variable de entorno JAVA_HOME en caso de ser necesario, como se muestra en la siguiente figura, el cual debe contener el valor que corresponde al directorio donde se instaló el JDK. Para acceder a las variables de entorno en Windows, se debe seguir la ruta: Panel de Control → Sistema → Configuración Avanzada → Variables de Entorno. A continuación, se muestra la configuración de la variable de entorno JAVA_HOME:



A.2.2. Instalación

La versión que se utiliza es la 6.1 de Pentaho BI, lo primero es dirigirse a la página oficial de Pentaho Community Edition (CE) y descargar los archivos correspondientes a Business Analytics Platform, Data Integration, Report Designer y Schema Workbench. En el siguiente link se puede descargar: <http://community.pentaho.com>

La instalación de cada herramienta simplemente consiste en descomprimir los archivos Zip en un directorio de instalación de preferencia del usuario. Para el caso del proyecto se crea una carpeta en el disco C llamada Pentaho y dentro de ella se descomprime los archivos, como se visualiza en la siguiente imagen:



Nombre	Fecha de modificación	Tipo	Tamaño
biserver-ce	29/8/2016 22:40	Carpeta de archivos	
data-integration	23/8/2016 0:51	Carpeta de archivos	
report-designer	7/4/2016 14:29	Carpeta de archivos	
schema-workbench	7/4/2016 12:48	Carpeta de archivos	

A.2.2. Ejecución

El servidor de Pentaho se puede iniciar y detener manualmente mediante los scripts start-pentaho.bat y stop-pentaho.bat que se encuentran dentro de la carpeta biserver-ce. Además, este está pre configurado para iniciar sobre el contenedor de Apache Tomcat, el cual se encuentra dentro de la misma carpeta. Cuando se inicie el servidor, este será accesible desde un navegador a través de la url por defecto: <http://localhost:8080/pentaho>. Para poder cambiar el puerto por defecto donde se ejecuta, se debe dirigir a la carpeta donde se descomprime el biserver-ce y buscar la siguiente ruta: tomcat → conf, dentro de esta carpeta se encuentra un fichero server.xml, donde se puede modificar el puerto en la siguiente sección del fichero:

```
<Connector URIEncoding="UTF-8" port="8090" protocol="HTTP/1.1"
connectionTimeout="20000"
redirectPort="8443" />
```

Para el caso del proyecto, se cambia el puerto a 8090, ahora el servidor inicia en el navegador a través de la url: <http://localhost:8090/pentaho>.

Para el proceso de extracción, transformación y carga de datos se utiliza Data Integration, también llamado Kettle el cual se inicia mediante la ejecución del archivo Spoon.bat que se encuentra dentro de la carpeta descomprimida de Data Integration.

La herramienta de Schema Workbench utilizada para la construcción de los cubos dimensionales, se inicia mediante la ejecución del archivo workbench.bat que se encuentra dentro de la carpeta de Schema Workbench.

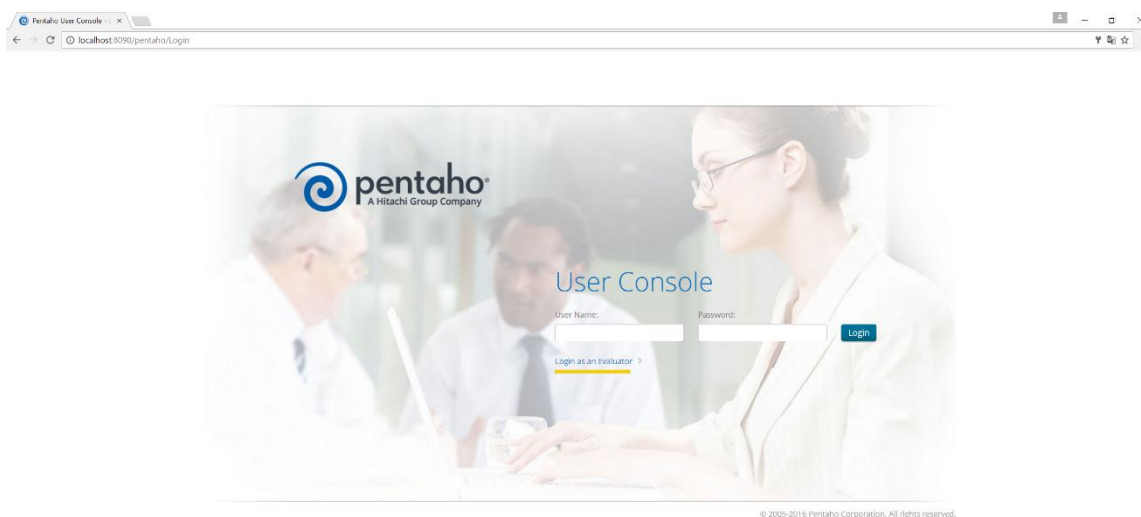
Finalmente, para ejecutar la herramienta de Report Designer para la generación de reportes personalizados, se debe ejecutar el archivo report-designer.bat que se encuentra dentro de la carpeta Report Designer.

Anexo B

MANUAL DE USUARIO: GENERACIÓN DE REPORTE

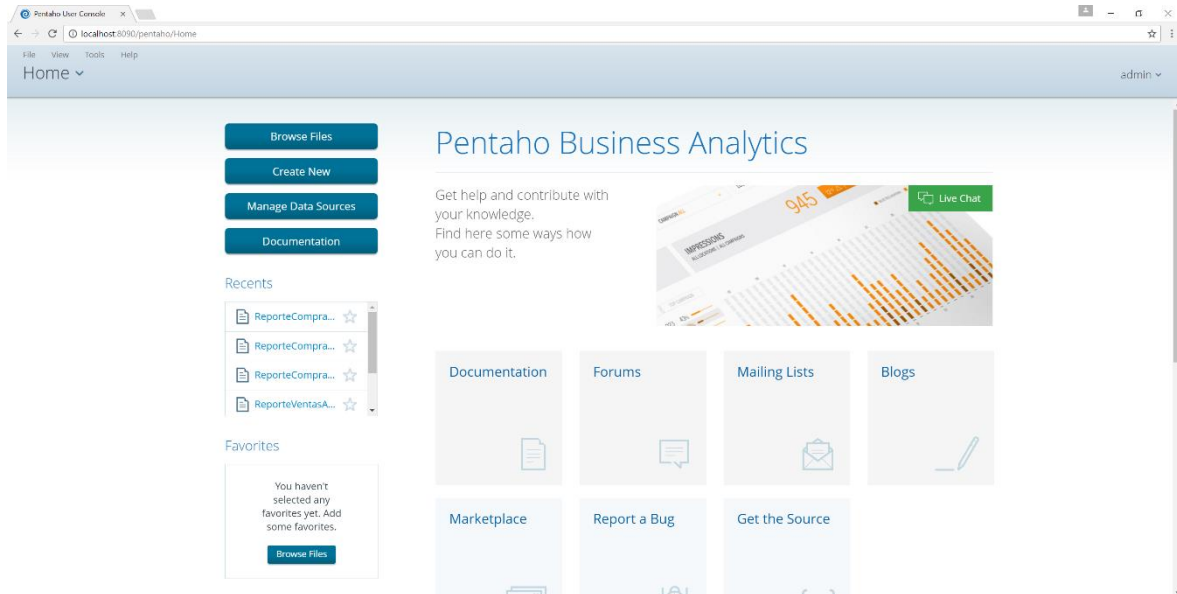
La generación de reportes personalizado que provee la Suite de Pentaho Community Edition (CE) permite al usuario obtener información de una manera fácil, resumida y gráficamente, con el fin de ayudar a la toma de decisiones dentro de una empresa. Los reportes que se presentan en este manual, fueron previamente establecidos según las necesidades consultadas al gerente general de la empresa. Para la generación de reportes es necesario iniciar el servidor de Pentaho, para más información revisar el Anexo A: Manual técnico. A continuación, se describe el proceso para generar un reporte:

Paso 1: Abrir un navegador: Microsoft Edge, Google Chrome, Firefox, Opera o cualquier otro navegador con el que este familiarizado y digitalizar en el url: <http://localhost:8090>²⁵, se abre la pantalla de inicio de la consola de usuario de Pentaho, como se muestra en la siguiente imagen:

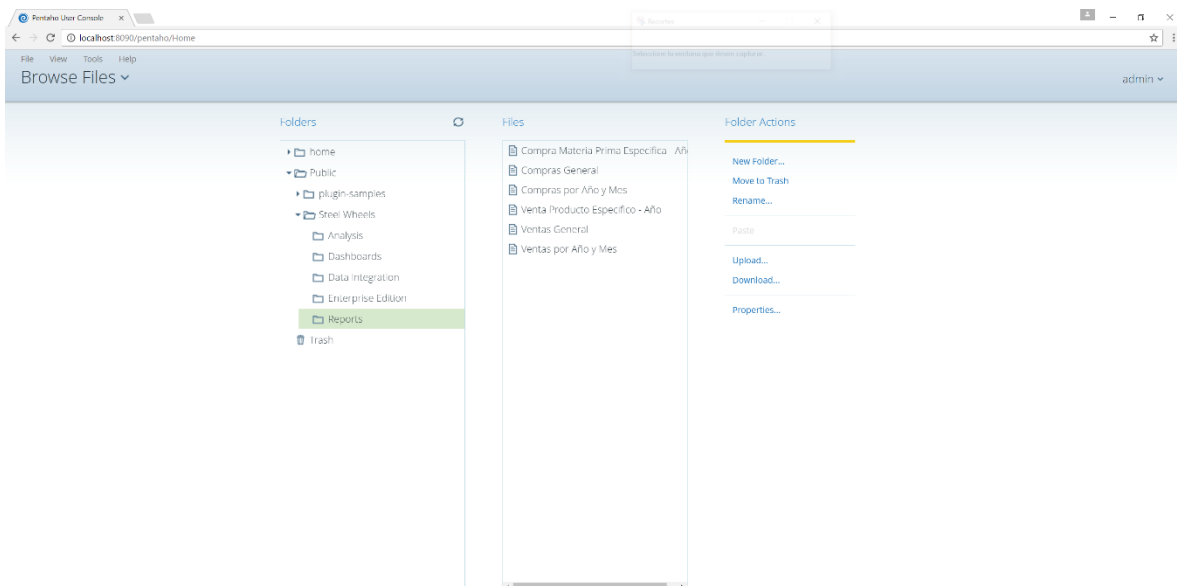


²⁵ El valor depende de la configuración que se haga al servidor de Pentaho, para más información revisar el manual técnico.

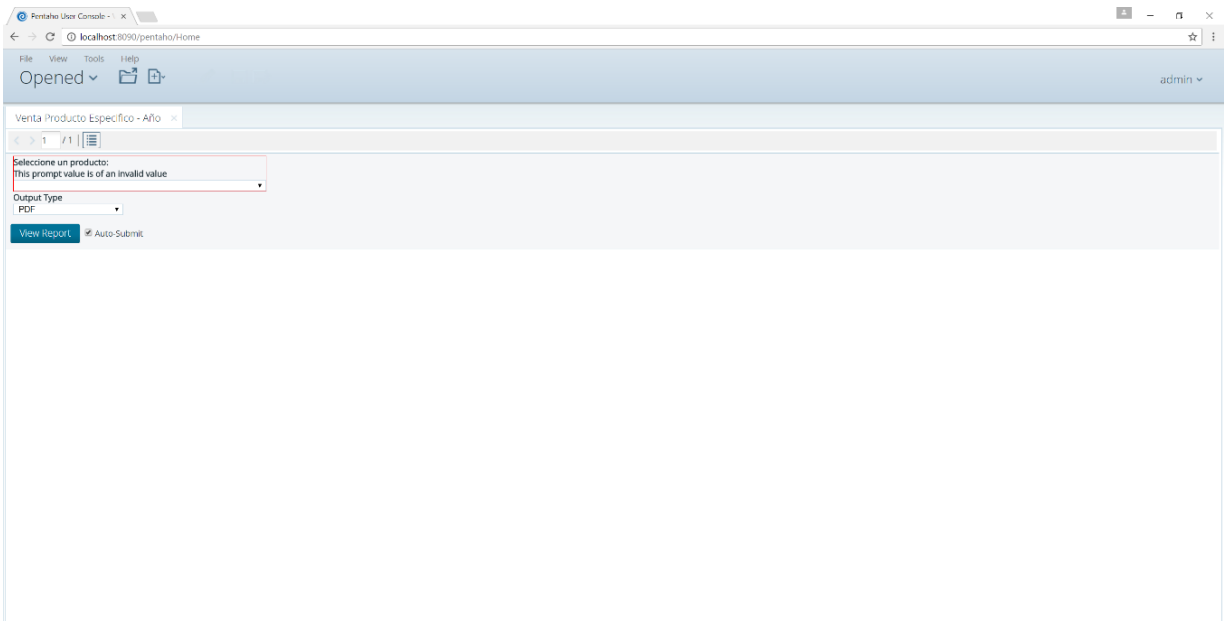
Paso 2: Ingresar el usuario y contraseña del rol de administrador que por defecto es usuario: admin y password: password, y dar clic en *Login* para redireccionar a la pantalla de inicio de la consola de Pentaho.



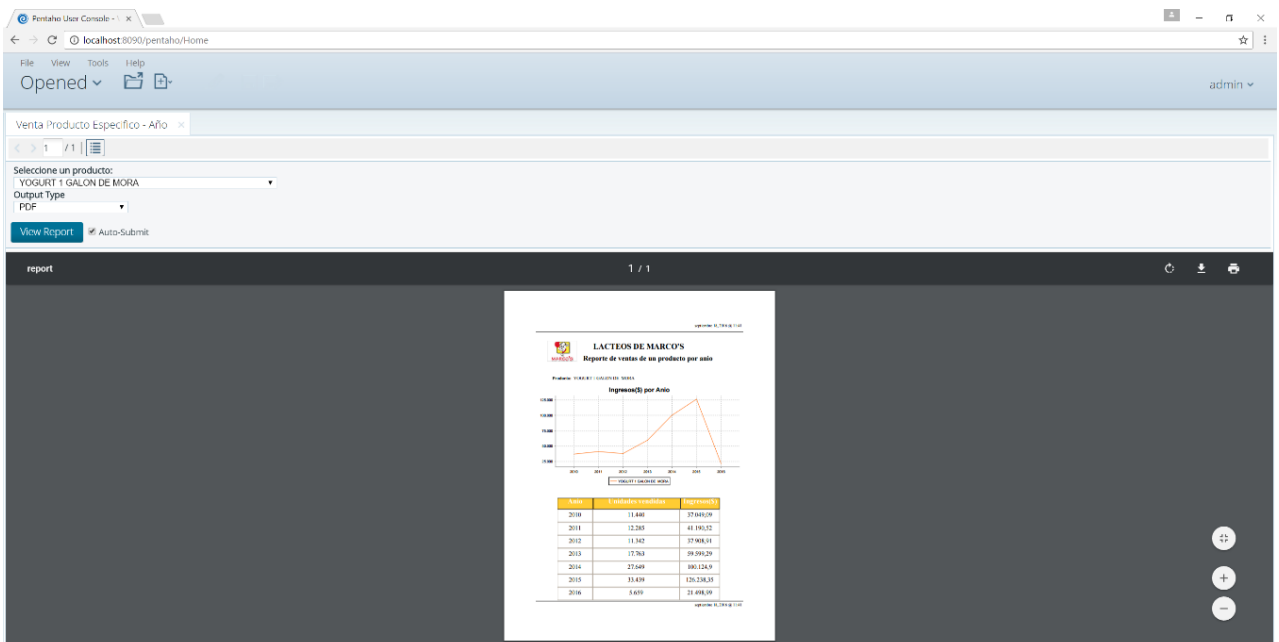
Paso 3: En la parte superior izquierda se encuentra la opción *Browse Files*, dar clic y dirigirse a la siguiente ruta: Public → Steel Wheels → Report. Ahí se muestran los reportes previamente establecidos según las necesidades de información del departamento financiero de la empresa.



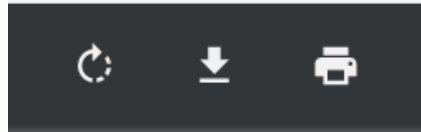
Paso 4: Para el ejemplo, se muestra como generar el reporte de ventas de un producto específico por año, para ello se selecciona el archivo de *Ventas Producto Especifico – Año* y en la parte superior derecha seleccionar la opción *Open*, el cual redirige al reporte y muestra las opciones que el usuario puede ingresar como parámetro, que en este caso es *Seleccione un producto* y el tipo de formato de salida.



Paso 5: Seleccionar el producto y formato de salida y dar clic en la opción View Report, en este caso se seleccionó el producto YOGURT 1 GALON DE MORA y el formato PDF.



El archivo generado se puede girar, descargar e imprimir; las opciones se encuentran en la parte superior derecha del reporte.



Finalmente, si el usuario desea generar otro tipo de reporte que se encuentran en la carpeta de *Report*, debe seguir los pasos a partir del número 4, teniendo en cuenta que los parámetros de ingreso para generar el reporte cambian dependiendo del tipo de reporte que se desea generar.

Anexo C



Quito, 19 de septiembre del 2016

INGENIERO
GUSTAVO CHAFLA
DECANO DE LA FACULTAD DE INGENIERIA
PONTIFICIA UNIVERSIDAD CATOLICA DEL ECUADOR
PRESENTE. -

De mis consideraciones

Yo, Ing. Marco Vinicio Proaño Chiluita propietario de la empresa LÁCTEOS DE MARCO'S, ubicada en la provincia de Tungurahua, Pillaro, Km. 1 vía La Primavera, informo a ustedes que el tema de disertación "ANÁLISIS COMPARATIVO DE HERRAMIENTAS OPEN SOURCE PARA SOLUCIONES DE INTELIGENCIA DE NEGOCIOS (BI) Y SU APLICACIÓN PRÁCTICA PARA LA TOMA DE DECISIONES EN LA EMPRESA LÁCTEOS DE MARCO'S", realizada por los señores Adrián Remigio Carrión Albán con cédula de identidad N° 0704721505 y Mark Christopher Torres Sosa con cédula de identidad N° 1713851648, ha concluido y han cumplido con lo propuesto en el proyecto.

Se realizó una demostración sobre la generación de reportes personalizados y exploración de la información a través de la utilización de la suite de Pentaho, evidenciando la gran ventaja que conlleva una solución de Inteligencia de Negocios (BI) para la empresa. Dando por terminado el tema de disertación propuesto por los estudiantes con la empresa.

Atentamente,

Ing. Marco Vinicio Proaño Chiluita