

**PONTIFICA UNIVERSIDAD CATÓLICA DEL ECUADOR SEDE  
ESMERALDAS**



**CARRERA:**

**INGENIERÍA DE SISTEMAS Y COMPUTACIÓN  
PREVIO AL GRADO ACADÉMICO DE INGENIERÍA DE SISTEMAS Y  
COMPUTACIÓN**

**TEMA DE INVESTIGACIÓN:**

**SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL APOYO DE LA TOMA  
DE DECISIONES EN TEMAS DE CONTAMINACIÓN MEDIOAMBIENTAL DE  
LOS RECURSOS AIRE, AGUA Y SUELO.**

**LINEA DE INVESTIGACIÓN:**

**AUTOMATISMOS Y APLICACIONES INTELIGENTES**

**PREVIO A LA OBTENCIÓN DE TÍTULO DE:**

**INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

**AUTOR:**

**LEMA SIVINTA MERY GUADALUPE**

**ASESOR:**

**PICO VALENCIA PABLO (PhD)**

**ESMERALDAS, 2022**

## TRIBUNAL DE GRADUACIÓN

**Título:** Sistema de inteligencia de negocios para el apoyo de la toma de decisiones en temas de contaminación medioambiental de los recursos aire, agua y suelo.

**Autor(a):** Mery Guadalupe Lema Sivinta

Ph.D. Pablo Pico Valencia

f. \_\_\_\_\_

**Asesor**

Mgt. Manuel Nevárez Toledo

f. \_\_\_\_\_

**Lector #1**

Mgt. Jaime Sayago Heredia

f. \_\_\_\_\_

**Lector #2**

Mgt. Xavier Quiñónez Ku

f. \_\_\_\_\_

**Coordinador de carrera**

## **AUTORÍA**

Yo, **Lema Sivinta Mery Guadalupe** con número de cédula de identidad **0803606011** manifiesto que mediante la presente investigación sobre el tema “**SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL APOYO DE LA TOMA DE DECISIONES EN TEMAS DE CONTAMINACIÓN MEDIOAMBIENTAL DE LOS RECURSOS AIRE, AGUA Y SUELO.**” Los resultados obtenidos como tesis de grado, previo a la obtención del título de “**INGENIERA EN SISTEMA y COMPUTACIÓN**” son de total responsabilidad del autor, y que se ha respetado las fuentes de información consultadas, realizando las citas correspondientes y los resultados alcanzados son totalmente personal, único y legítimo. Al mismo tiempo declaro que todo el contenido incluyendo resultados, discusión, conclusión, recomendación y otros efectos legales y académicos que se desglosan, son y serán exclusiva responsabilidad legal y académica del autor y de la PUCESE.

---

Lema Sivinta Mery Guadalupe

0803606011

## **DEDICATORIA**

*Este trabajo de lo dedico con cariño y felicidad a Dios por haber sido mi guía y participar en todos los episodios de mi vida.*

*A mis padres, quienes con sacrificio supieron motivarme moral y materialmente para la culminación de mi vida profesional. ¡Gracias sin ustedes no hubiera sido posible llegar a este sueño!*

*A mis abuelos quienes con consejos supieron impulsarme para continuar con mis objetivos y así llegar al éxito alcanzado.*

*Para todo ustedes dedico este título con amor, cariño y humildad.*

***Mery Lema***

## **AGRADECIMIENTO**

*A Dios por amarme y bendecirme todos los días de mi vida a pesar de las adversidades que tuve en el largo de vida universitaria. Por haberme brindado las fuerzas para llegar a la meta propuesta y a la vez compartir este momento de felicidad con mis padres por ser el pilar fundamental de mi vida ¡Gracias por todo el apoyo queridos padres los amo!*

*En especial, quiero agradecer a mi asesor PhD. Pablo Pico Valencia, por compartir sus conocimientos y guiarme en el proceso de la presente tesis.*

**DOCENTES:** *Quienes con sabiduría supieron transmitir sus conocimientos para mi formación académica disciplinario y sembrar una semilla que a la postre dará sus frutos, para una profesional integra y capaz.*

**PUCESE:** *Por brindarme la oportunidad para mi formación académica y alcanzar la Educación Superior.*

***Mery Lema***

## Índice general

TRIBUNAL DE GRADUACIÓN	I
AUTORÍA	II
DEDICATORIA	III
AGRADECIMIENTO	IV
RESUMEN	VIII
ABSTRACT	IX
INTRODUCCIÓN	10
Presentación del problema-----	10
Planteamiento de problema-----	11
Justificación -----	12
Objetivos -----	13
General -----	13
Específicos -----	13
Publicación asociada al trabajo -----	13
CAPÍTULO I: MARCO TEÓRICO	14
1.1. Bases teóricas -----	14
1.1.1. Inteligencia de Negocios-----	14
1.1.2. Datawarehouse -----	17
1.1.3. Contaminación Medioambiental-----	22
1.1.4. Suite Pentaho -----	26
1.2. Antecedentes -----	28
1.3. Fundamentos legales-----	30
CAPITULO II: METODOLOGÍA	32
2.1. Delimitación del estudio-----	32
2.2. Tipos de estudio -----	32
2.3. Métodos de investigación -----	33
2.4. Operacionalización de variables -----	33
2.5. Población y muestra-----	36
2.6. Técnicas e instrumentos de recolección de datos -----	36
2.7. Técnicas de procesamiento y análisis de datos-----	37
2.8. Normas éticas -----	37
CAPITULO III: RESULTADOS	38
3.2. Arquitectura de flujo de datos del sistema BI -----	39
3.2.1. Fuentes de datos -----	39
3.2.2. Base de datos STAGE -----	39
3.2.3. Datawarehouse-----	39

3.3. Procesamientos ETL -----	41
3.4. Respuestas a las preguntas de negocio -----	43
3.3. Cuadros de mando ( <i>dashboard</i> ) -----	53
CAPITULO IV: DISCUSIÓN	55
CAPITULO V: CONCLUSIONES Y RECOMENDACIONES	57
5.1. Conclusiones -----	57
5.2 Recomendaciones -----	58
Referencias	59
Anexo	66
Anexo 1. Ficha de observación de la Herramienta Suites de Pentaho -----	67
Anexo 2. Ficha de preguntas de instrumento de la entrevista -----	68
Anexo 3. Tutorial de diseño de cubo de Olap y Reporte. -----	69

## Índice de tablas

<b>TABLA 1.</b> MATERIALES DE MÁXIMO PERMISIBLE DEL AIRE [50]. -----	23
<b>TABLA 2.</b> PRINCIPALES CONTAMINANTES DEL AGUA [55]. -----	24
<b>TABLA 3.</b> MATERIALES DE MÁXIMO PERMISIBLE DEL SUELO [56].-----	24
<b>TABLA 4.</b> VARIABLES E INDICADORES SUJETOS A ESTUDIO.-----	34
<b>TABLA 5.</b> FORMULACIÓN DE LAS PREGUNTAS DE EN BI-----	38

## Índice de figuras

<b>FIGURA 1.</b> SOPORTE DE LA INTELIGENCIA DE NEGOCIOS [17]. -----	14
<b>FIGURA 2.</b> FASES DE LA TOMA DE DECISIONES [21]. -----	15
<b>FIGURA 3.</b> FASES DE LA TOMA DE DECISIONES [21]. -----	16
<b>FIGURA 4.</b> ESQUEMA DE DATAWAREHOUSE [29]. -----	17
<b>FIGURA 5.</b> ESQUEMA DE DATAMART [33]. -----	18
<b>FIGURA 6.</b> ARQUITECTURA DE LOS PROCESOS DE ETL [34]. -----	19
<b>FIGURA 7.</b> EJEMPLO DE CUBO OLAP [37].-----	19
<b>FIGURA 8.</b> ARQUITECTURA DE FLUJO DE DATOS [41]. -----	20
<b>FIGURA 9.</b> METODOLOGÍA RALPH KIMBALL [44]. -----	21
<b>FIGURA 10.</b> ARQUITECTURA BILL INMON [47]. -----	22
<b>FIGURA 11.</b> ARQUITECTURA DE PENTHAHO BI SUITE [62]. -----	26
<b>FIGURA 12.</b> BASE DE DATOS DBCONTAMINANTE.-----	39
<b>FIGURA 13.</b> BASE DE DATOS STAGE. -----	40
<b>FIGURA 14.</b> BASE DE DATOS DATAWAREHOUSE -----	40
<b>FIGURA 15.</b> ETL PARA LA CARGA DE LA TABLA CONTAMINANTES DESDE LA BASE DE DATOS DBCONTAMINANTES HACIA STAGE. -----	41
<b>FIGURA 16.</b> PROCESO ETL PARA CARGA DE DATOS DE LA DIMENSIÓN PARAMETRO DE STAGE AL DATAWAREHOUSE. -----	42
<b>FIGURA 17.</b> PROCESO ETL PARA LA GENERACIÓN AUTOMÁTICA DE LA DIMENSIÓN TIEMPO. -----	42
<b>FIGURA 18.</b> PROCESO ELT DE LA TABLA DE HECHO. -----	42
<b>FIGURA 19.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 1 -----	43
<b>FIGURA 20.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 2-----	44
<b>FIGURA 21.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 3-----	45
<b>FIGURA 22.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 4-----	46
<b>FIGURA 23.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 5-----	47
<b>FIGURA 24.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 6-----	48
<b>FIGURA 25.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 7-----	49
<b>FIGURA 26.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 8-----	50
<b>FIGURA 27.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 9-----	51
<b>FIGURA 28.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 10 -----	52
<b>FIGURA 29.</b> CUBO OLAP Y RESULTADOS A LA PREGUNTA DE NEGOCIO 11 -----	53
<b>FIGURA 30.</b> DASHBOARD Y REPORTE-----	54
<b>FIGURA 31.</b> CONTROL DE MONITOREO-----	54

## RESUMEN

En la vida cotidiana del ser humano, éste enfrenta a la incertidumbre de los nuevos avances tecnológicos que le permiten detectar si su estilo de vida en pleno siglo XXI es realmente el correcto o la causante de llevarle al desastre. Entre ella se destaca aquellas innovaciones que se enfocan al cuidado de la salud, en un contexto poblacional al cuidado de sus factores ambientales tales como es las fuentes de agua, la fertilidad del suelo y la fuerza del aire que respira. Por ello, en esta investigación, pensando en el medio ambiente, se ha propuesto aplicar un sistema de inteligencia de negocios para gestionar datos ambientales, y suministrar información y conocimiento para favorecer la toma de decisiones en conexión con los controles de la contaminación ambiental.

Para comenzar, en la metodología se realizó una revisión bibliográfica acerca de las diferentes infraestructuras, así como de las fuentes de datos más utilizadas para el proceso de la toma de decisiones. A través de una entrevista realizada a un especialista en Gestión Ambiental, se determinó qué normativas y parámetros, así como de fuentes de datos que se podrían tomar para usar por parte del sistema.

Para la prueba se utilizó una cantidad de 40 registros comprendidos entre 2012 hasta 2021 en el Ecuador. Hay que remarcar que fue de una fuente de datos con valores totalmente verídicos que contiene los tres aspectos en los que se basa este estudio (agua, suelo y aire).

Los resultados obtenidos a partir del análisis de los datos de los recursos planteados permitieron determinar los índices de contaminación ambiental del aire, agua y suelo, como se tenía planificado. No obstante, los datos solo sirvieron para probar el sistema.

En conclusión, a través de esta investigación se determinó que la utilización de Pentaho Suite para el proceso del análisis de negocio tiene un rango de credibilidad y funcionalidad.

## **ABSTRACT**

In the daily life of the human being, he faces the uncertainty of the new technological advances that allow him to detect if his lifestyle in the XXI century is really the correct one or the cause of leading him to disaster. Among them, those innovations that focus on health care stand out, in a population context, taking care of its environmental factors such as water sources, soil fertility and the force of the air it breathes. For this reason, in this research, thinking about the environment, it has been proposed to apply a business intelligence system to manage environmental data, and provide information and knowledge to favor decision-making in connection with environmental pollution controls.

To begin with, in the methodology, a bibliographic review was carried out about the different infrastructures, as well as the most used data sources for the decision-making process. Through an interview with a specialist in Environmental Management, it was determined what regulations and parameters, as well as data sources that could be taken to be used by the system.

For the test, an amount of 40 records from 2012 to 2021 in Ecuador was used. It should be noted that it was from a data source with totally true values that contains the three aspects on which this study is based (water, soil and air).

The results obtained from the analysis of the data of the proposed resources made it possible to determine the rates of environmental contamination of the air, water and soil, as planned. However, the data was only used to test the system.

In conclusion, through this investigation it was determined that the use of Pentaho Suite for the business analysis process has a range of credibility and functionalit

# INTRODUCCIÓN

## Presentación del problema

En los actuales años se han comprobado grandes avances en el campo tecnológico, de tal manera que la Inteligencia Artificial (IA) se ha desarrollado sustancialmente. Varias técnicas de la IA tales como el aprendizaje automático también conocido como aprendizaje artificial (ML, por sus inicios en inglés *Machine Learning*) [1], la minería de datos (MD) [2], y la inteligencia de negocios (BI, por sus inicios en inglés) [3], están en pleno auge gracias a la proliferación de datos que se han originado en el ciberespacio [4].

La Inteligencia de Negocios permite la integración, análisis y presentación de datos e información con el propósito de brindar un mejor soporte a la toma de decisiones principalmente en el dominio empresarial [5]. Un ejemplo de su uso en el sector empresarial podría ser el estudio y posterior análisis de sus clientes, con el fin de mejorar la toma de decisiones en cuanto a la apertura de una nueva sucursal o a su vez el cierre de alguna de las que ya existen [6].

Aunque la Inteligencia de Negocios ha sido ampliamente aplicada en las empresas, también se ha extendido a otros dominios como la educación, la industria, la salud y el Medio Ambiente. En el ámbito educativo, la inteligencia de negocios ha permitido manejar eficientemente las calificaciones de los estudiantes para integrar el funcionamiento de indicador clave de rendimiento (KPI), que motiva al estudiante a tener éxito en sus estudios [7]. También en el campo de gestión industrial, la Inteligencia de Negocios se ha utilizado para analizar la disposición de los procesos de la formación [8].

En cuanto a la salud, la Inteligencia de Negocios ha sido esencial para crear un complemento que mejore la toma de decisiones basadas en fuentes confiables, de tal manera que permitan a los médicos brindar atención de calidad y un tratamiento adecuado a sus pacientes [9]. Finalmente, en el campo del Medio Ambiente, la Inteligencia de Negocios ha permitido el almacenamiento de datos sobre los gases contaminantes para favorecer a la toma de decisiones en el ámbito de calidad ambiental [10].

La Inteligencia de Negocios en el sector del Medio Ambiente, puede aprovechar los datos generados por las estaciones de observación que se encargan de analizar la importancia del aire, agua, suelo, ruido, entre otras, con el fin de integrarlos, cruzarlos y finalmente analizarlos; de tal manera que se pueda proporcionar información relevante a la hora de la toma de decisiones en términos de administración ambiental. La presente

investigación se centró en usar las técnicas de Inteligencia de Negocios (específicamente las técnicas de análisis OLAP, procesamiento analítico en línea) para gestionar datos ambientales y suministrar información y conocimiento para favorecer la toma de decisiones en conexión con los controles de la contaminación ambiental.

### **Planteamiento de problema**

Actualmente, existen varias organizaciones que se preocupan por el mantenimiento constante y el cuidado del Medio Ambiente. Por tal motivo las autoridades se enfocan en el desarrollo y ejecución de distintos tipos de controles que consideran el seguimiento ambiental y para lograrlo se debe considerar la implementación de plataformas de observación de importancia del aire, agua y suelo, de modo que sensores especializados permitan conocer los niveles de contaminación existentes en las diferentes ciudades. Sin embargo, no es suficiente solo contar con datos, también se requiere de sistemas que permitan la estandarización, análisis y resumen de los datos para que los responsables de la toma de decisiones en el área ambiental puedan respaldar sus decisiones en base a experiencias pasadas.

Muchos de los datos capturados por las plataformas de monitoreo de contaminación ambiental se encuentran almacenadas en diferentes sistemas de información y formatos, lo que ocasiona que el análisis de las mismas resulte complejo, evidenciando la importancia de contar con herramientas especializadas en el medio ambiente, de tal forma que permitan llevar a cabo tareas de extracción, innovación y carga de datos de un modelo estándar.

La generación de datos de calidad ambiental involucra grandes volúmenes de datos, por lo que es necesario contar con herramientas especializadas que permitan resumir y visualizar la información. El no contar con estos mecanismos de visualización genera cierta dificultad para los usuarios que no son especializados en TIC al momento de entender la información en relación con los índices de contaminación medioambiental. Esto se debe a que la información no procesada no tendría ninguna relevancia a la hora de la toma de decisiones por parte de las autoridades responsables de la gestión del Medio Ambiente [11].

Los recursos del Medio Ambiente son importantes para la vida y muchos de ellos son no renovables, por lo que su cuidado es obligatorio. En base a los planteamientos anteriores, mediante la presente investigación se pretende responder a la siguiente

interrogante: ¿Son idóneos los sistemas de Inteligencia de Negocios para la toma de decisiones en torno a la calidad medioambiental? Adicionalmente, se plantea investigar ¿cuál es el alcance de los sistemas de inteligencia de negocios en el campo medioambiental? y a su vez determinar si éstos son aptos o no para ayudar a los profesionales de gestión ambiental y autoridades gubernamentales a proteger recursos como el aire, el agua y el suelo.

## **Justificación**

La ciencia de datos aplicada a los negocios ha ganado mucha atención en las comunidades de expertos de la tecnología de la información durante las dos últimas décadas. No obstante, las empresas se enfrentan a nuevos desafíos, no solo en la gestión de inteligencia empresarial, sino también en análisis de datos, que requiere nuevos enfoques para obtener conocimientos de contenidos contextualizados [12]. En este sentido, la Inteligencia de Negocios juega un papel importante en la toma de decisiones.

En el área ambiental, a través de sistemas de Inteligencia de Negocios es posible monitorear e identificar problemas ambientales como la contaminación del aire, el agua y el suelo [13]. Su uso permite resumir grandes volúmenes de datos con el fin de presentar un panorama del estado de estos recursos a departamentos que velan por su preservación. Este es un aspecto importante que motivó a la ejecución de este estudio.

El tema de investigación trae consigo una nueva modalidad para agilizar el procedimiento de toma de decisiones en el ámbito de contaminación medio ambiental ya que permite emplear instrumentos que garanticen la obtención de los datos de una forma óptima para generar la información [14] que puede ser comprendida por profesionales de un ámbito distinto al tecnológico como es el caso de profesionales en políticas públicas, ambiente, gestión, entre otros.

Por lo anteriormente expuesto, la importancia de llevar a cabo esta investigación radica en la utilidad que brindan los sistemas de inteligencia de negocios en la toma de decisiones en temas de contaminación medioambiental, esto debido a que la contaminación ambiental es un tema alarmante en todo el mundo, puesto que vivimos colapsados de algunas enfermedades causadas por el agua que consumimos periódicamente, y el aire contaminado que respiramos a diario [15]. Por tal razón, las autoridades requieren de herramientas que permitan tomar decisiones ante el impacto negativo que tiene la polución del aire, agua y suelo, para de esta manera ofrecer a la

ciudadanía un entorno que les permita vivir saludable, de tal modo que no se afecte su calidad de vida [15].

## **Objetivos**

### **General**

Desarrollar un sistema de inteligencia de negocios mediante la Suite de Pentaho para el apoyo en la toma de decisiones en aspectos de contaminación del aire, agua y suelo.

### **Específicos**

- a. Identificar los principales aspectos involucrados en la toma de decisiones de contaminación medioambiental para la valoración de los parámetros a monitorear.
- b. Conocer los principales componentes para el diseño de sistemas de inteligencia de negocios utilizando las herramientas Suite de Pentaho.
- c. Crear un sistema de inteligencia de negocios para la gestión de datos de contaminación medioambiental utilizando la Suite de Pentaho y la base de datos PostgreSQL.
- d. Evaluar la funcionalidad del sistema de negocios creado usando fuentes de datos disponibles de forma pública (local, nacional o internacional), o en su defecto datos generados de forma aleatoria.

## **Publicación asociada al trabajo**

Los resultados del trabajo planteado en este documento fueron publicados en el Libro de Actas del fruto del XII Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2021). Los datos de la publicación fueron presentados en el congreso y se describe como sigue:

Mery Lema, Pablo Pico-Valencia, Juan A. Holgado-Terriza. (2021). **Sistema de Inteligencia de Negocios para el Apoyo a la Toma de Decisiones de Calidad del Aire**. En: XII Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2021). Obras Colectivas de Tecnología 34. Editorial: Editorial Universidad de Alcalá. Páginas: 223-231. ISBN: 978-84-18979-68-2. (superó proceso de revisión por pares).

# CAPÍTULO I: MARCO TEÓRICO

## 1.1. Bases teóricas

En este apartado se describe de forma detallada las bases conceptuales relacionadas con la inteligencia de negocios, la polución del aire, agua y suelo, y finalmente, se describen las herramientas que conforman la Suite de Pentaho, una plataforma especializada en el progreso de métodos de inteligencia de negocios [16].

### 1.1.1. Inteligencia de Negocios

La inteligencia de negocios (conocida como BI por sus iniciales en inglés) es un grupo de estrategias, operaciones y materiales orientadas a la gestión y establecimiento de intelecto mediante los estudios de datos vigentes en una organización o compañía [17]. Es precisamente en la administración del conocimiento donde se sustentan estas estrategias que permiten seguir un conjunto de acciones que la empresa alcanza a emprender, y que le otorgan una ventaja sobre sus competidores, especialmente porque el precio incorporado a los productos o servicios que son resultado de estas acciones, evolucionan con eficiencia en su producción y alto nivel de desempeño que escasamente pueden ser replicadas por aquellas que no tienen estas estrategias o procesos [17].

Como operación de inteligencia de negocios, las maniobras que se incluyen en los sistemas de información, los métodos de descubrimiento y el progreso de los procedimientos permiten inventar un proceso de toma de decisiones inteligente [17], [18]. Con el propósito de comprender lo antes mencionado en la Figura 1 se muestra un esquema del proceso del soporte de la inteligencia de negocios y algunos de los elementos estratégicos empleados [19].

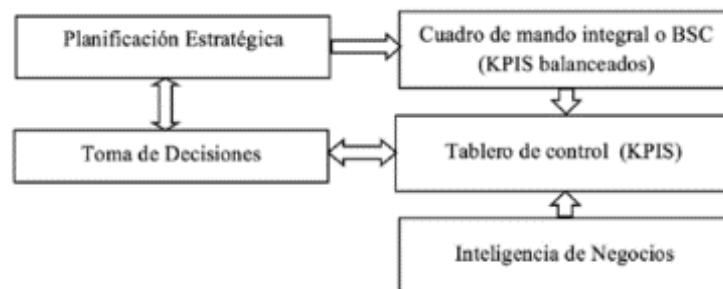


Figura 1. Soporte de la Inteligencia de Negocios [17].

El soporte de inteligencia de negocios consiste en diferentes fases que se fundamentan en la toma de decisiones que está enlazada a la planificación de estrategias que son fundamentales para BI, con el fin de tener un cuadro de mando integrado que permita unificar con un tablero de control, lo cual favorece a la toma de decisiones a diversos niveles organizacionales.

### 1.1.1.1. Tipos de decisiones

La toma de decisiones empresariales conforma un desarrollo de opción entre las diferentes elecciones factibles [20]. A partir de la opción de una alternativa se logrará pasar al proceso del alcance de actividades, aumento de las estrategias y formulación de programas y gestión de presupuestos [21].

Las decisiones adecuadas son de gran importancia, ya que se requiere de la utilización de los medios disponibles de un modo eficiente. La existencia de recursos limitados exige cierto nivel de responsabilidad puesto que se deben aprovechar al máximo todas las oportunidades en los diferentes niveles de la empresa. Las decisiones particulares respectivas a cada nivel de compromiso afectan de un modo continuo al progreso de todo el sistema organizativo y al cumplimiento de sus objetivos [20].

Algunas decisiones pueden ser de rutina o intrascendentes. Como se observa la Figura 2, el ciclo para realizar la toma de decisiones consiste en llevar a cabo siete fases, las cuales parten de la definición del problema hasta llegar a la fase de seguimiento y control [21].



**Figura 2.** Fases de la toma de decisiones [21].

### 1.1.1.2. Niveles de decisión

Los niveles de decisiones que se pueden tomar en una empresa se muestran en la Figura 3. Los tres niveles de decisiones dependen del nivel jerárquico de la empresa. Las más altas corresponden al nivel estratégico y las de más bajo nivel corresponden a decisiones de nivel operativo [22].



**Figura 3.** Fases de la toma de decisiones [21].

- **Nivel estratégico.** El nivel de estrategia se refiere al proceso estratégico en los servicios, su compromiso es responder a la duración de la clasificación, por lo tanto, sus decisiones poseen efectos a corto, mediano y largo plazo [4]. El nivel estratégico accede a establecer un tiempo explícito en las disposiciones de eficiencia, producción, rotación de inventarios, magnitud de ventas, desvíos de los fines estratégicos de un comercio, departamento o producción, entre otras [4].
- **Nivel táctico.** El nivel táctico se encarga de procesos de diseño de servicios que existen por integración funcional de la correspondencia y la combinación de las áreas de colocación. Este nivel se complementa con aquellos recursos que ayudan en la inteligencia de negocio (BI) a la toma de decisiones a corto o largo plazo [4]. En este nivel se brinda soporte a la combinación de acciones y administración de archivos determinados para proporcionar consultas sobre información contenida en el sistema, además de proveer de informes y facilitar el encargo independiente de la información por parte de los niveles intermedios de la distribución [23].
- **Nivel operativo.** Finalmente, el nivel operativo se encarga de comprender las técnicas importantes que afirman la actividad diaria de la colocación. El efecto de sus operaciones y las decisiones son visibles en el corto plazo. Este nivel interactúa con los usuarios y busca asegurar la actividad diaria de la colocación para proveer datos que ayudan a tomar mejores decisiones estratégicas [4].

### 1.1.2. Datawarehouse

Un datawarehouse es una agrupación de datos temáticos, integrados, variantes en el tiempo y no volátiles, que se localizan en la estructuración para sostener la obligación empresarial orientada a la toma de decisiones.

Para la creación de un datawarehouse se requieren procesos para separar datos de las intervenciones de manera periódica de la empresa, originada en diferentes intervalos, y subsistemas, y sumarlos y acumularlos en un almacén de datos (repositorio) de manera que se pueda acudir a ellos cada vez que el usuario lo requiera [24].

Como se mencionó antes, un datawarehouse puede ser considerado como un almacenamiento de datos que suministra un enfoque global, normal e integrado de los datos de la colocación, independiente de cómo se lleguen a utilizar después por los clientes o usuarios. Así, un datawarehouse debe cumplir con las siguientes características: estable, coherente, fiable e histórico.

Al contener un ámbito global de la colocación con un extenso alcance histórico, el tamaño de datos que debe gestionar un datawarehouse puede ser muy grande (centenas de Terabytes) [25], [26]. En la Figura 4, se puede observar un ejemplo de la estructura de un datawarehouse, el mismo que a través de técnicas de extracción, innovación y carga, integra los datos de las fuentes de información transaccionales que una organización emplea o gestiona dentro de su negocio [27], [28].

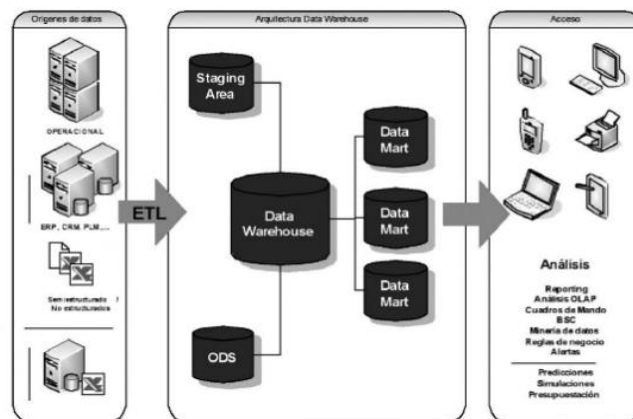


Figura 4. Esquema de datawarehouse [29].

### 1.1.2.1. Datamart

Un datamart, a diferencia de un datawarehouse que es un repositorio centralizado de toda la organización, es una base de datos que almacena datos de un ámbito específicos del negocio (i.e., ventas). En distintos puntos, un datamart se suministra a partir de los datos de un datawarehouse o integra por sí mismo una recopilación de diferentes orígenes de información [30], [31]. La forma como se integran los datamarts para constituir un datawarehouse se ilustra en la Figura 5 [32].

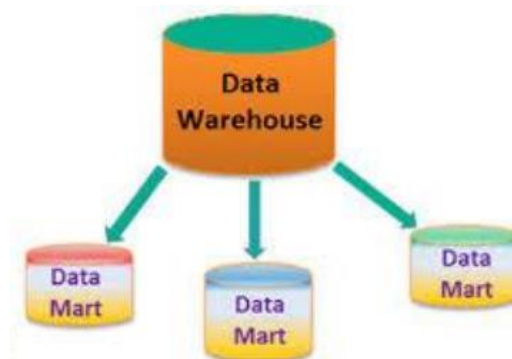
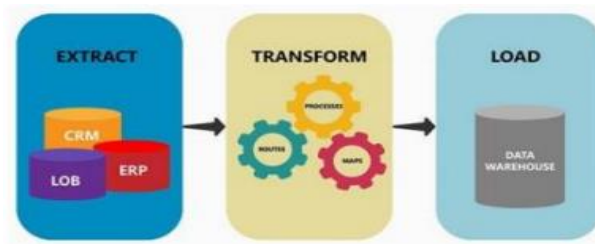


Figura 5. Esquema de datamart [33].

### 1.1.2.2. Procesos ETL

Los procesos ETL consisten en ejecutar tres funciones de datos, esto es, obtener datos de un origen los cuales son transformados aplicando normas, y son transformados para su estandarización y que de esta manera sean cargados en un datamart o un datawarehouse [34].

Como se observa en la Figura 6, los procesos ETL permiten organizar los datos de fuentes heterogéneas. En este sentido, los procesos ETL permiten llevar a cabo acciones enfocadas en trasladar datos de uno o infinitos principios; inicializar y depurar los datos en las ocasiones necesarias; atribuir los datos a bases de datos, ya sea en un datamart o un datawarehouse; los datos logran ser examinados una vez localizados o establecidos en la aplicación [34]. De igual manera, los procesos ETL permiten otras operaciones útiles en repositorios como son: fortalecer, migrar y sincronizar bases de datos efectivos; sincronizar en otras técnicas operacionales (i.e., ERP); migrar bases de datos entre versiones [35].



**Figura 6.** Arquitectura de los procesos de ETL [34].

### 1.1.2.3. Análisis OLAP

El desarrollo analítico en línea, acreditado como OLAP por sus iniciales en inglés (*On-Line Analytical Processing*), es conocido como el estudio pluridimensional de datos. El estudio pluridimensional de datos es un método analítico que facilita la observación de los datos en un modelo de cubo dimensional (o hiperdimensional), proporcionando la opción, análisis y toma de decisiones [36].

En la Figura 7, se muestra un cubo OLAP en que se cruzan tres variables en una organización. Las tres variables analizadas corresponden al producto, el año y la zona geográfica en que se ha distribuido o vendido un producto. En este sentido, este cubo OLAP permite responder a preguntas de negocio de la siguiente índole: ¿cómo los clientes han aceptado los productos que la organización produce de acuerdo con los últimos años de vida de la empresa y de acuerdo con la zona territorial en la que se encuentran ubicadas las tiendas?



**Figura 7.** Ejemplo de cubo OLAP [37].

De esta forma, los cubos OLAP suministran la inteligencia de simulaciones, estudio y observaciones de grandes grupos de datos accediendo a la manipulación, recuperación y composición de éstos por centro de consultas, documentación y cuadros de mando de fácil uso por parte del usuario [36].

#### 1.1.2.4. Arquitectura de flujo de datos

La arquitectura de flujo de datos está orientada en situar el proceso de análisis de estrategias con las distintas herramientas de BI [38], [39]. Además, para que el proceso de la información sea más entendible o manejable en la toma de decisiones, la obtención de la información constituye un elemento primordial y más aún cuando se trata de los recursos de inteligencia de negocios [40]. Posteriormente, se alcanza a observar en la Figura 8, el flujo a seguir para el diseño de un sistema de inteligencia de negocios.

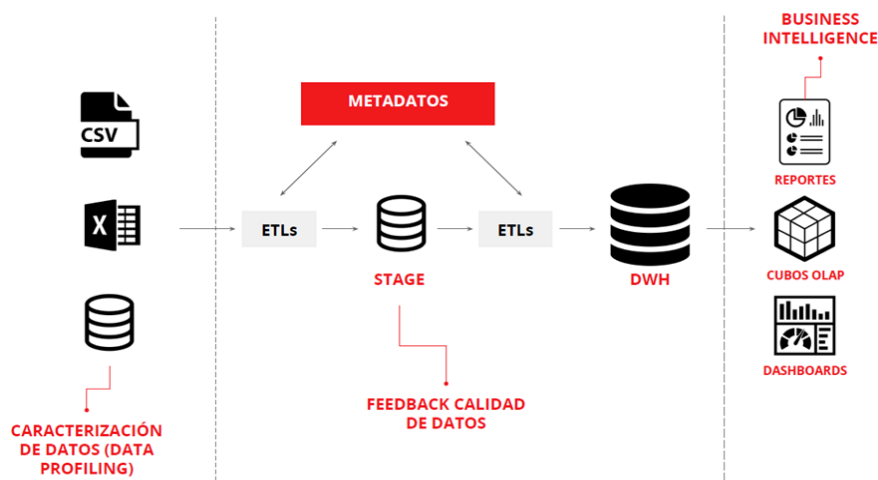


Figura 8. Arquitectura de flujo de datos [41].

- **Los datos detallados:** Habitualmente, los datos extensos pueden ser muy voluminosos y exigir un apartado artificial para administrar y negociar [23]. Es fundamental previo a procesar los datos ejecutar un proceso de caracterización de datos para analizar la lógica del negocio y sobre todo analizar posibles anomalías en los propios datos almacenados.
- **Datawarehouse.** Mediante el uso de proceso ETL es necesario normalizar los datos de las fuentes heterogéneas para almacenarlas en un repositorio, datawarehouse, y a partir de ellos, realizar un análisis de los datos. Previo a almacenar los datos en el datawarehouse se recomienda pasar por un repositorio *stage* y en él aplicar mecanismos ETL que permitan eliminar cualquier anomalía que persista de las técnicas ETL aplicadas en las fuentes de datos heterogéneas.
- **Inteligencia de negocio:** Forman un resultado de estudio y una síntesis de la información argumentado en el método de decisión [23]. Para resumir los datos en un método de inteligencia de negocios emplean análisis OLAP, y a partir de dichos

resultados los datos son presentados mediante el uso de reportes, cuadros de mando, entro otros [23].

#### 1.1.2.5. Metodologías de desarrollo de sistema de BI

Las metodologías constituyen el proceso para diseñar el sistema de inteligencia de negocios; esto es, construcción del repositorio y el diseño de mecanismos de análisis de datos. Hacia el crecimiento de sistema de inteligencia de negocios, se han manifestado dos enfoques principales, estos son: el paradigma de Kimball y el paradigma de Inmon [42]. De manera general, a continuación, se describen ambas metodologías.

- **Metodología de Ralph Kimball.** Ralph Kimball es considerado el inventor del modelo dimensional y pionero en datawarehouse e inteligencia de negocios. La metodología de Kimball se fundamenta en el periodo de vida de la magnitud del negocio y se centra en construir una infraestructura de información centralizada (almacén de datos, datawarehouse o DW). Como se muestra en la Figura 9, Kimball propone aplicar procesos ETL de las fuentes de datos con la finalidad de nutrir el repositorio centralizado (datawarehouse), y a partir de los datos almacenados en dicho repositorio, se aplican los mecanismos de analítica de datos (i.e., OLAP). Éstos últimos permiten resumir los datos y presentan información ventajosa para la toma de decisiones a distintos niveles organizacionales; pero principalmente a nivel estratégico [43].

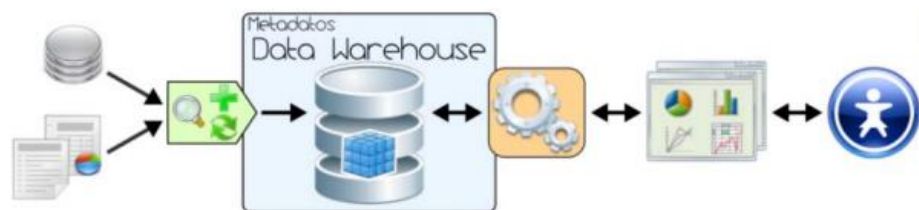
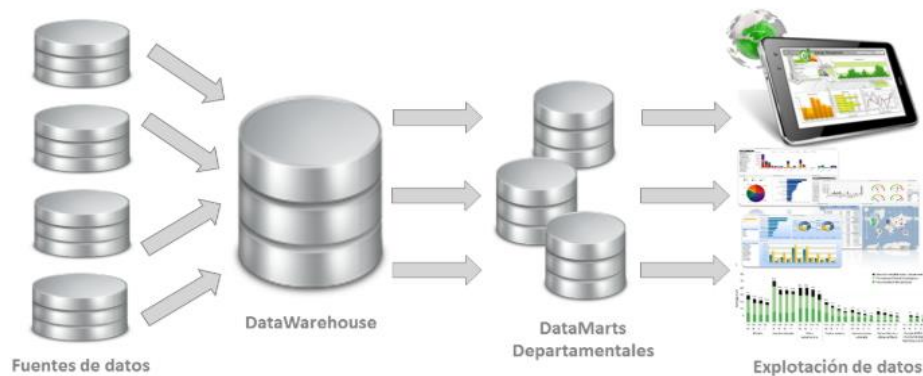


Figura 9. Metodología Ralph Kimball [44].

- **Metodología de Bill Inmon.** Es una metodología *top-down* que plantea desarrollar un almacén de datos empezando con la arquitectura del modelo de retención de datos empresarial, que determina la importancia de las áreas temáticas y la colectividad con las que opera la empresa [45]. Se plantea entonces, un almacenamiento que agrupa los datos de los distintos métodos funcionales de las estructuras para que éstos

queden validados e incorporados en la sola base de datos [46]. La lógica de esta metodología se ilustra en la Figura 10 [47].



**Figura 10.** Arquitectura Bill Inmon [47].

A diferencia de Kimball, todas las operaciones de analítica de datos, reportes y resúmenes de la información se las realiza desde una parte del datawarehouse, es decir, desde los datamarts que contienen información específica departamental.

### **1.1.3. Contaminación Medioambiental**

La contaminación medioambiental causa cambios en el medio ambiente. Ésta puede perjudicar de modo leve o grave al entorno o a su vez arruinarlo por completo. Además, se debe tener en cuenta el factor tiempo, pues cuanto más continuo es el daño puede afectar elementos vitales para la existencia de los seres vivos en la Tierra, sea en el suelo, agua o aire [48].

#### **1.1.3.1. Recursos ambientales**

Los recursos ambientales son activos sobre los cuales los individuos practican diversas demandas para conseguir un nivel de bienestar, teniendo en cuenta que existen limitaciones y términos. Además, la gestión de los recursos naturales analiza cuál debe ser la mejor conducta de los agentes sociales, en sentido amplio, respecto al uso, explotación y conservación [49]. Los recursos ambientales se dividen en recursos renovables y no renovables. A continuación, se describen cómo los recursos como el aire, agua y suelo son recursos ambientales sujetos a investigación, debido a que generalmente se encuentran contaminados.

- **Contaminación del aire:** Se produce por toda sustancia no esperada que llega a la atmósfera, la principal complicación es la colectividad actual y varias de sus innovaciones industriales. Además, la contaminación del aire es generada en la ciudad moderna, emanando diferentes gases y materia de partículas o masas minúsculas que pueden ser perjudiciales para la salud humana y el ambiente [50], [51].
- **Contaminación del agua:** Los recursos naturales como el agua son indispensables para la vida, es una obligación fundamental para la salud, y como tal se debe entender que es uno de los derechos humanos básicos. El agua se ha transformado en un bien muy valioso necesario, es un apoyo a la vida y además el progreso financiero está subordinado a la reservación del agua [52].
- **Contaminación del suelo:** Es el resultado de la actividad humana, que contiene empresas industriales de polución como las papelerías, térmicas, cremación y vertido de desechos industriales y urbanos, además de las antiguas zonas mineras activas, el uso de plaguicidas o tráfico de fertilizantes y herbicidas [53], [54].

### 1.1.3.2. Contaminantes

Los contaminantes químicos afectan directamente a la naturaleza y particularmente a bienes naturales como el aire, agua y suelo que son los más perjudicados por el exceso de contaminación. Con el propósito de controlar los niveles de contaminación de estos elementos, importantes para la vida, se han desarrollado normativas de regulación en las que se determinan los máximos permisibles de los contaminantes. Sobre estos máximos permisibles es posible desarrollar normas de regulación. En la Tabla 1,

Tabla 2 y Tabla 3 se muestran respectivamente los principales contaminantes del aire, agua y suelo. Estos contaminantes son cruciales para determinar la calidad del aire, agua y suelo, por lo que fueron considerados como datos importantes para la implementación del sistema de inteligencia de negocios.

**Tabla 1.** Materiales de máximo permisible del aire [50].

Contaminantes y Periodo de tiempo	Unidad	Alerta	Alarma	Emergencia
Monóxido de carbono	ug/m <sup>2</sup>	15000	30000	40000

Ozono	ug/m <sup>2</sup>	200	400	600
Dióxido de nitrógeno	ug/m <sup>2</sup>	1000	2000	3000
Dióxido de azufre	ug/m <sup>2</sup>	200	1000	1800
Material particulado PM 10	ug/m <sup>2</sup>	250	400	500
Material particulado PM 2,5	ug/m <sup>2</sup>	150	250	350

**Tabla 2.** Principales contaminantes del agua [55].

<b>Parámetro</b>	<b>Expresado como</b>	<b>Unidad</b>	<b>Máximo Permissible</b>
Aceites y Grasas	Sustancias solubles en hexano	mg/l	0,3
Aluminio	Al	mg/l	0,2
Amoniac	N-Amoniacal	mg/l	1,0
Amonio	NH <sub>4</sub>	mg/l	0,05
Arsénico (total)	As	mg/l	0,05
Bario	Ba	mg/l	1,0
Cadmio	Cd	mg/l	0,01
Cianuro (total)	CN <sup>-</sup>	mg/l	0,1
Cloruro	Cl	mg/l	250
Cobre	Cu	mg/l	1,0
Coliformes Totales	nmp/100 ml		3000
Coliformes Fecales	nmp/100 ml		600
Color	color real	unidades de color	100
Compuestos fenólicos	Fenol	mg/l	0,002
Cromo hexavalente	Cr <sup>+6</sup>	mg/l	0,05
Demanda Bioquímica de Oxígeno (5 días)	DBO <sub>5</sub>	mg/l	2,0
Dureza	CaCO <sub>3</sub>	mg/l	500

**Tabla 3.** Materiales de máximo permissible del suelo [56].

<b>Parámetros</b>	<b>Unidades (Concentración en peso seco de suelo)</b>	<b>Valor</b>
Conductividad	dS/m	2

pH		6 a 8
Relación de Adsorción de Sodio (Índice SAR)		4
Arsénico (inorgánico)	mg/kg	5
Azufre (elemental)	mg/kg	250
Bario	mg/kg	200
Boro(soluble en agua caliente)	mg/kg	1
Cadmio	mg/kg	0,5
Cobre	mg/kg	30

Es importante señalar que cada país tiene su propia normativa de regulación y consecuentemente, sus máximos permisibles, por lo cual se recalca que los datos considerados corresponden a la actual normativa vigente en el Ecuador.

### **1.1.3.3. Normativas de gestión**

La normativa establecida en el Ecuador con los máximos permisibles de contaminación contempla los bienes como aire, agua y suelo. Las normativas referentes al cuidado del aire son las técnicas que se rigen en la Ley de gestión Ambiental, las cuales buscan prevenir y controlar la contaminación ambiental estableciendo las siguientes normas: la importancia del aire ambiente, los límites permisibles de los contaminantes críticos y contaminantes no convencionales del aire ambiente, los sistemas y procedimientos para especificar los contaminantes en el aire ambiente [50], [55].

Las normativas del agua se basan en la Ley de Gestión Ambiental y del Reglamento a la Ley de Gestión Ambiental para la Prevención y Control de la Contaminación Ambiental. Esta normativa se enfoca en preservar la importancia del recurso del agua para proteger y salvar la integridad de la persona, de los ambientes y sus correlaciones, y del ambiente en general. Por ello, se contempla esta normativa técnica que también establece aspectos como: los términos permisibles, preparaciones y prohibiciones para las descargas en cuerpos de aguas y ordenanza de alcantarillado; las posiciones de la naturaleza de las aguas para sus distintos manejos; procedimientos y técnicas para establecer el aspecto de contaminantes en el agua [57], con los elementos de máximo límite predecible.

De igual manera, la normativa del suelo se basa en las leyes del Ministerio del Ambiente y la Ley de Gestión Ambiental, la cual regula a las instituciones tanto comerciales como industriales el desarrollo de acciones que poseen el potencial de alterar los recursos del suelo [58].

#### 1.1.4. Suite Pentaho

La suite de Pentaho incluye herramientas de inteligencia de negocios, reportes, estudios, cuadros de mando, minería de datos e integración de datos, que brindan una secuencia de servicios precisos entre los que están la acreditación, cronogramas de tareas, confianza y servicios Web [59]. También es la plataforma general de inteligencia de negocios líder de recursos BI de acceso libre, centrada en técnicas que logran ser cómodas y personalizadas [60].

La arquitectura de Pentaho implementa los componentes de BI tal como se puede observar en la Figura 11 [61].

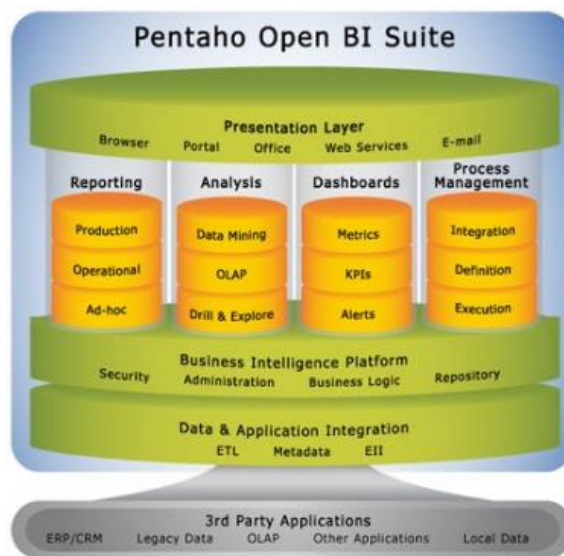


Figura 11. Arquitectura de Pentaho BI Suite [62].

- **Pentaho Business Analytics.** Permite al usuario de negocios acceder de forma intuitiva, explorar y analizar todos los datos, lo que les permite hacer decisiones basadas en información que positivamente puede impactar el desempeño de sus organizaciones. La colección de componentes de análisis en *Pentaho Business Analytics* permite visualizaciones de tendencias de datos creando informes

estáticos de una fuente de datos de análisis, atravesando un cubo de análisis, mostrando los puntos de datos se comparan mediante el uso de gráficos, y monitorear el estado de ciertas tendencias y umbrales con cuadros de mando [63].

- **Pentaho Report Designer.** La herramienta *Pentaho Report Designer* es un instrumento especializado en el diseño de reporte y se conecta a la base de datos recopilando los datos necesarios para compilar el informe utilizando consultas en lenguaje SQL [64]. Pentaho proporciona una herramienta de generación de informes (Report Designer), la cual puede crear informes de análisis profesionales de acuerdo con las necesidades del usuario y admite exportar a formatos como Excel, PDF y otros. Además, permite editar el informe con el diseño pudiendo establecer los parámetros de entrada, a través de los parámetros en los datos de informes del filtro de origen puede pasar condiciones de consultas interactivas [31].
- **Pentaho Data Integration.** Pentaho Data Integration (PDI) o Kettle ETL es una herramienta de código abierto enfocada en integrar los datos en el archivo de datos para luego presentarlo con precisión y puntualidad. Los procesos de integración en inteligencia de negocio son bien conocidos como un proceso Extraer Transformación Cargando (ETL, por sus siglas en inglés), además el proceso ETL cambia los datos del procesamiento transaccional en línea [65]. Esta herramienta tiene una larga trayectoria, solidez y robustez que la convierten en una herramienta muy recomendada, permitiendo la creación de transformaciones que automatizan los procesos de estandarización de datos. Funciona de una manera muy sencilla e intuitiva [66]. El diseño de Pentaho Data Integración es similar a la arquitectura [67].
- **Pentaho Data Mining.** Maneja algoritmos complejos para el análisis de datos a través de procesos que permiten descubrir datos valiosos para la toma de decisiones [63]. Lo que la hace tan atractiva a muchos usuarios comerciales y lo distingue de los procesos de BI regulares, ya que cuenta con herramientas para informar y analizar el desempeño pasado y comparar los objetivos con medidas reales. Tiene la capacidad de hacer predicciones sobre el desempeño futuro dentro de ciertos límites de confianza [68].

## 1.2. Antecedentes

Para identificar los antecedentes de la investigación se utilizaron artículos recuperados de bibliotecas digitales y bases de datos científicas de alto impacto, las cuales son reconocidas por parte de la comunidad académica y que se encuentran indexadas en: Scopus, IEEE Xplore y ACM. La indagación de los estudios se realizó mediante la aplicación de un protocolo de búsqueda científica la cual estuvo fundamentada en el uso de una cadena de búsqueda especializada que incluyó los siguientes términos: ("*business intelligent*" OR "*bi*") and ("*pollution*" OR "*environment*") and ("*monitoring*"). Dicha cadena fue aplicada en las bases de datos documentales y bibliotecas digitales ya mencionadas, filtrando estudios publicados entre 2015 y 2020. Luego de analizar los estudios recuperados se seleccionaron seis artículos donde se utiliza la técnica de inteligencia de negocios para el estudio de contaminación del agua, suelo y aire; además de identificar los análisis ambientales con los recursos ya mencionados.

La primera investigación propuesta por Toma et al. [69] "*IoT Solution for Smart Cities' Pollution Monitoring and the Security Challenges*" plantea que mitigar los riesgos de contaminación del aire es un factor importante en el calentamiento global que amenazan directamente la salud, por lo que se debe reflexionar sobre el riesgo de la contaminación del aire [69]. Se propone un sistema de vigilancia de contaminación en tiempo real, incluidos los protocolos de comunicación de Internet de las cosas (IoT). En el estudio se realizó la adquisición y transmisión de datos a través de canales de comunicación, al igual que se gestionó la seguridad de los datos.

La segunda investigación propuesta por Desai et al. [70] "*IoT based air pollution monitoring and predictor system on Beagle Bone Black*", planteó medir el nivel de dióxido de carbono y monóxido de carbono existente en el aire, junto con la ubicación del sistema de posicionamiento global (GPS) mediante el uso de sensores de detección y cargas en los servicios en la nube de Azure. También integrado junto con los sensores de gas que se utilizan para la adquisición de datos [70]. Los datos se representan mediante la herramienta Power BI, los datos del sensor de gas calibrado se obtienen de los sensores y son cargados con éxito en la nube. Se utilizan los datos almacenados en la nube por diferentes servicios para que los datos sean significativos. El sistema propuesto está

implementado y es útil para monitorear y reducir la contaminación en una ciudad inteligente, evitando así la contaminación.

La tercera investigación propuesta por Khot et al. [71] “*Survey on Air Pollution Monitoring Systems*” propone un método de vigilancia del aire en tiempo real que requiere características como la medición exacta de los parámetros y el análisis de los mismos para la toma de decisiones de manera oportuna [71]. Se puede ver la representación en tiempo real del escenario actual. A través de este tipo de aplicaciones lo que permite realizar evaluaciones de impacto en la salud.

La cuarta investigación propuesta por Haghparast et al. [72] “*Comprehensive Environmental Monitoring based on Stations of Environmental Pollutants (Air, Water and Soil) in Tehran*”, se basa en estaciones de monitoreo de contaminantes ambientales (aire, agua y suelo) y control del contenido de dichos contaminantes en Teherán. Dada la naturaleza de los factores y elementos investigados en este estudio, el software ArcGis fue aplicado para el análisis de datos y suministros de mapas digitales básicos [72]. Se manejó el modelo de lógica difusa para el análisis de los datos con el método de comparación por pares entre el aire, agua y suelo, en función de la importancia y la preferencia.

La quinta investigación propuesta por Crisóstomo [73] “*Implementación De Power Bi Para El Análisis De Información En La Productividad En El Laboratorio Clínico Del Hospital Central De La Fuerza Aérea Del Perú De Lima-2017*”, desarrollado en el laboratorio del Hospital Central de la FAP se fortalece a partir de datos de los pacientes, exámenes, servicios, analizadores, procesamiento de las muestras, etc. Se implementa el Power BI para el análisis de información adquirida [73].

La sexta investigación propuesta por Baralis et al. [74] “*Analyzing air pollution on the urban environment*”, propone el análisis de datos basado en inteligencia empresarial, metodologías y tecnologías abiertas, para soportar diferentes análisis específicos de datos sobre contaminación del aire [74]. Para analizar el problema de diferentes facetas, mediciones de contaminación del aire se enriquecen con información adicional como datos meteorológicos y de tráfico que se recopilan a través de redes de sensores disponibles en el contexto de la ciudad inteligente. Este conjunto de datos integrados se analiza periódicamente para generar paneles informativos basados en una selección de

claves indicadores de desempeño (KPI) y tableros que proporcionan información útil sobre el alcance de los contaminantes.

### **1.3. Fundamentos legales**

Dentro de los fundamentos legales que garantizan la fiabilidad y viabilidad respecto al desarrollo de este proyecto de investigación; es necesario enfatizar en la legislación vigente del Ecuador así como: la Ley Orgánica de Salud [75], Código Orgánico Ambiental [76], Constitución de la República del Ecuador [77], y la Ley Orgánica de Educación Superior [78].

De acuerdo con la Ley Orgánica de Salud en el Capítulo tercer del derecho a la salud y su protección [75]; que en el Art.95, especifica que “El sector privado y público está obligado a proporcionar a la población, información adecuada y veraz respecto del impacto ambiental y sus consecuencias para la salud individual y colectiva [75]”. El presente estudio propone proveer de una herramienta que permita a los ciudadanos conocer los niveles de contaminación ambiental, además de tener como finalidad que las autoridades tomen decisiones asertivas en relación con los niveles de contaminación, de tal manera que se pueda impedir la contaminación de las propiedades de los recursos como el aire, agua y suelo.

De acuerdo con el Código Orgánico Ambiental del Título III de protección frente a riesgos y amenazas ambientales para la salud, Sección primera de calidad del aire y contaminación acústica y visual [76]; que el Art. 91, determina que se debe “Mantener la calidad del aire y evitar, controlar y mitigar la contaminación del mismo y la contaminación acústica y visual que afecta a la salud individual y colectiva [76]”. En este sentido la investigación ayudó a resumir los datos a través del análisis OLAP, permitirá a los gestores de la ciudad monitorear en cada momento los niveles de contaminación de los recursos como el aire, agua y suelo, y a la aplicación de medidas de control en caso de ser pertinente.

Además, del convenio con la Ley Orgánica de Educación Superior en el Título I de Ámbito, objeto, fines y principios del sistema de educación superior; Capítulo 2 Fines de la educación superior [78]; que en el Art. 8, estipula que se debe “Fomentar y ejecutar programas de investigación de carácter científico, tecnológico y pedagógico que

coadyuven al mejoramiento y protección del ambiente [78]”. LA investigación tiene el propósito de facilitar una alternativa de apoyo a la resolución de la problemática, fomentada por el impacto medio ambiental y los factores que causan un deterioro en la salud del ser humano. Se pretendió que el método planteado sirva para la toma de decisiones acertadas que favorezcan el buen vivir de las personas. No obstante, es importante señalar que el sistema no controla, tan solo proporciona información para que las autoridades tomen decisiones más acertadas basadas en datos.

## **CAPÍTULO II: METODOLOGÍA**

### **2.1. Delimitación del estudio**

La investigación se desarrolló en la provincia de Esmeraldas puesto que en esta ciudad se asientan varias empresas que afectan a la contaminación medioambiental (i.e., empresas industriales, empresas palmicultores, empresas mineras, entre otras) y consecuentemente, afectan a los bienes ambientales como el aire, agua y suelo, por causa de sus actividades diarias. Se intentó buscar fuentes de datos de la ciudad de Esmeraldas, sin embargo, al no haber podido contar con ellos, se usaron datos del contexto nacional.

Asimismo, la investigación se efectuó durante el segundo semestre del presente año 2022.

### **2.2. Tipos de estudio**

El tipo de análisis es mixto, es decir, es una investigación cuantitativa y cualitativa [79]. Es de tipo cualitativo porque permitió determinar las características de estos contaminantes acorde a la norma ambiental del Ecuador y específicamente de la ciudad de Esmeraldas. Esto con la finalidad de que el sistema se adapte al contexto de dicha ciudad.

Por otro lado, la investigación también posee un enfoque cuantitativo debido a que se tomaron valores de los niveles de contaminación ambiental de los contaminantes que se relacionan con la polución del aire, agua y suelo, para en base a ellos evaluar el nivel de contaminación que tiene la ciudad de Esmeraldas en cada momento [80]. No fue un requisito contar con datos reales como se ha comentado anteriormente, se planteó que la herramienta a desarrollar se adapte a futuro a los datos cuando la ciudad de Esmeraldas; estos no estaban disponibles a la fecha en la que se ejecutó el proyecto.

Finalmente, la investigación es de tipo experimental debido al uso de las herramientas de BI para la toma de decisiones en base a los valores conseguidos del set de datos de los niveles de contaminación en los diferentes recursos como el aire, agua y suelo, con el fin

de saber las causas y efectos del fenómeno que impacta en la contaminación medioambiental.

### **2.3. Métodos de investigación**

Los métodos que se aplicaron en el plan de investigación fueron el método deductivo, inductivo y experimental.

El método inductivo, parte de asuntos individuales para alcanzar a una propuesta general, se empleó para establecer las conclusiones del estudio y llevar a cabo el proceso de toma de decisiones en base a los datos usados por la aplicación de inteligencia de negocio.

En relación con el método deductivo este proporciona las generalizaciones, de forma lenta se empleó en los casos particulares con el propósito de llevar los conceptos y buenas prácticas de los sistemas de inteligencia de negocio, generalmente aplicados al ámbito empresarial, al campo ambiental, y específicamente a la gestión de la polución del aire, agua y suelo.

Finalmente, se empleó el método experimental con el fin de llevar a cabo cada una de las etapas del período del diseño y progreso del sistema de inteligencia de negocios en el campo medioambiental. Cada etapa del flujo de datos que sigue un sistema de inteligencia de negocios, implementada con la suite Pentaho, fue probado, de tal modo que se evidenció que los datos del datawarehouse son consistentes e íntegros. Además, en base a los datos se pudo validar cada uno de los mecanismos de análisis de datos OLAP, de manera que, en el contexto real de Esmeraldas se pueda proyectar a futuro cuando se cuente con redes de sensores que capturen datos reales se permitan tomar decisiones contextualizadas y acertadas.

### **2.4. Operacionalización de variables**

La investigación planteada buscó investigar tres variables fundamentales, la infraestructura de inteligencia de negocios que se requiere para el sistema de observación de la contaminación ambiental, así como, la variable fuente de datos que permitió conocer los datos con los que se debe disponer para la aplicación de los procesos ETL y almacenarlos en el datawarehouse. Esto es importante debido a que son datos técnicos y

siendo importante investigar su naturaleza y su comportamiento en relación con las propiedades del aire, agua y suelo.

Por consiguiente, como se muestra en la Tabla 4, se planteó también investigar la variable del método de inteligencia de negocios con la finalidad de determinar los requerimientos con relación a los datos, análisis OLAP (formulación de preguntas de negocio), reportes, y cuadros de mando. Estas variables, así como sus indicadores se listan en la tabla antes mencionada.

**Tabla 4.** Variables e indicadores sujetos a estudio.

<b>Variable</b>	<b>Dimensión</b>	<b>Indicador</b>	<b>Tipo variable</b>
<b>Infraestructura de BI</b>	Tecnología hardware	Servidor de integración de datos (ETL)	Cuantitativa y cualitativa
		Servidor de base de datos (datawarehouse)	Cuantitativa y cualitativa
		Servidor de BI (despliegues de OLAP)	Cuantitativa y cualitativa
	Tecnología software	Herramienta de caracterización de datos	Cuantitativa y cualitativa
		Herramienta de integración de datos	Cuantitativa y cualitativa
		Herramienta de análisis OLAP	Cuantitativa y cualitativa
		Herramienta de reporte	Cuantitativa y cualitativa
		Herramienta de cuadros de mando	Cuantitativa y cualitativa
<b>Fuentes de datos</b>	Datos	Fuentes de datos	Cualitativo
		Set de datos	Cualitativo
		Formatos	Cualitativo
		Metadatos	Cualitativo
		Integridad de datos	Cualitativo
	Tipos	Contaminantes del aire	Cuantitativa y cualitativa
		Contaminantes del agua	Cuantitativa y cualitativa
		Contaminantes del suelo	Cuantitativa y cualitativa
	Máximos permisibles en hogares	Contaminación del aire	Cuantitativo
		Contaminación del agua	Cuantitativo
Contaminación el suelo		Cuantitativo	
<b>Sistemas de BI</b>	Módulo de cuadros de mando	Interfaz gráfica de usuario	Cuantitativo
		Efectividad	Cuantitativo

A continuación, se describen los indicadores mencionados en la Tabla 1 de manera que se pueda entender de una forma más clara sus implicaciones y en algunos casos su alcance:

- **Infraestructura de BI:** es la infraestructura de hardware que sirve para efectuar el método de inteligencia de negocios.

- Servidor de integración de datos (ETL): es el servidor donde se ejecutarán los procesos ETL para estandarizar los datos en el datawarehouse.
  - Servidor de base de datos (datawarehouse): es el servidor para realizar el almacenamiento del datawarehouse que mantiene grandes cantidades de datos de contaminación del aire, agua y suelo.
  - Servidor de BI (despliegues de OLAP): es el servidor donde se ejecutaron los análisis OLAP a partir de los datos del datawarehouse.
  - Herramienta de caracterización de datos: Son requisitos tecnológicos de la herramienta que permite acceder, interpretar y examinar la integridad de la información, así como las anomalías de las fuentes de datos útiles para nutrir el datawarehouse.
  - Herramienta de integración de datos: Requisitos tecnológicos de la herramienta para diseñar y ejecutar los procesos ETL que cargan los datos desde las fuentes al datawarehouse.
  - Herramienta de análisis OLAP: Requisitos tecnológicos de la herramienta especializados en analizar grandes cantidades de datos usando cubos OLAP.
  - Herramienta de reporte: Requisitos tecnológicos de la herramienta para resumir la información usando reportes.
  - Herramienta de cuadros de mando: Requisitos tecnológicos de la herramienta que permiten llevar a cabo el diseño y despliegue de los cuadros de control de sistema a partir de los cubos OLAP y el datawarehouse.
- **Fuente de datos:** son los datos que sirven al sistema para nutrir al datawarehouse.
    - Fuentes de datos: Corresponden a las fuentes u orígenes de datos que pueden estar almacenados como parte de sistemas transaccionales, repositorios públicos, entre otros.
    - Set de datos: Son los datos de contaminación del aire, agua y suelo propiamente dicho.
    - Formato: Extensión y esquema de los sets de datos a emplear en los procesos ETL.
    - Metadatos: Son datos acerca de los cuales se suministra información a los sets de datos a emplear en el sistema.
    - Contaminantes del aire: Datos químicos de estado sólido y gases en el aire que son emitidos por autos y empresas industriales.

- Contaminantes del agua: Datos químicos, físicos, entre otros, que son dañinos para el agua de consumo y de uso agrícola.
- Contaminantes de suelo: Datos químicos de las sustancias que son arrojadas en el suelo afectando su fertilidad.
- **Sistemas de BI:** Es el uso de estrategias y herramientas que sirven para la transformación de la información en conocimiento.
  - Interfaz de usuario: es el medio donde el usuario puede interactuar con una máquina.
  - Efectividad: es la capacidad de realizar diferentes funciones para el cumplimiento de los objetivos para el que fue creado un componente o interfaz.

## 2.5. Población y muestra

En el transcurso de la indagación se realizó la búsqueda de un dataset que contenga datos generados por estaciones de monitoreo ambiental de los recursos tales como: aire, agua y suelo. Como datos sujetos a análisis se considerarán los contaminantes de los tres recursos antes especificados, descritos en la Tabla 1,

Tabla 2 y Tabla 3.

## 2.6. Técnicas e instrumentos de recolección de datos

Las técnicas de investigación que se aplicaron en el proyecto fueron principalmente la entrevista y el diseño experimental. A continuación, se describe como se aplicaron y la herramienta de recolección de datos que se ha diseñado para su aplicación [81].

- **Entrevista.** La técnica de la entrevista se aplicó en el transcurso del año 2022, durante el mes de septiembre, con la finalidad de recolectar los datos sobre los químicos que afectan al aire, agua y suelo de la ciudad de Esmeraldas. Se contó con el apoyo de un especialista en Gestión Ambiental, de manera que aporte información sobre las diferentes fuentes de contaminación, las formas de monitoreo y en la toma de decisiones en dicha área. Adicionalmente, se buscó determinar las principales preguntas de negocios que el sistema de inteligencia de negocios debe responder, para de esta manera garantizar la toma de decisiones

acertadas. El instrumento utilizado para la realización de la entrevista fue un cuestionario compuesto de 5 preguntas abiertas. Para obtener información detallada acerca del cuestionario se puede observar el

- **Anexo 2.**
- **Diseño experimental.** El diseño experimental se aplicó en el transcurso del año 2022, en el mes de febrero, con el propósito de poner en marcha la Suite Pentaho para el progreso del sistema de inteligencia de negocios propuesto [59]. Se diseñó un cubo OLAP para responder a cada una de las preguntas de negocios planteadas (posteriormente especificadas) y realizar el despliegue de cuadros de mando [82]. Además, se verificó la calidad de los datos generados por los cubos OLAP, que son cruciales para la toma de buenas decisiones. El experimento se realizó con el apoyo de una ficha que se puede observar en el **Anexo 1**.

## **2.7. Técnicas de procesamiento y análisis de datos**

El método de procesamiento de datos que se utilizó en la investigación fue la estadística descriptiva. Los cubos OLAP emplean funciones de agregado sobre los datos del datawarehouse tales como: suma, promedio, mínimo, máximo y cuenta. Todas ellas, estudiadas en la Estadística Descriptiva y usadas para resumir datos.

También está implícita la ejecución de consultas que incluyen grandes cantidades de datos que son procesados mediante la utilización de estructuras multidimensional, SQL y cubos OLAP.

## **2.8. Normas éticas**

El proyecto se desarrolló bajo el reglamento de grado de la Pontificia Universidad Católica del Ecuador Sede Esmeraldas (PUCESE) [83]. De igual manera, se planteó respetar el derecho de autor de la información obtenida como reportes, análisis y tipo de toma de decisiones. La investigación es totalmente inédita y los datos publicados de los niveles de contaminación medioambiental fueron usados con fines académicos si se llegara a encontrar datos reales que puedan causar impacto negativo a determinadas empresas.

## CAPITULO III: RESULTADOS

Para obtener los resultados se diseñó una base de datos y se cargaron los datos a partir de múltiples conjuntos de datos de calidad ambiental ya que no existe una base de datos pública de contaminación en la ciudad de Esmeraldas. A continuación, se procedió a crear los procesos ETL y se continuó con la fase de creación y diseño de la base de datos *stage* y datawarehouse. A partir del datawarehouse se plantearon los cubos OLAP y análisis multidimensionales que habilitaron las tareas de inteligencia de negocios. Todos estos aspectos siguen el esquema de la arquitectura de flujo de datos del capítulo Marco Teórico.

### 3.1. Análisis y preguntas de negocio

La formulación de las preguntas de negocio se realizó en base a la base de datos “DbContaminate”. Se plantearon 11 preguntas a las que el sistema de BI debe dar respuesta. Estas preguntas se especifican en la **Tabla 5**.

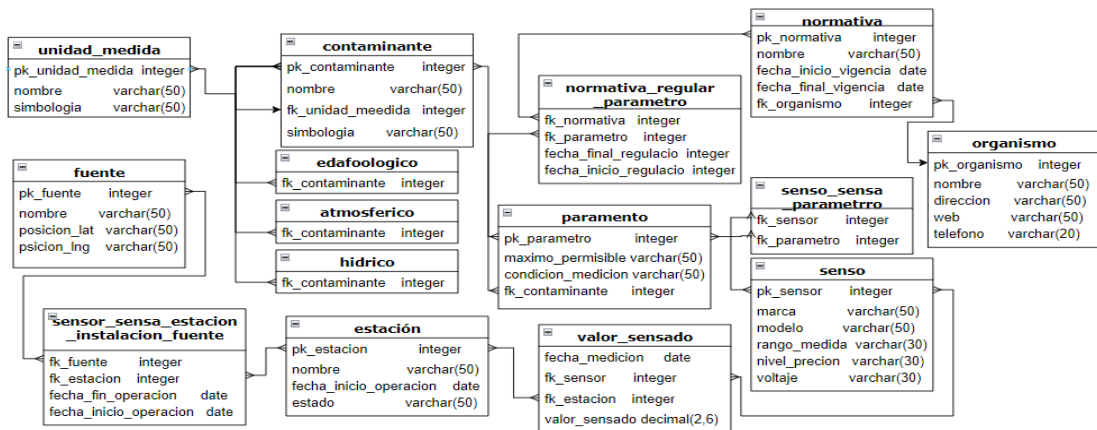
**Tabla 5.** Formulación de las preguntas de en BI

#	Preguntas de negocio
1.	¿En qué estación se han reflejado el menor índice de contaminantes (hídrico, edafológico, atmosférico) por (día, semana, mes, trimestre, año)?
2.	¿En qué estación se han reflejado el mayor índice de contaminantes (hídrico, edafológico, atmosférico) por (día, semana, mes, trimestre, año)?
3.	¿Cuál es la fuente de mayor de contaminantes (hídrico, edafológico, atmosférico) por (día, semana, mes, trimestre, año)?
4.	¿Cuál es la fuente de menor de contaminantes (hídrico, edafológico, atmosférico) por (día, semana, mes, trimestre, año)?
5.	¿Cuál es el contaminante (hídrico, edafológico, atmosférico) de menor índice por (día, semana, mes, trimestre, año)?
6.	¿Cuál es el contaminante (hídrico, edafológico, atmosférico) de mayor índice por (día, semana, mes, trimestre, año)?
7.	¿En qué fecha se han alcanzado valores superiores al máximo permisible de contaminantes (hídrico, edafológico, atmosférico)?
8.	¿En qué fecha se han alcanzado valores óptimos (calidad) al máximo permisible de contaminantes (hídrico, edafológico, atmosférico)?
9.	¿Cuál es el promedio de contaminante atmosférico (Co2, O3, So2, No2, PM10, PM2.5) por (día, semana, mes, trimestre, año)?
10.	¿Cuál es el promedio de contaminante edafológico (PH, DS, As, S, Bo, Br, Col, Cu) por (día, semana, mes, trimestre, año)?
11.	¿Cuál es el promedio de contaminante hídrico (Al, N-Amoniacal, NH4, As, Ba, Cd, Cn, Cl, Cu, Fenol, Cr+6, DBO5, CaCO3) por (día, semana, mes, trimestre, año)?

## 3.2. Arquitectura de flujo de datos del sistema BI

### 3.2.1. Fuentes de datos

Los datos obtenidos de fuentes externas fueron normalizados y almacenados en una base de datos. El diseño de dicha base de datos (modelo físico) se describe en la **Figura 12**. En resumen, las tablas que la base de datos integra corresponden a las necesarias para guardar los datos de los contaminantes del aire, agua y suelo, y aspectos relacionados como normativas y organismos reguladores. Todo se resume en las 15 tablas relacionadas.



**Figura 12.** Base de datos DbContaminante.

### 3.2.2. Base de datos STAGE

Siguiendo el flujo de datos de la arquitectura del sistema de BI, a continuación, en la **Figura 13**, se muestra el diseño del modelo de la base de datos STAGE. Esta base de datos es un primer paso para normalizar los datos previo a su almacenamiento en el repositorio (datawarehouse). Cualquier tipo de error o inconsistencia se puede corregir y que los datos del datawarehouse almacene datos de calidad.

### 3.2.3. Datawarehouse

El modelo de datos del datawarehouse se observa en la **Figura 14**. Este modelo sigue las bases del modelo propuesto por Kimball y es la base para diseñar los cubos OLAP y posteriormente el análisis multidimensional.

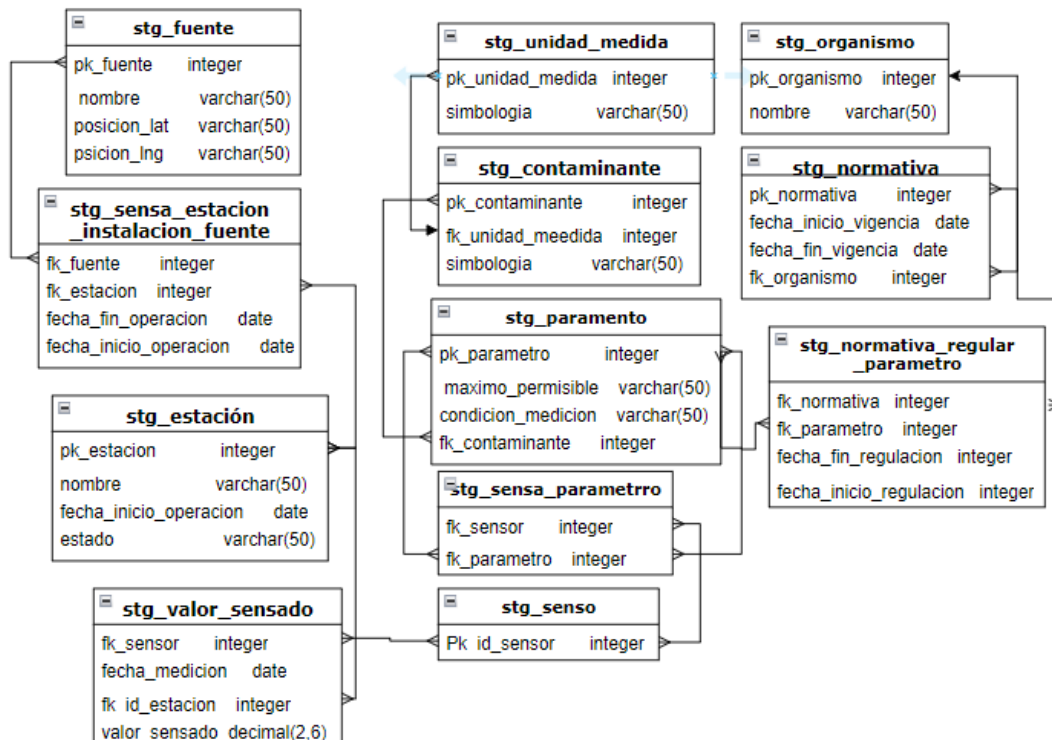


Figura 13. base de datos STAGE.

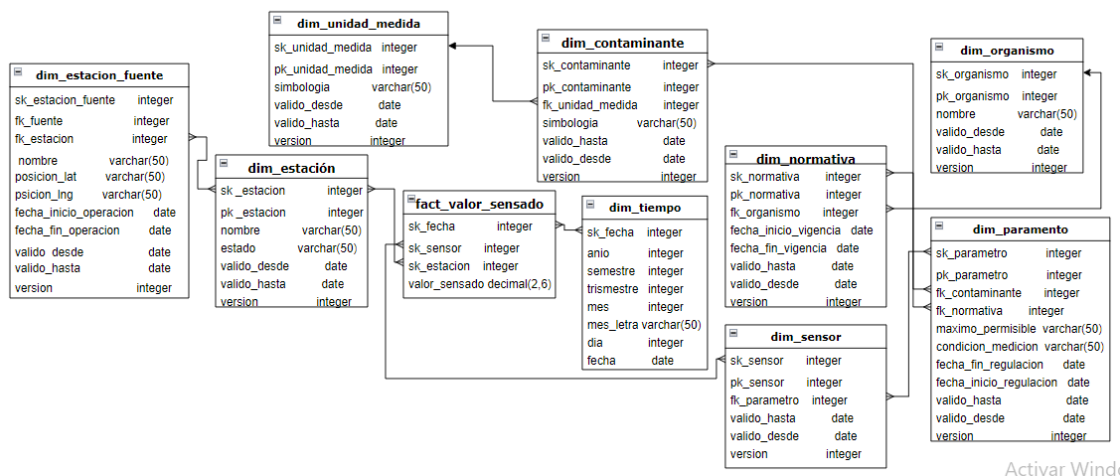


Figura 14. Base de datos Datawarehouse

A continuación, se describe el alcance de cada una de las dimensiones que conforma el modelo del datawarehouse, que sigue los lineamientos del modelo copo de nieve.

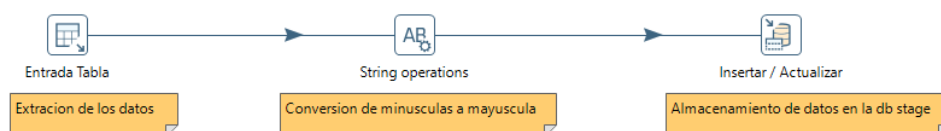
- DIM\_ESTACION\_FUENTE. Almacena los datos respecto a las fuentes de medición de datos ambientales.
- DIM\_ESTACION. Almacena datos recibidos por las estaciones como el nombre y el estado de la estación para medir los contaminantes.

- DIM\_FACT\_VALOR\_SENSOR. almacena los valores del sensor, como sk de fecha, sensor y estación para medir los contaminantes.
- DIM\_TIEMPO. almacenar datos temporales de forma estática (año, semestre, trimestre, mes y día) para la medición de contaminantes.
- DIM\_SENSOR. almacena los datos recepto de sensor como marca, modelo para la medición de los contaminantes.
- DIM\_PARAMETRO. almacenamiento de datos de acuerdo con los parámetros del nivel máximo permitido para la medición de contaminantes.
- DIM\_ORGANISMO. Almacena los datos recibidos de las agencias como identificadores de medición de contaminantes
- DIM\_CONTAMINATE. Almacena datos de mediciones atmosféricas, hídricas e hidrológicas de contaminantes.
- DIM\_UNIDAD\_MEDIDA. Almacena datos en unidades de medida de datos ambientales.

### 3.3. Procesamientos ETL

#### Carga de datos en la base de datos STAGE

Previo a almacenar los datos en el datawarehouse se empleó una base de datos denominada STAGE. En esta base de datos se normalizaron los datos según la estructura del datawarehouse, cuidando los datos relevantes para llevar a cabo una adecuada toma de decisiones. Adicionalmente, se cuidó la calidad de los datos almacenados. En la **¡Error! No se encuentra el origen de la referencia.Figura 15** se puede observar el esquema del proceso ETL para cargar los datos de la base de datos DbContaminante hacia STAGE. Cada uno de estos procesos depende de la información. No obstante, en vista de que la base de datos está normalizada, los procesos ETL no tuvieron alta complejidad.



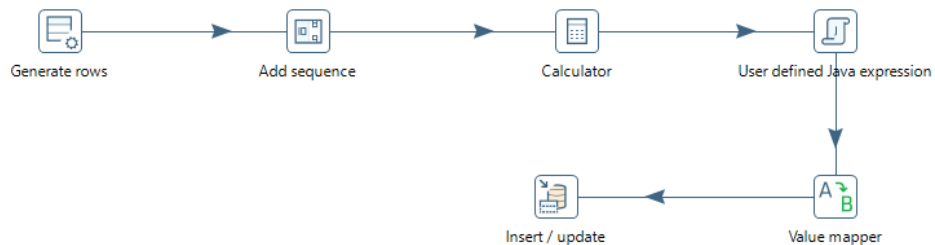
**Figura 15.** ETL para la carga de la tabla contaminantes desde la base de datos DbContaminantes hacia STAGE.

## Carga de datos de la base de datos STAGE hacia el datawarehouse

Mediante procesos ETL se pulieron aspectos relacionados con los datos a almacenar los datos del datawarehouse. En este proceso se eliminó cualquier anomalía en los propios datos previo a su almacenamiento en el repositorio. Este proceso es importante porque en base a los datos cargados se generan los resúmenes y se responde a las preguntas de negocio. El proceso ETL para cada dimensión fue similar. En la **Figura 16** se muestra un ejemplo de uno de los procesos ETL para la carga de la dimensión PARAMETRO. No obstante, la dimensión tiempo, se programó para que generara automáticamente periodos de fechas en base a los cuales se realizarán los análisis. EL proceso contemplado se ilustra en la **Figura 17**.



**Figura 16.** Proceso ETL para carga de datos de la dimensión PARAMETRO de STAGE al datawarehouse.



**Figura 17.** Proceso ETL para la generación automática de la dimensión Tiempo.

Asimismo, fue necesario crear un proceso ETL para pasar los datos a la tabla de hechos del datawarehouse. En la **Figura 18** se muestra el esquema de un proceso ETL para pasar los datos de STAGE a la tabla de hechos del repositorio.

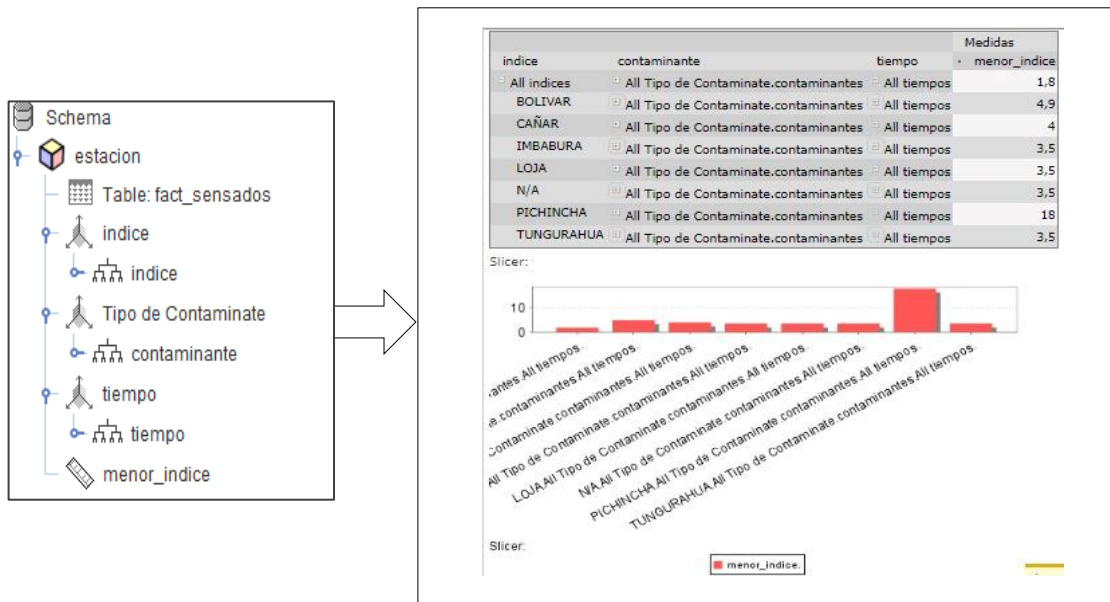


**Figura 18.** Proceso ELT de la tabla de hecho.

### 3.4. Respuestas a las preguntas de negocio

**Pregunta 1:** ¿En qué estación se han reflejado el menor índice de contaminantes (hídrico, edafológico, atmosférico) por (día, semana, mes, trimestre, año)?

Este cubo OLAP se creó en base a tres dimensiones, la primera fue la dimensión de índice, la segunda tipo\_contaminante, y la tercera el tiempo. Los niveles jerárquicos hacen referencia a la tabla DIM\_CONTAMINATE, y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de en qué estación se han reflejado el menor índice de contaminante, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite ver el menor índice. A continuación, se presenta el esquema de este cubo y los datos de respuesta en la **Figura 19**.



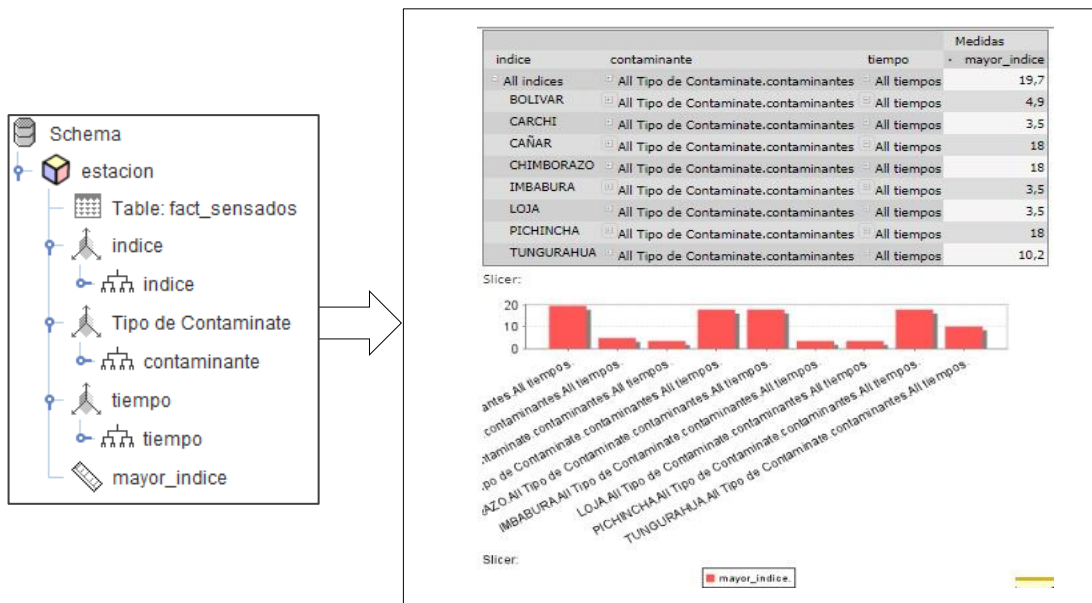
**Figura 19.** Cubo OLAP y resultados a la pregunta de negocio 1

Los resultados muestran que el menor índice es de 1.8, seguidos de las estaciones de Imbabura, Loja, Tungurahua, cuyo índice es de 3.5. Se encontró que Pichincha tienen un índice de 18. Dichos índices son en relación con la fecha es decir ya sea por día, semana, mes, trimestre o año como se puede observar en la **Figura 19**.

**Pregunta 2:** ¿En qué estación se han reflejado el mayor índice de contaminantes (hídrico, edafológico, atmosférico) por (día, semana, mes, trimestre, año)?

Este cubo se creó en base a tres dimensiones, la primera fue la dimensión de índice, la segunda tipo\_contaminante, y la tercera el tiempo, éstas están compuestas por niveles de

jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_CONTAMINATE, y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de en qué estación se han reflejado el mayor índice de contaminante, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite ver el mayor índice. A continuación, se presenta el esquema de este cubo para responder a la pregunta de negocios 2 en la **Figura 20**.



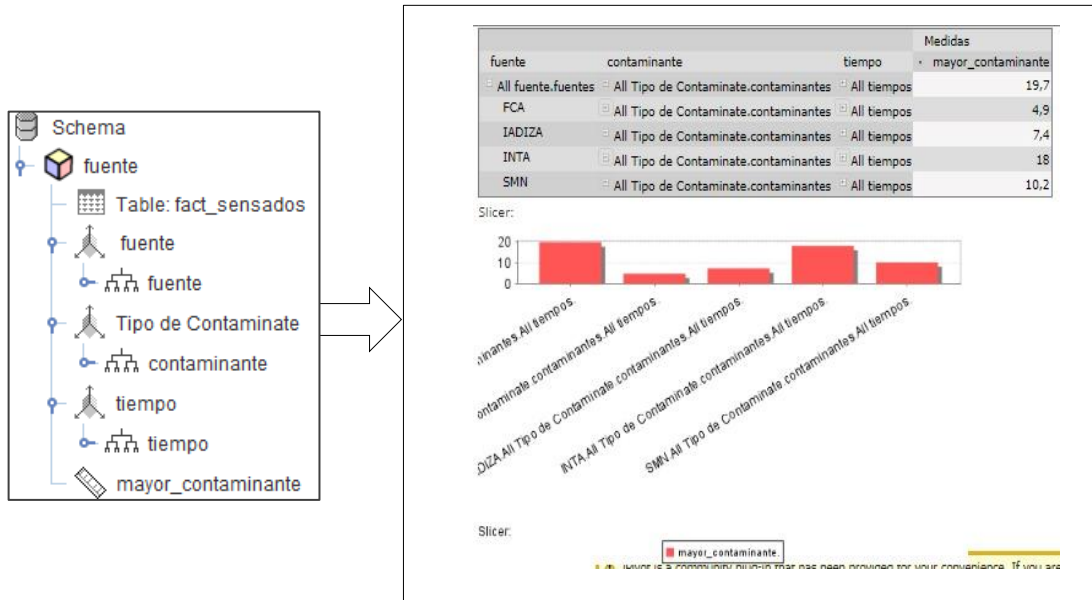
**Figura 20.**Cubo OLAP y resultados a la pregunta de negocio 2

En la **Figura 20** se realizó el desplazamiento del cubo 2, como se puede observar que el mayor índice es de 19.7, seguidos de las estaciones de Pichincha, Cañar, Chimborazo, cuyo mayor índice es de 18.

**Pregunta 3: Cuál es la fuente de mayor de contaminantes (hídrico, edafológico, atmosférico) por (día, semana, mes, trimestre, año)**

Este cubo se creó en base a tres dimensiones, la primera fue la dimensión de fuente, la segunda tipo\_contaminante, y la tercera el tiempo, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_ESTACION\_FUENTE, y DIM\_CONTAMINANTE. Debido a que se necesita responder a una pregunta acerca de cuál es la fuente de mayor contaminación, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas

para cada una de las dimensiones. Además, se agregó una medida que permite ver el mayor contaminante. A continuación, se presenta el esquema de este cubo para responder a la pregunta de negocios 3. El cubo descrito se ilustra en la **Figura 21**.

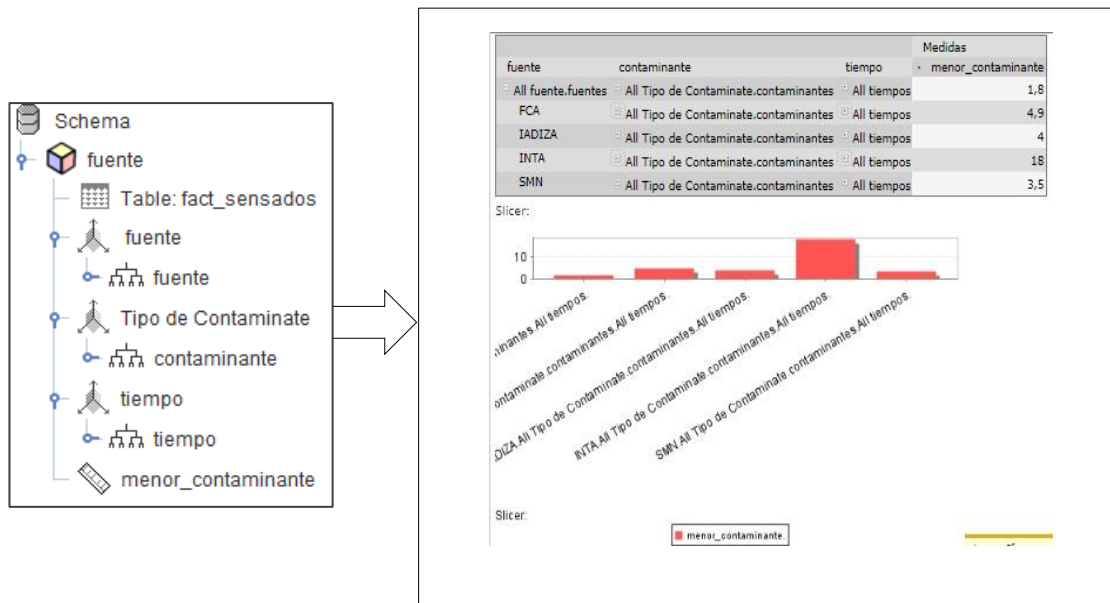


**Figura 21.** Cubo OLAP y resultados a la pregunta de negocio 3

En la **Figura 21**, se realizó el desplazamiento del cubo 3, como se puede observar que el mayor contaminante fue de 19.7, seguidos de las fuentes de INTA, SMN, cuyo mayor contaminante es de 18 y 10.2. Se encontró que FCA y IADIZA tienen un mayor contaminante de 4.9 y 7.4. Dichos son en relación con la fecha es decir ya sea por día, semana, mes, trimestre o año.

**Pregunta 4: Cuál es la fuente de menor de contaminantes (hídrico, edafológico, atmosférico) por (día, semana, mes, trimestre, año)**

Este cubo se creó en base a tres dimensiones, la primera fue la dimensión de fuente, la segunda tipo\_contaminante, y la tercera el tiempo, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_ESTACION\_FUENTE, y DIM\_CONTAMINANTE. Debido a que se necesita responder a una pregunta acerca de cuál es la fuente de mayor contaminación, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite ver el menor contaminante. A continuación, se presenta el esquema de este cubo para responder a la pregunta de negocios 4 según se muestra en la **Figura 22**.

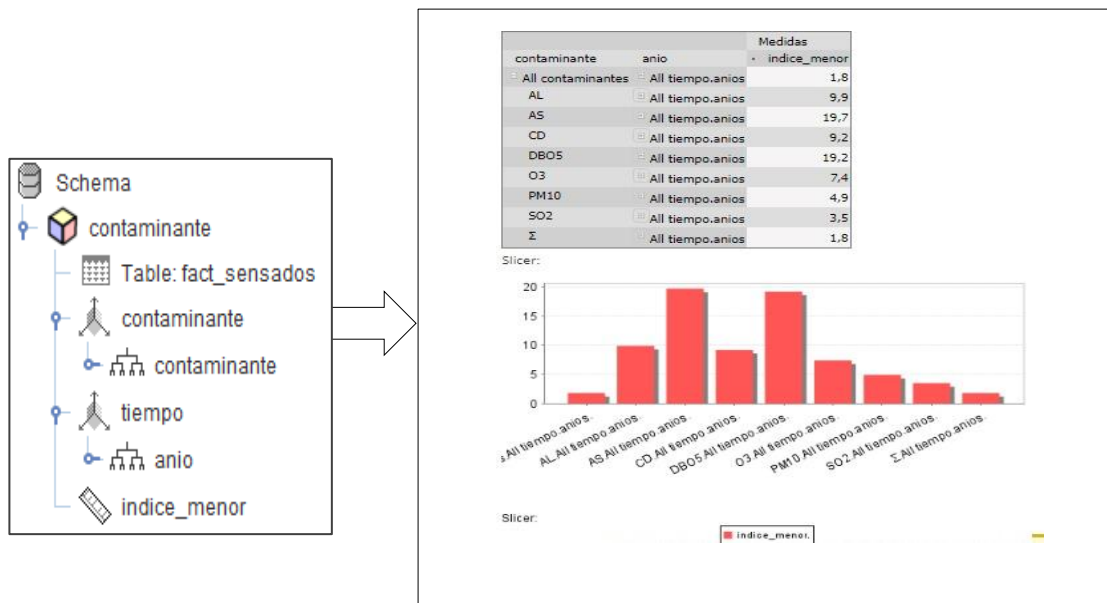


**Figura 22.** Cubo OLAP y resultados a la pregunta de negocio 4

En la **Figura 22**, se realizó el desplazamiento del cubo 4, como se puede observar que el menor contaminante fue de 1.8, seguidos de las fuentes de IADIZA, SMN, cuyo menor contaminante es de 3.5 y 4. Dichos son en relación con la fecha es decir ya sea por día, semana, mes, trimestre o año.

**Pregunta 5: Cuál es el contaminante (hídrico, edafológico, atmosférico) de menor índice por (día, semana, mes, trimestre, año)**

Este cubo se creó en base a dos dimensiones, la primera fue la dimensión contaminante y la segunda es la dimensión tiempo, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_CONTAMINANTE y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de cuál es el contaminante de menor índice, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite observar el menor índice. A continuación, se presenta el esquema de este cubo para responder a la pregunta de negocios 5 (**Figura 23**).

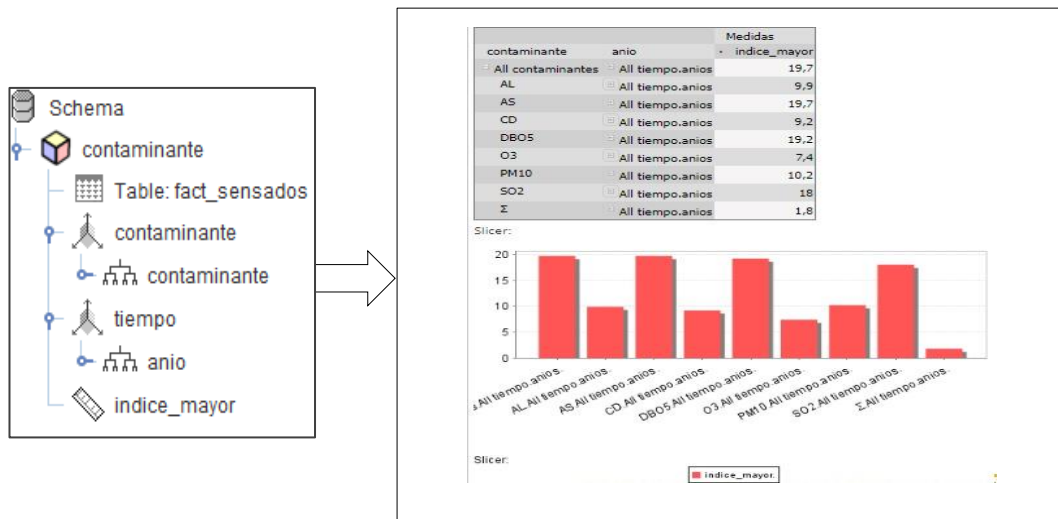


**Figura 23.** Cubo OLAP y resultados a la pregunta de negocio 5

En la **Figura 23** se realizó el desplazamiento del cubo 5, como se puede observar que el menor índice fue de 1.8, seguidos de los contaminantes de SO2, PM10, Σ cuyo menor índice es de 3.5, 4.9 y 1.8. Como se puede apreciar en el diagrama de barras.

**Pregunta 6: Cuál es el contaminante (hídrico, edafológico, atmosférico) de mayor índice por (día, semana, mes, trimestre, año)**

Este cubo se creó en base a dos dimensiones, la primera fue la dimensión contaminante y la segunda es la dimensión tiempo, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_CONTAMINANTE y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de cuál es el contaminante de mayor índice, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite observar el mayor índice. A continuación, en la **Figura 24**, se presenta el esquema de este cubo para responder a la pregunta de negocios 6.

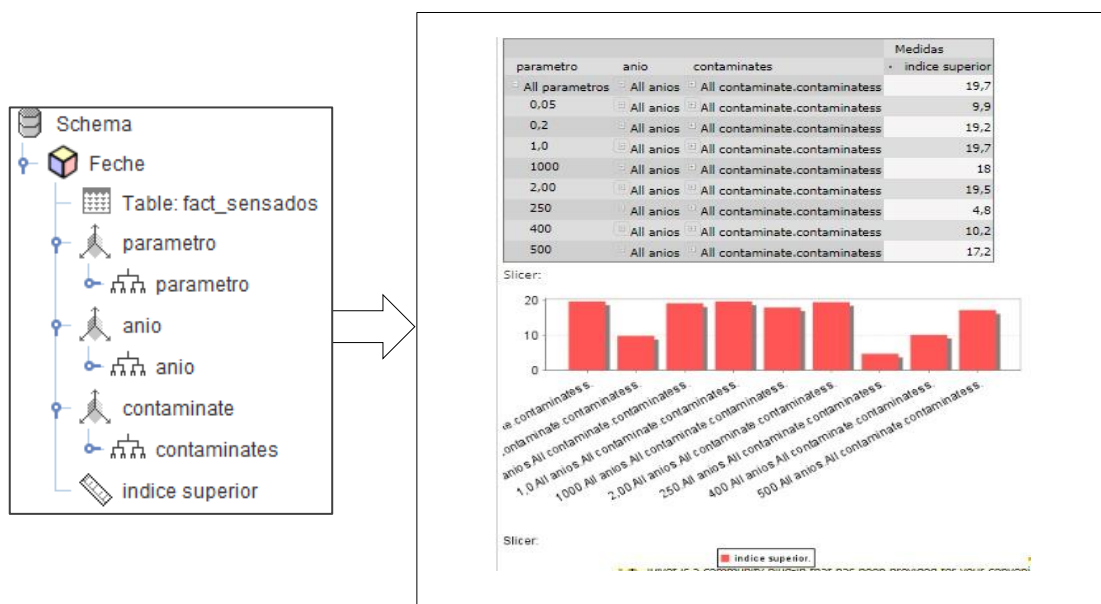


**Figura 24.** Cubo OLAP y resultados a la pregunta de negocio 6

En la **Figura 24** se realizó el desplazamiento del cubo 6, como se puede observar que el mayor índice fue de 19.7, seguidos de los contaminantes de SO2, DBO5, AS cuyo mayor índice es de 18, 19.2 y 19.7. Como se puede apreciar en el diagrama de barras.

**Pregunta 7: En qué fecha se han alcanzado valores superiores al máximo permisible de contaminantes (hídrico, edafológico, atmosférico)**

Este cubo se creó en base a dos dimensiones, la primera fue la dimensión parámetro y la segunda es la dimensión fecha, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_PARAMETRO y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de en qué fecha se han alcanzado valores superiores al máximo permisible del contaminante, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite observar el índice superior. A continuación, se presenta el esquema de este cubo para responder a la pregunta de negocios 7, según la **Figura 25**.

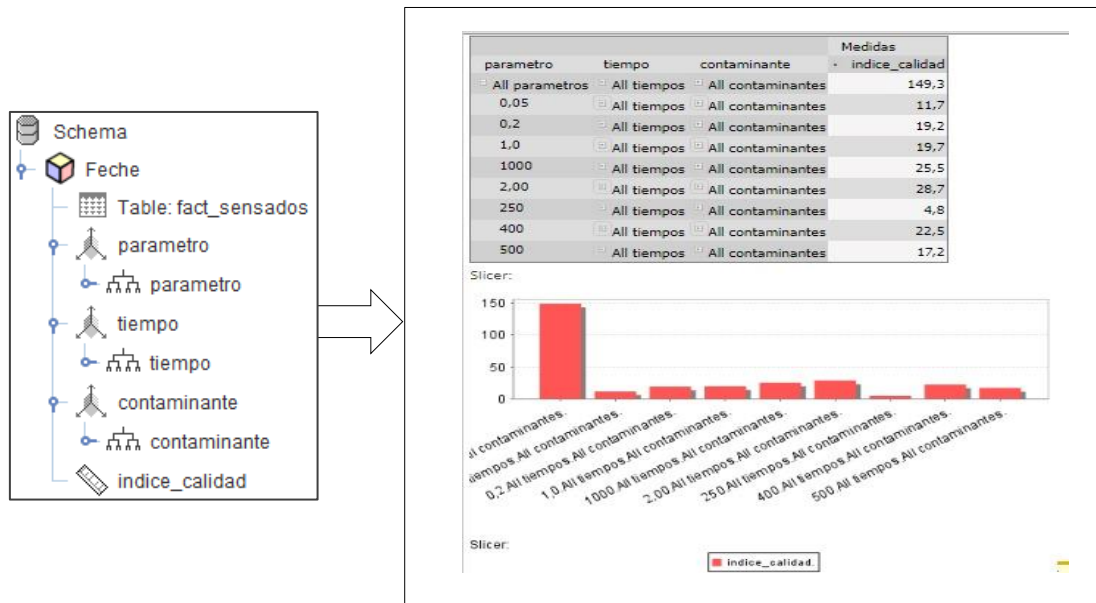


**Figura 25.** Cubo OLAP y resultados a la pregunta de negocio 7

En la **Figura 25** se realizó el desplazamiento del cubo 7, como se puede observar que el índice superior fue de 19.7, seguidos de los parámetros de 1.0, 2.00, 0.2 cuyo mayor índice es de 19.7, 19.5 y 19.2. Como se puede apreciar en el diagrama de barras.

**Pregunta 8: En qué fecha se han alcanzado valores óptimos (calidad) al máximo permisible de contaminantes (hídrico, edafológico, atmosférico)**

Este cubo se creó en base a dos dimensiones, la primera fue la dimensión parámetro y la segunda es la dimensión fecha, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_PARAMETRO y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de en qué fecha se han alcanzado valores óptimos al máximo permisible del contaminante, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite observar el índice de calidad. A continuación, se presenta el esquema de este cubo para responder a la pregunta de negocios 8 (**Figura 26**).

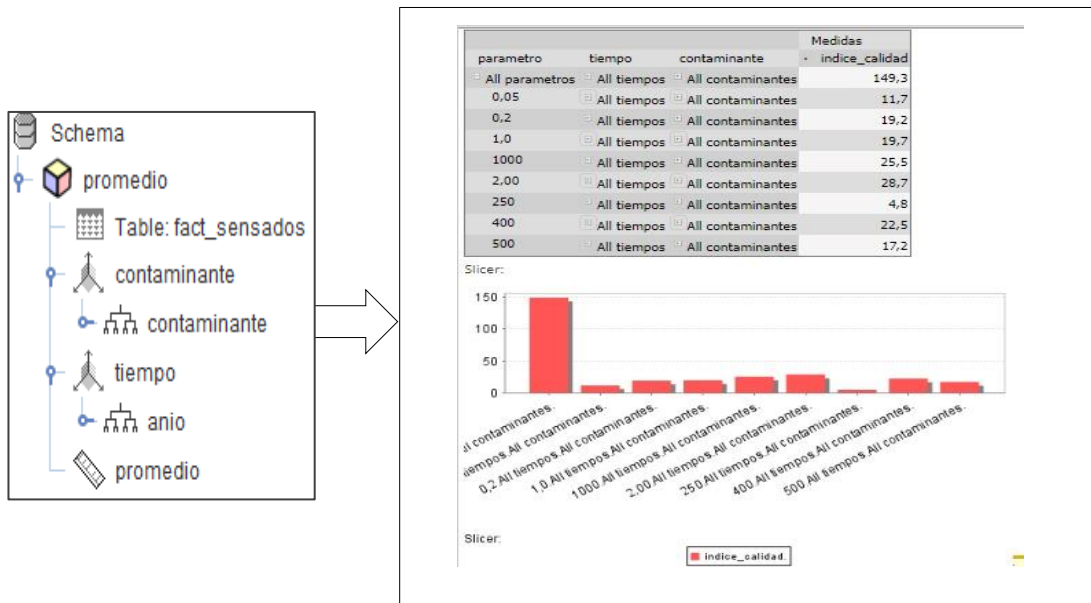


**Figura 26.** Cubo OLAP y resultados a la pregunta de negocio 8

En la **Figura 26** se realizó el desplazamiento del cubo 8, como se puede observar que el índice de calidad fue de 149.3, seguidos de los parámetros de 2.00 cuyo mayor índice es de 28.7. Como se puede apreciar en el diagrama de barras.

**Pregunta 9: Cuál es el promedio de contaminante atmosférico (Co2, O3, So2, No2, PM10, PM2.5) por (día, semana, mes, trimestre, año)**

Este cubo se creó en base a dos dimensiones, la primera fue la dimensión contaminante y la segunda es la dimensión tiempo, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_CONTAMINANTE y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de cuál es el promedio de contaminación atmosférica, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite observar el promedio. A continuación, en la **Figura 27**, se presenta el esquema de este cubo para responder a la pregunta de negocios 9.

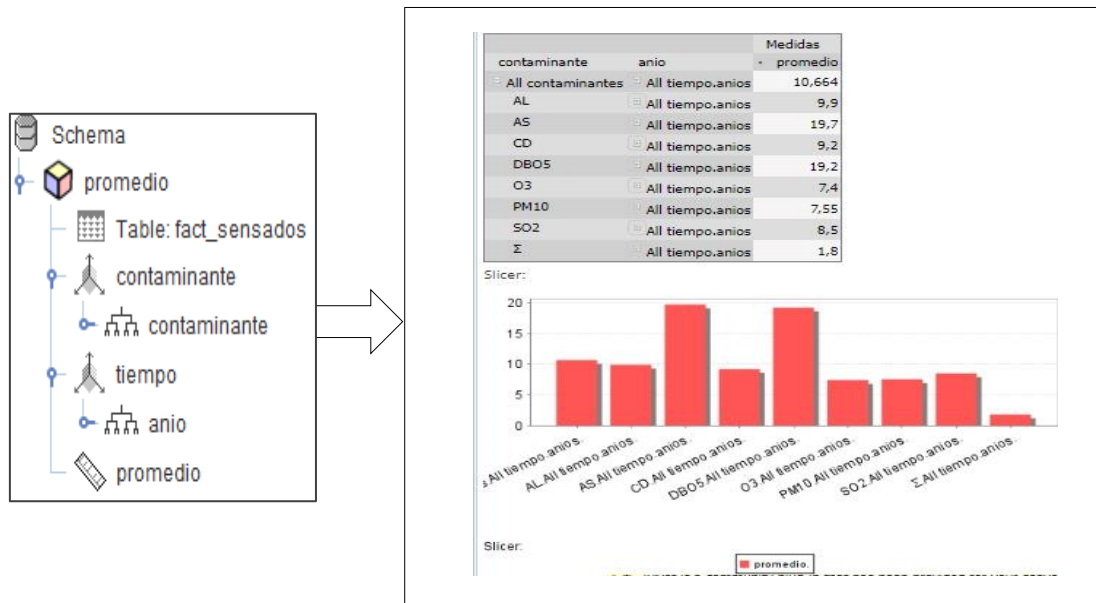


**Figura 27.** Cubo OLAP y resultados a la pregunta de negocio 9

En la **Figura 27** se realizó el desplazamiento del cubo 9, como se puede observar el promedio de contaminante atmosférico de AL fue de 9.9, el de O3 es de 7.4, el PM10 es de 7.55 y el S02 fue de 8.5, dando un promedio de contaminantes de 10.664, como se puede apreciar en la imagen del desplazamiento.

**Pregunta 10: Cuál es el promedio de contaminante edafológico (PH, DS, As, S, Bo, Br, Col, Cu) por (día, semana, mes, trimestre, año)**

Este cubo se creó en base a dos dimensiones, la primera fue la dimensión contaminante y la segunda es la dimensión tiempo, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_CONTAMINANTE y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de cuál es el promedio de contaminación edafológico, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite observar el promedio. A continuación, se presenta el esquema de este cubo para responder a la pregunta de negocios 10 (**Figura 28**).

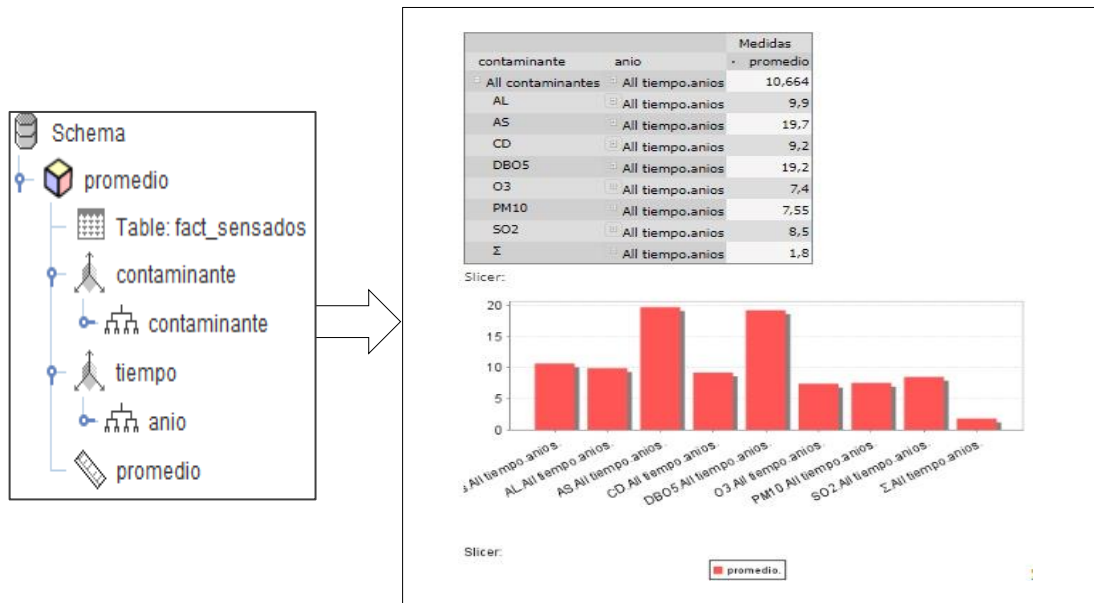


**Figura 28.** Cubo OLAP y resultados a la pregunta de negocio 10

En la **Figura 28** se realizó el desplazamiento del cubo 10, como se puede observar el promedio de contaminante edafológico de AL fue de 9.9, el de O3 es de 7.4, el PM10 es de 7.55 y el S02 fue de 8.5, dando un promedio de contaminantes de 10.664, como se puede apreciar en la imagen del desplazamiento.

**Pregunta 11: Cuál es el promedio de contaminante hídrico (Al, N-Amoniaco, NH4, As, Ba, Cd, Cn, Cl, Cu, Fenol, Cr+6, DBO5, CaCO3) por (día, semana, mes, trimestre, año)**

Este cubo se creó en base a dos dimensiones, la primera fue la dimensión contaminante y la segunda es la dimensión tiempo, están compuestas por niveles de jerarquía, cada uno de estos niveles hace referencia a la tabla DIM\_CONTAMINANTE y DIM\_TIEMPO. Debido a que se necesita responder a una pregunta acerca de cuál es el promedio de contaminación hídrico, es necesario incluir la tabla de hechos FAC\_SENSADOS en la cual se establecen las claves subrogadas para cada una de las dimensiones. Además, se agregó una medida que permite observar el promedio. A continuación, se presenta el esquema de este cubo para responder a la pregunta de negocios 11, según el esquema de la **Figura 29**.



**Figura 29.** Cubo OLAP y resultados a la pregunta de negocio 11

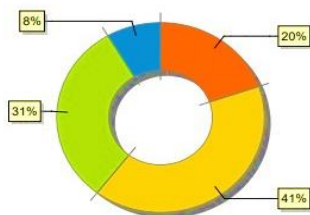
En la **Figura 29** se realizó el desplazamiento del cubo 11, como se puede observar el promedio de contaminante hídrico de AL fue de 9.9, el de O3 es de 7.4, el PM10 es de 7.55 y el S02 fue de 8.5, dando un promedio de contaminantes de 10.664, como se puede apreciar en la imagen del desplazamiento.

### 3.3. Cuadros de mando (*dashboard*)

El diseño de los cuadros de mando como se puede observar en la **Figura 30**, es otra forma de representar las respuestas a las preguntas de negocio donde se muestran las principales métricas que interviene en la consecuencia de los objetivos de las preguntas.

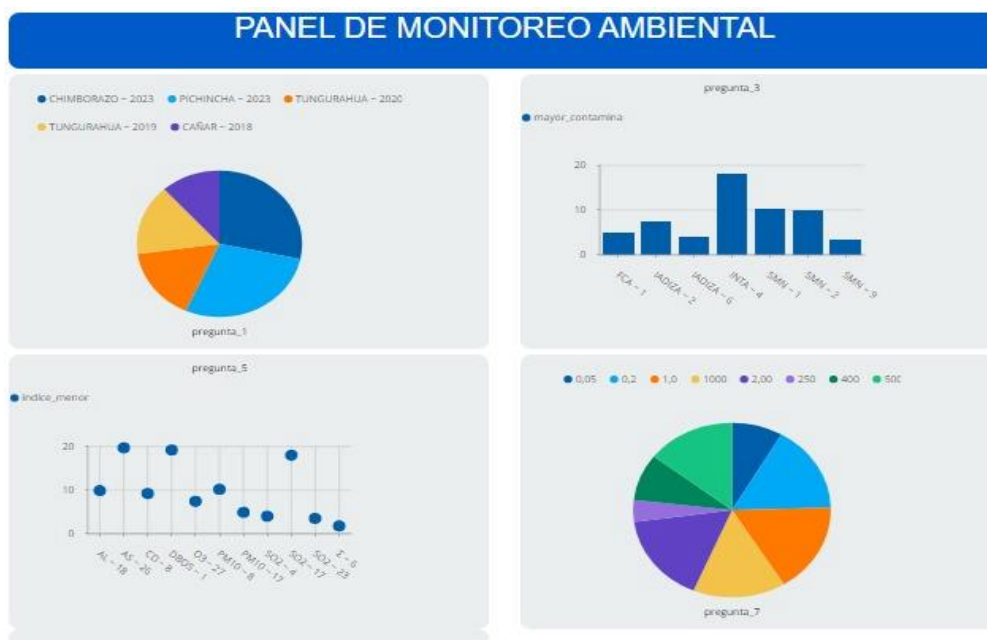
También se puede observar un ejemplo del diseño de reporte para crear un resumen de datos que proporciona el sistema de inteligencia de negocios a través del datawarehouse. Éstos consultas que se pueden obtener de manera impresa.

FUENTE	INDICE	ANIO
IADIZA	4	2.017
SMN	3,5	2.023
INTA	18	2.023
IADIZA	7,4	2.018
FCA	4,9	2.022
SMN	10,2	2.020
SMN	9,9	2.019



**Figura 30.** Reporte del sistema

Los resultados proporcionados por el sistema se pueden ver a través de un cuadro de mando integral como el ilustrado en la **Figura 31**. A través de este cuadro de mando es posible monitorear los datos y actualizar los resultados acordes a como se carguen los datos en el datawarehouse, fuente principal de datos del sistema de inteligencia de negocios.



**Figura 31.** Control de monitoreo

## CAPÍTULO IV: DISCUSIÓN

A continuación, se realiza una comparación y discusión obtenida en este estudio y en los antecedentes descritos en el capítulo Marco Teórico.

En la investigación realizada por Toma et al. [69] se desarrolló un sistema de monitoreo de contaminación ambiental en zonas urbanas, la cual comprendía el uso de una red de sensores de IoT para la obtención de la información. Dicha información era almacenada en una base de datos no relacional Mongo DB, y por último se visualizó en una plataforma desarrollada con Apache. En comparación, de este estudio con la investigación desarrollada, en esta investigación se decidió utilizar una base de datos relacional PostgreSQL por su facilidad de administración y por ser compatible con las herramientas de la Suite de Pentaho. En este sentido, la base de datos PostgreSQL no representó conflictos para la implementación de las fuentes de datos del sistema desarrollado.

Por otro lado, en la investigación propuesta por Desai et al. [70] y Crisóstomo [73], los autores plantearon un enfoque similar al trabajo propuesto por Toma et al. [69], el primer estudio utilizó los servicios de la plataforma de Azure para el almacenamiento de la información y los servicios de aprendizaje automático para un sistema de predicción. En cambio, el segundo estudio utilizó la base de datos de SQL Server porque los datos referenciales ya se encontraban en este tipo de base de datos. En ambos trabajos se utilizó la herramienta de Power BI para el procesamiento y presentación de los datos. El estudio de investigación desarrollado, a diferencia de los estudios descritos, aplicó el proceso de ingeniería de negocio utilizando la Suite de Pentaho. Aunque Pentaho es más complejo que Power BI, esta herramienta se distribuye bajo licencia libre y permitió el procesamiento de datos, formación de los cubos OLAP para responder a las preguntas de negocios planteada, y por último, permitió de manera exitosa la presentación de los resultados a través de gráficos estadísticos a través de cuadros de mando.

Siguiendo la discusión con los estudios de los antecedentes, en el estudio realizado por Haghparast et al. [72] se creó un sistema de monitoreo de factores de contaminación ambiental (aire, agua y suelo) a través de la herramienta ArcGIS para el mapeo de zonas de mayor riesgo y de la lógica difusa para la toma de decisiones. El estudio de investigación desarrollado tuvo el mismo enfoque en cuanto a la selección de los factores ambientales a evaluar, donde se presentó que el contaminante atmosférico era mayor en

comparación a los demás criterios. En este sentido, tienen un gran parecido los estudios, no obstante, en el caso de la propuesta desarrollada, todo se contextualizó a la normativa ecuatoriana, algo que facilitará llevar el sistema al contexto de Esmeraldas, al momento de contar con datos de calidad ambiental.

Finalmente, en el estudio propuesto por Baralis et al. [74] y por Khot et al. [71] los esfuerzos se centraron en definir cuáles son los factores ambientales en el aire que deberían ser considerados para el monitoreo de los contaminantes similar a la investigación de Haghparast et al. [72], no obstante, el estudio de investigación propuesto tomó como base estos parámetros de contaminación ambiental y ayudó a definir los indicadores ambientales a considerar, esto es, PM2.5, CO2, O3, NO2, SO2, PM10, AL, NH3, NH4, AS, BA, CD, CN-, CL, CU, FENOL, CR+6, CACO3, SAR, S y B. Cabe destacar que la inclusión de los indicadores de las categorías de contaminantes hídricos y edafológico, ayudan en la precisión para la toma de decisiones.

# CAPITULO V: CONCLUSIONES Y RECOMENDACIONES

## 5.1. Conclusiones

Se puede concluir que la investigación permite identificar los principales aspectos que involucra a la toma de decisiones de contaminación medioambiental, el cual se evaluó, analizó y determinó las normativas actuales y sus máximos permisibles de contaminación con los respectivos valores de cada uno de los recursos como aire, agua y suelo.

Los principales componentes del manejo de BI fueron determinar la fuente de información de la base de datos del sistema desarrollado, que permitió el diseño del modelo de datawarehouse, bajo el modelo copo de nieve, el cual permitió moldear cubos OLAP orientados a responder las preguntas de negocio planteadas.

La implementación de BI es muy recomendable, ya que permite la toma de decisiones con los datos obtenidos de la investigación, es importante llevar o contar con un sistema de BI que ayuda a la toma de decisiones en lo que respecta al tema de investigación como es la contaminación del aire, agua y suelo. Al contar con un nuevo sistema se tiene información oportuna y veraz que favorecerá para que las autoridades se centren en realizar un control sobre la contaminación del aire, agua y suelo.

Las herramientas de BI de la Suite de Pentaho permitieron extraer y analizar los datos con técnicas como son las ETL, es decir permite extraer, transformar y cargar datos de fuentes heterogéneas. Esto ayudó mucho a la hora de crear la base de datos STAGE y el DATWAREHOUSE, sobre todo porque no se contaba con datos normalizados. Los datos que se encontraron, contextualizados a Ecuador, fueron cargados de manera exitosa en la base de datos del sistema, y consecuentemente, se pudo aplicar el análisis multidimensional, proceso típico de un sistema de inteligencia de negocios.

Importante mencionar que los cubos OLAP son procedimientos multidimensionales que proporciona el análisis rápido de los datos, el cual permite obtener los reportes como resultado de recopilación de los datos para la evaluación unitaria e integrales de la investigación propuesta. No obstante, en caso de que el sistema tenga un crecimiento exponencial es importante contar con un servidor dedicado al almacenamiento de datos y al análisis multidimensional.

## **5.2 Recomendaciones**

Es recomendable que el base de datos del Datawarehouse sea más voluminoso, es decir que cuente con datos de más años, esto facilita para que el trabajo de inteligencia de negocio sea más complejo y que las decisiones lleguen a ser más acertadas si es el caso de que la información se use para la toma de decisiones en un contexto real. Además, contar con más datos ayudaría para que se aplique Minería de Datos, y así, descubrir patrones de comportamiento de la contaminación en un contexto específico.

Ademas, es importante que la información y datos almacenados dentro de sistemas de información general de datos sea del contexto de Esmeraldas. Esto ayudaría específicamente a los ciudadanos de Esmeraldas que tienen la problemática de la contaminación ambiental y conviven con ella. No obstante, hay una dependencia de datos capturados por redes de monitoreo.

## Referencias

- [1] C. Gonzalez-Cruz *et al.*, “Machine Learning in Melanoma Diagnosis. Limitations About to be Overcome,” *Actas Dermosifiliogr.*, vol. 111, no. 4, pp. 313–316, 2020, doi: 10.1016/j.ad.2019.09.002.
- [2] P. Espejo, C. Morales, and E. Galindo, “Minería De Datos Sobre Objetos De Aprendizaje,” *Comput. Av. energía y plasmas*, no. 81, p. 8358, 2012, [Online]. Available: <https://helvia.uco.es/bitstream/handle/10396/20507/2020000002142.pdf?sequence=1&isAllowed=y>
- [3] C. Barrera, J. González, and G. Cáceres, “Toma de decisiones en el sector turismo mediante el uso de Sistemas de Información Geográfica e inteligencia de negocios,” *Rev. Científica*, vol. 38, no. 2, pp. 160–173, 2020, doi: 10.14483/23448350.15997.
- [4] H. Zamora Carrillo, N. Novoa Torres, and D. R. Bermúdez Huérfano, “Nociones, consideraciones y ventajas de la inteligencia de negocios BI,” *Rev. vínculos*, vol. 16, no. 2, pp. 280–287, 2019, doi: 10.14483/2322939x.15592.
- [5] I. William and G. Guerrero, “Arquitectura Empresarial – Dominios Y Beneficios,” pp. 87–92, 2016.
- [6] F. Rodriguez, *Desarrollo Regional Y Territorio . Nuevos Planteamientos Y Perspectivas*. 2007.
- [7] J. Sluijter and M. Otten, “Business intelligence (BI) for personalized student dashboards,” *ACM Int. Conf. Proceeding Ser.*, pp. 562–563, 2017, doi: 10.1145/3027385.3029458.
- [8] D. G. Martínez, “Optimización Del Mantenimiento Industrial Mediante Técnicas Bi , Aplicación De Un Cuadro De Mandos Integral,” *Bus. intelligence*, p. 53, 2019, [Online]. Available: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/96547/7/gonmardTFG0619memoria.pdf>
- [9] J. Palacios, E. Medina, J. Ochoa, and M. Torres, “Business Intelligence Aplicado Al Sector Salud,” *Rev. Arbitr. Interdiscip. Koinonía*, vol. 5, no. 3, p. 622, 2020, doi: 10.35381/r.k.v5i3.914.
- [10] E. Sevryukova, A. Volkov, G. Kuznetsov, A. Golovlev, and D. Yakovenko, “Development Of The Automated Environment Monitoring System,” *Proc. 2018 IEEE Conf. Russ. Young Res. Electr. Electron. Eng. ElConRus 2018*, vol. 2018-Janua, pp. 1936–1939, 2018, doi: 10.1109/EIConRus.2018.8317488.
- [11] D. Sikora-Fernandez, “Factores De Desarrollo De Las Ciudades Inteligentes,” *Rev. Univ. Geogr.*, vol. 26, no. 1, pp. 135–152, 2017, [Online]. Available: [http://www.scielo.org.ar/scielo.php?script=sci\\_abstract&pid=S1852-42652017000100007](http://www.scielo.org.ar/scielo.php?script=sci_abstract&pid=S1852-42652017000100007)
- [12] G. Melo, A. Marotta, and F. Serra, “Sistema de Business Intelligence para análisis de datos de calidad del aire de la Intendencia de Montevideo,” *Univ. la República (Uruguay). Fac. Ing.*, pp. 5–165, 2019, [Online]. Available: <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/22889>

- [13] P. Guillermo, J. Moreno, and J. Lopez, “Estudio De Contaminantes Emergentes Urbanos : Técnicas Analíticas Y Tecnologías De Tratamiento,” *Ing. Química y Ambient.*, p. 95, 2020, [Online]. Available: <https://repositorio.upct.es/handle/10317/8963>
- [14] Y. Reyes and L. Nuñez, “La Inteligencia De Negocio Como Apoyo A La Toma De Decisiones En El Ámbito Académico,” *L*, vol. 3, no. 2, pp. 2013–2014, 2015, [Online]. Available: <https://www.upo.es/revistas/index.php/gecontec/article/download/1745/1427/5545>
- [15] L. Tello and A. Sánchez, “La Contaminación Ambiental En Los Acuíferos De Ecuador. Necesidad De Su Reversión Desde Las Políticas Públicas Con Enfoque Bioético,” *Rev. Iberoam. Bioeconomía y Cambio Climático*, vol. 5, no. 9, pp. 1053–1102, 2019, doi: 10.5377/ribcc.v5i9.7946.
- [16] A. Silva, O. Espino-Barros, L. López-Reyes, and P. López, *La Biodiversidad En Colima, Estudio De Estado*, no. August. 2016. doi: 10.13140/2.1.1293.6321.
- [17] M. Altamira and F. Aguirre, “Inteligentes Negocio,” p. 11, 2018, [Online]. Available: <https://www.gestiopolis.com/inteligencia-los-negocios/>
- [18] E. Ahumada-Tello and J. Perusquia-Velasco, “Inteligencia De Negocios: Estrategia Para El Desarrollo De Competitividad En Empresas De Base Tecnológica,” *Contaduria y Adm.*, vol. 61, no. 1, pp. 127–158, 2016, doi: 10.1016/j.cya.2015.09.006.
- [19] D. M. Cordero Guzman and G. Rodríguez López, “La Inteligencia De Negocios: Una Estrategia Para La Gestión De Las Empresas Productivas.,” *Cienc. Unemi*, vol. 10, no. 23, p. 40, 2017, doi: 10.29076/issn.2528-7737vol10iss23.2017pp40-48p.
- [20] M. Alva, M. Callan, and A. Yim, “Solución De Inteligencia De Negocios Para Mejorar La Toma De Decisiones En La Dirección Ejecutiva Del Hospital La Caleta,” vol. 6, p. 104, 2019, [Online]. Available: <https://1library.co/document/yjd6wxky-solucion-inteligencia-negocios-mejorar-decisiones-direccion-ejecutiva-hospital.html>
- [21] C. Pacci Ayala, “Aplicando Inteligencia De Negocios De Autoservicio, Utilizando Power Bi, Para La Toma De Decisiones Dentro De Una Pyme En La Región De Tacna.,” *Univ. Priv. Tacna*, p. 140, 2017, [Online]. Available: <https://docplayer.es/90439928-Universidad-privada-de-tacna-tesis.html>
- [22] L. Canos, C. Pons, M. Valero, J. Maheut, and C. Baum, “Gene Transfer Into Hematopoietic Cells: From Basic Science To Clinical Application.,” *Methods Mol. Biol.*, vol. 506, 2009, doi: 10.1007/978-1-59745-409-4.
- [23] B. Lopez Bernal, “Guia Para La Contruccion De Data Warehouse,” *J. Phys. A Math. Theor.*, vol. 44, no. 8, p. 133, 2011, doi: 10.1088/1751-8113/44/8/085201.
- [24] A. Villanueva and W. Flores, “Sistema Para La Toma De Decisiones Para La Inteligencia De Negocios Del Área Comercial De La Empresa Ingram Micro S.A.,” *Univ. César Vallejo*, pp. 6–88, 2017, [Online]. Available: [https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/16387/Villanueva\\_MA.pdf?sequence=1&isAllowed=y](https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/16387/Villanueva_MA.pdf?sequence=1&isAllowed=y)

- [25] J. Sánchez Espinoza and C. Canelo Sotelo, “Modelo De Data Warehouse Con Aplicacion De Inteligencia De Negocios Para Las Pymes,” *Cienc. Desarro.*, vol. 6, no. 21, pp. 113–123, Jun. 2019, doi: 10.33326/26176033.2017.21.737.
- [26] G. Zegarra Fuentes, “Solución De Inteligencia De Negocios Orientada A Mejorar La Toma De Decisiones En Las Operaciones Mineras De Extracción Y Metalurgia De Hochschild Mining,” pp. 1–155, 2015, [Online]. Available: <https://1library.co/document/y8gjw3wz-solucion-inteligencia-orientada-decisiones-operaciones-extraccion-metalurgia-hochschild.html>
- [27] A. Cabrera Torres, E. Moran Cabrera, and R. Rodriguez Anormaliza, “Uso De La Tecnología Data Warehouse En Unidades Educativas De Nivel Medio: Consideraciones Teóricas,” *Cienc. Unemi*, vol. 7, no. 11, p. 51, 2015, doi: 10.29076/issn.2528-7737vol7iss11.2014pp51-57p.
- [28] N. Fernández Carrión and C. López Rodríguez, “La Influencia De La Inteligencia De Negocios En El Análisis De Información De Ventas De La Importadora Y Distribuidora Jiménez E.I.R.L, En La Ciudad De Nueva Cajamarca,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2018.
- [29] J. Hernández Rubicelia, R. Téllez Villegas, and J. Jaramillo Rocha, “Data Warehouse Para La Toma De Decisiones En El Departamento De Tutorías,” *Jóvenes en la Cienc. Jóvenes Investig. Vol. 3, No.1*, vol. 3, no. 1, pp. 498–500, 2017, [Online]. Available: <https://www.jovenesenlaciencia.ugto.mx/index.php/jovenesenlaciencia/article/view/875/pdf1>
- [30] J. Yalan Castillo and L. Palomino Paniora, “Implementacion De Un Datamart Como Un Solcuion De Inteligencia De Negocios Para El Area De Logistica De T-Impluso,” *Rev. Investig. Sist. E Informática*, vol. 10, no. 1, pp. 53–63, 2012, doi: <https://doi.org/10.15381/risi.v10i1.5713>.
- [31] J. Wang and N. Wang, “An Open-Source Solution Of Business Intelligence In China Railway Passenger Transportation Decision Support,” vol. 30, pp. 135–140, 2016, doi: 10.2991/iconfem-16.2016.23.
- [32] C. . Buñay Cujilema and B. Hidalgo Ponce, “Desarrollo De Un Sistema Aplicando Inteligencia De Negocios Al Sgia-Alpa Para Facilitar La Toma De Decisiones,” 2019, [Online]. Available: <http://dspace.espace.edu.ec/handle/123456789/13005>
- [33] K. Damiano Espinoza and H. Robalino Gomez, “Implementación De Una Data Mart Para Mejorar La Toma De Decisiones En El Área De Continuidad Del Negocio, Banco Ripley,” 2020, [Online]. Available: [https://repositorio.utp.edu.pe/bitstream/handle/20.500.12867/3128/Karla\\_Damiano\\_Trabajo\\_de\\_Suficiencia\\_Profesional\\_Titulo\\_Profesional\\_2020.pdf?sequence=4&isAllowed=y](https://repositorio.utp.edu.pe/bitstream/handle/20.500.12867/3128/Karla_Damiano_Trabajo_de_Suficiencia_Profesional_Titulo_Profesional_2020.pdf?sequence=4&isAllowed=y)
- [34] A. Wong Ángeles and M. Suxe Ramírez, *An Update On The Pathophysiology Of Osteoarthritis*, vol. 59, no. 5–6. 2016. doi: 10.1016/j.rehab.2016.07.004.
- [35] A. Falcón de la Cruz, *Implementación De Un Datamart Para La Gestión De Indicadores De Deserción Universitaria Relacionados A La Calidad Docente Administrativa De La Modalidad Cpe De La Universidad Científica Del Sur*. 2020. [Online]. Available: <https://repositorio.cientifica.edu.pe/handle/UCS/1132>

- [36] Z. Guzmán and H. Aguilar, “Construcción De Cubo Olap En Microsoft Analysis Services Y Microsoft Excel,” *Rev. Investig. en Tecnol. la Inf.*, vol. 8, no. 15, pp. 41–49, 2020, doi: 10.36825/riti.08.15.005.
- [37] J. Bernal Vargas, M. . Veliz Peñafiel, and M. Vinueza Morales, “Diseño De Un Procesamiento Analítico En Línea (Cubo Olap) Para La Toma De Decisiones Gerenciales En El Área De Ventas Del Supermercado ‘Valdiviezo’ Ubicado En El Cantón La Troncal,” vol. 21, no. 1, pp. 1–9, 2020, [Online]. Available: <http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203>
- [38] L. Silva Solano, “Business Intelligence: Un Balance Para Su Implementación,” *InnovaG*, no. 3, pp. 27–36, 2017, [Online]. Available: <http://revistas.pucp.edu.pe/index.php/innovag/article/view/19742>
- [39] R. Berlanga, M. Aramburu, I. Lanza, D. Llido, M. Llego., and I. Sanz, “Dynamic Slod-Bi : Infraestructura Din´Amica De Inteligencia De Negocio Social,” *Intel. Neg. Soc.*, pp. 1–4, 2018, [Online]. Available: <http://repositori.uji.es/xmlui/handle/10234/189273>
- [40] C. Pereda Medina, M. Cabrera Sanchez, and A. Ullon Ramirez, “Solución De Inteligencia De Negocios (Bi) Para Mejorar El Análisis De La Información En Los Procesos De Ventas De La Empresa Ingenieros En Accion S.R.L. Utilizando La Arquitectura De Pentaho BI,” *Univ. Priv. Antenor Orrego*, p. 157, 2016, [Online]. Available: <https://hdl.handle.net/20.500.12759/4765>
- [41] A. Montero, J. Vega, and M. Ampuero, “Solución De Inteligencia De Negocio Para Métricas De Gestión De Proyectos,” *Rev. Cuba. Ciencias Informáticas*, vol. 9, no. 0, pp. 85–97, 2015, [Online]. Available: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2227-18992015000500006](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992015000500006)
- [42] M. Herrera, F. Ospina, and C. Mejia, “Integración De Etapas Técnicas De Metodologías Bi Con El Método De Gestión Lean Analytics: Una Metodología Para Obtención De Conocimiento Milena Herrera Santos Felipe Ospina Valencia Director,” *Interciencia*, vol. 489, no. 20, pp. 313–335, 2018.
- [43] J. Acosta, D. Flórez, and C. Diaz, “Diseño E Implementación De Prototipo Bi Utilizando Una Herramienta De Big Data Para Empresas Pymes Distribuidoras De Tecnología,” p. 87, 2015, [Online]. Available: [https://repository.ucatolica.edu.co/bitstream/10983/2543/1/PROYECTO\\_FINAL\\_.pdf](https://repository.ucatolica.edu.co/bitstream/10983/2543/1/PROYECTO_FINAL_.pdf)
- [44] E. Leonard and Y. Castro, “Metodologías Para Desarrollar Almacén De Datos,” vol. 7, pp. 1–12, 2013, [Online]. Available: <http://www.redalyc.org/pdf/1939/193930080003.pdf>
- [45] E. Sevilla and C. Zepeda, “Guía Metodológica Para La Definición Y Desarrollo De Un Data Warehouse,” p. 174, 2003, [Online]. Available: <http://biblioteca.uam.edu.ni/xmlui/bitstream/handle/721007/1445/00902630.pdf?sequence=1&isAllowed=y>
- [46] R. Guillen, “Sistema De Soporte De Decisiones Con Tecnología Data Warehouse Para La Gestión De La Información De La Empresa Mallku Import Sac,” pp. 1–172, 2017, [Online]. Available:

[http://repositorio.unap.edu.pe/bitstream/handle/UNAP/3410/Hermosa\\_Mendoza\\_Gisela del Carmen.pdf?sequence=1&isAllowed=y](http://repositorio.unap.edu.pe/bitstream/handle/UNAP/3410/Hermosa_Mendoza_Gisela%20del%20Carmen.pdf?sequence=1&isAllowed=y)

- [47] E. Macas, F. Cevallos, and W. Bustamante, "Data Warehouse: Análisis Multidimensional De Bafici Utilizando Power Pivot," *Rev. Espac.*, vol. 39, no. 34, pp. 2–3, 2018, [Online]. Available: <http://www.revistaespacios.com/a18v39n34/18393424.html>
- [48] J. Orellana, "Contaminación," *Ing. Sanit. UTN - FRRO*, pp. 1–27, 2005.
- [49] R. Gómez López, "Los Recursos Ambientales : Una Reflexión Sobre Su Gestión En El Desarrollo Del Futuro," pp. 1–15, 2017, [Online]. Available: <https://old.aecr.org/web/congresos/2005/ponencias/p14.pdf>
- [50] Ministerio del Ambiente de Ecuador, "Norma De Calidad Del Aire Ambiente O Nivel De Inmision." pp. 1–16, 2011. [Online]. Available: <http://www.ambiente.gob.ec/wp-content/uploads/downloads/2012/10/Acuerdo-50-NCA.pdf>
- [51] M. Rodriguez, Z. Cahua, and M. Choque, "Diagnóstico Ambiental," pp. 1–457, 2020, [Online]. Available: <https://fdocuments.mx/document/gobierno-regional-de-puno-siarminamgob-gerencia-regional-de-recursos-naturales.html?page=1>
- [52] M. Girbau Garcia, "La Contaminación Del Agua," *Ciencias la Tierra y del Medio Ambient.*, pp. 1–5, 2012, [Online]. Available: <http://www1.ceit.es/asignaturas/ecologia/Hipertexto/00General/IndiceGral.html>
- [53] J. Ortega *et al.*, "Threats, Challenges And Opportunities For Paediatric Environmental Health In Europe, Latin America And The Caribbean," *An. Pediatr.*, vol. 90, no. 2, pp. 124.e1-124.e11, 2018, doi: 10.1016/j.anpedi.2018.11.015.
- [54] J. A. Ortega-García *et al.*, "Amenazas, desafíos y oportunidades para la salud medioambiental pediátrica en Europa, América Latina y el Caribe," *An. Pediatría*, vol. 90, no. 2, pp. 124.e1-124.e11, 2018, doi: 10.1016/j.anpedi.2018.11.015.
- [55] M. Pereyra San Martín, "Decreto No. 3516 (Vigencia Y Aplicabilidad Del Texto Unificado De Legislación Secundaria Del Ministerio Del Ambiente)," vol. 3516, no. 3516, 2014.
- [56] R. Sataloff, M. Johns, and K. Kost, "Norma Tecnica De Suelo," 2017, [Online]. Available: [https://www.cip.org.ec/attachments/article/1357/NORMA SUELO.pdf](https://www.cip.org.ec/attachments/article/1357/NORMA%20SUELO.pdf)
- [57] Ministerio del Ambiente., "Norma De Calidad Ambiental Y De Descarga De Efluentes : Recurso Agua," *Texto unificado Legis. Secund. del Minist. del Ambient.*, pp. 8–9, 2011.
- [58] L. ambiental del Ecuador, "Normativa Tecnica Del Suelo," *J. Phys. A Math. Theor.*, vol. 44, no. 8, p. 13, 2013, doi: 10.1088/1751-8113/44/8/085201.
- [59] H. Portal Uipan, D. Quispe Alcca, and L. Camacho Colan, "Implementación De Business Intelligence Para Mejorar El Proceso De Toma De Decisiones En El Área De Solucionaes De La Empresa Telefónica Del Perú S.A.A.," pp. 1–145, 2018, [Online]. Available: <http://repositorio.autonoma.edu.pe/handle/AUTONOMA/515>

- [60] E. Naranjo and I. Sáez, “Pentaho: Software Líder De Inteligencia De Negocio De Código Abierto,” *Telemática*, vol. 10, no. 2, 2011.
- [61] D. Dueñas Bustinza, “Optimización En El Proceso De Limpieza De Datos Para La Construcción De Un Datawarehouse Usando Herramientas Open Source,” pp. 2–55, 2018, [Online]. Available: <https://1library.co/document/y9g2m1lq-optimizacion-proceso-limpieza-construccion-datawarehouse-usando-herramientas-source.html>
- [62] S. Sabina Durá, “Sistemas De Información Bi: Estado Actual Y Herramientas De Software Libre,” pp. 1–19, 2011, [Online]. Available: <http://openaccess.uoc.edu/webapps/o2/handle/10609/8175>
- [63] D. Tarnaveanu, “Pentaho Business Analytics A Business Intelligence Open Source Alternative,” *Database Syst. J.*, vol. 3, no. 3, pp. 23–34, 2012.
- [64] N. Leite, I. Pedrosa, and J. Bernardino, “Open Source Business Intelligence On A Sme: A Case Study Using Pentaho,” *Iber. Conf. Inf. Syst. Technol. Cist.*, vol. 2019-June, no. June, pp. 1–7, 2019, doi: 10.23919/CISTI.2019.8760740.
- [65] R. Salaki, J. Waworuntu, and I. Tangkawarow, “Extract Transformation Loading From Oltp To Olap Data Using Pentaho Data Integration,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 128, no. 1, pp. 0–8, 2016, doi: 10.1088/1757-899X/128/1/012020.
- [66] V. Parra, A. Mohammad, A. Syed, and M. Halgamuge, “Pentaho And Jaspersoft: A Comparative Study Of Business Intelligence Open Source Tools Processing Big Data To Evaluate Performances,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 10, pp. 20–29, 2016, doi: 10.14569/ijacsa.2016.071003.
- [67] D. Yaguachi and J. Jumbo, “Diseño De Un Prototipo De Un Data Warehouse Para Una Empresa De Medicina Prepagada,” *Ing. EN Sist. INFORMÁTICOS*, 2019.
- [68] R. Bouman and J. Van Dongen, *Pentaho Solutions: Business Intelligence And Data Warehousing With Pentaho And Mysql*. 2009.
- [69] C. Toma, A. Alexandru, M. Popa, and A. Zamfiroiu, “Iot Solution For Smart Cities’ Pollution Monitoring And The Security Challenges,” *Sensors (Switzerland)*, vol. 19, no. 15, 2019, doi: 10.3390/s19153401.
- [70] N. Desai and A. Sahaya, “Iot Based Air Pollution Monitoring And Predictor System On Beagle Bone Black,” *2017 Int. Conf. Nextgen Electron. Technol. Silicon to Software, ICNETS2 2017*, pp. 367–370, 2017, doi: 10.1109/ICNETS2.2017.8067962.
- [71] R. Khot and V. Chitre, “Survey On Air Pollution Monitoring Systems,” *Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIECS 2017*, vol. 2018-Janua, pp. 1–4, 2018, doi: 10.1109/ICIECS.2017.8275846.
- [72] M. Haghparast, S. Hosseini, and J. Ghodousi, “Comprehensive Environmental Monitoring Based On Stations Of Environmental Pollutants ( Air , Water And Soil ) In Tehran,” vol. 4, no. 4, pp. 263–279, 2020, doi: 10.22097/eeer.2020.212750.1128.
- [73] C. Medrano Huayanay and W. Auccahuasi Aiquipa, “Implementación De Power Bi Para El Análisis De Información En La Productividad En El Laboratorio Clínico Del Hospital Central De La Fuerza Aérea Del Perú De Lima,” *Ing. Sist. E*

*Informática*, vol. 44, no. 8, p. 287, 2017, doi: 10.1088/1751-8113/44/8/085201.

- [74] E. Baralis, T. Cerquitelli, S. Chiusano, P. Garza, and M. Kavosifaris, “Analyzing Air Pollution On The Urban Environment,” *2016 39th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2016 - Proc.*, pp. 1464–1469, 2016, doi: 10.1109/MIPRO.2016.7522370.
- [75] Asamblea Nacional del Ecuador, “Ley Organica De Salud - Ecuador,” *Plataforma Prof. Investig. Jurídica*, p. 13, 2015, [Online]. Available: [https://www.salud.gob.ec/wp-content/uploads/2017/03/Ley Organica De Salud - Ecuador.pdf](https://www.salud.gob.ec/wp-content/uploads/2017/03/Ley%20Organica%20De%20Salud%20-%20Ecuador.pdf)
- [76] Ministerio de Salud Pública, “Codigo Organico De La Salud.” p. 172, 2016.
- [77] J. Calero, “Art.1 De La Constitucion Del Ecuador (Análisis) | Análisis De La Constitución De La República Del Ecuador,” *Iusrectusecart*, no. 449, pp. 1–219, 2015, [Online]. Available: [http://www.corporacionlideres.com/wp-content/uploads/2019/07/NormasSSO\\_Constitucion.pdf%0Ahttp://iusrectusecart.blogspot.com/2015/11/art1-de-la-constitucion-del-ecuador.html](http://www.corporacionlideres.com/wp-content/uploads/2019/07/NormasSSO_Constitucion.pdf%0Ahttp://iusrectusecart.blogspot.com/2015/11/art1-de-la-constitucion-del-ecuador.html)
- [78] Consejo de Educación Superior, “Ley Organica De Educacion Superior,” 2018, pp. 1–58, 2018.
- [79] I. Conference and Surfaces, “Libro De Resúmenes Book Of Abstracts,” *Vacuum*, pp. 21–23, 2016.
- [80] A. Federal, “Gaceta Oficial Del Distrito Federal,” no. 6, 2010.
- [81] J. Carquin, P. Anton de los Santos, R. Rengifo, G. Tejada, and J. Bazan, “Impacto De Un Prototipo De Inteligencia De Negocio Y Su Incidencia En El Procedimiento De Toma Decisiones Empresa - Yupris S . A,” 2020, [Online]. Available: [https://www.lareferencia.info/vufind/Record/PE\\_be9f9a02cf25ebc14a913c136f2d4fd7](https://www.lareferencia.info/vufind/Record/PE_be9f9a02cf25ebc14a913c136f2d4fd7)
- [82] J. Curto Díaz and J. Conesa Caralt, *Introducción Al Business Intelligence*, vol. 163. 2011. [Online]. Available: [http://uoc.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwfV1LSwMxEB6kPfg4qKj4qLA3T1u2m0m6C57sQxcF8QmeynQnKQVpxbr07P\\_zR5npYy0eCr mFJCQkk28-Zr4BUHE9Cv\\_ZBBEJbLlzRrsmaU5co6\\_ZptikGB3TjJbs4u1d8nCVvK6o5kyndR4OpHzGkD5KXmZmwUn8dvTWt-ohspJovsf0bXmVJAMUI8Rt2CzH7\\_y](http://uoc.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwfV1LSwMxEB6kPfg4qKj4qLA3T1u2m0m6C57sQxcF8QmeynQnKQVpxbr07P_zR5npYy0eCr mFJCQkk28-Zr4BUHE9Cv_ZBBEJbLlzRrsmaU5co6_ZptikGB3TjJbs4u1d8nCVvK6o5kyndR4OpHzGkD5KXmZmwUn8dvTWt-ohspJovsf0bXmVJAMUI8Rt2CzH7_y)
- [83] R. Intriago Farías and J. Velazco Vargas, “Análisis De La Pesca Artesanal Del Camarón Pomada En Las Costas De Limones Y Camarones En El Norte De Esmeraldas,” *Gestión Ambient. Tesis*, p. 58, 2018, [Online]. Available: [https://repositorio.pucese.edu.ec/bitstream/123456789/1477/1/FARÍAS INTRIAGO RAIXA ANNABELLA.pdf](https://repositorio.pucese.edu.ec/bitstream/123456789/1477/1/FARÍAS%20INTRIAGO%20RAIXA%20ANNABELLA.pdf)

# **Anexos**

## Anexo 1. Ficha de observación de la Herramienta Suites de Pentaho

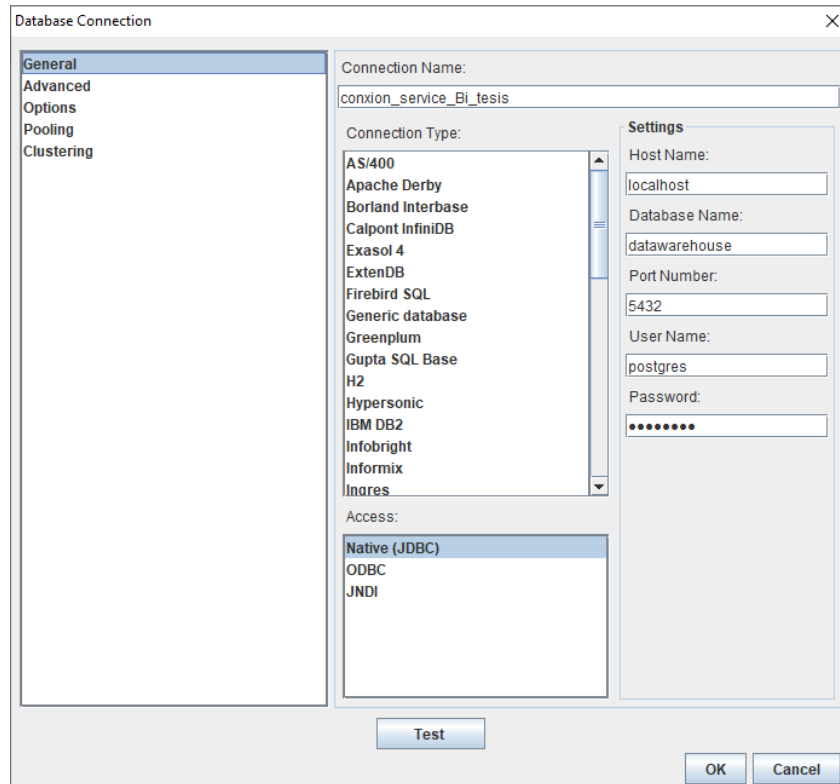
<b>Observación del funcionamiento de la plataforma Suites de pentaho</b>			
Nombre de la plataforma:			
Empresa propietaria:			
Fecha:			
<b>Característica</b>	<b>Alto</b>	<b>Medio</b>	<b>Malo</b>
Plataforma de análisis			
Integración de datos			
accesibilidad a la información.			
Aporta una seguridad extra a la información			
Toma de decisiones			
Diseño de reportes			

## Anexo 2. Ficha de preguntas de instrumento de la entrevista

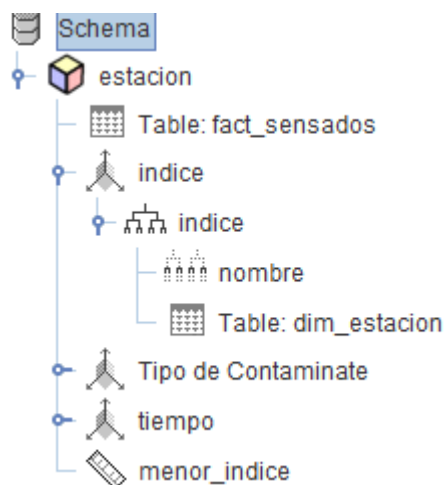
<b>Lugar de la entrevista:</b>		<b>Fecha de realización:</b>	
<b>Nombre del entrevistado:</b>			
<b>Rol en su comunidad:</b>			
¿Conoce usted cuál es la normativa se emplea actualmente en Ecuador para evaluar la contaminación del aire, agua y suelo?			
¿Tiene conocimiento de la existencia de datos sobre contaminación del aire, agua y suelo de la ciudad Esmeraldas?			
¿Cuáles son los tipos de parámetros que se utilizan para describir los contaminantes aire, agua y suelo? ¿cuáles de ellos son prioritarios en un sistema de monitoreo para una ciudad?			
A su criterio, ¿cómo deberían presentarse los datos de contaminación del aire, agua y suelo para la toma de decisiones?			
¿Qué información considera necesaria para la toma de decisiones en el ámbito ambiental, específicamente en la contaminación del aire, agua y suelo?			

### Anexo 3. Tutorial de diseño de cubo de Olap y Reporte

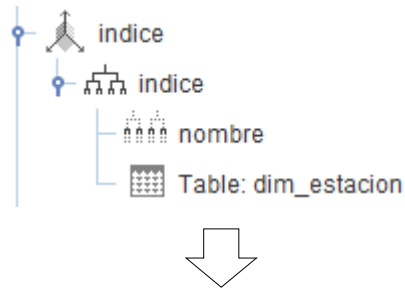
Para la creación del cubo Olap se debe establecer una conexión con la base de datos del Datawarehouse. Para obtener la información necesaria que responderá las diferentes preguntas de negocio.



El diseño del cubo se realiza de forma jerárquica, el esquema (SCHEMA) inicial es un nombre, una descripción y tiene la función de contenedor de cubos, donde el cubo estación se define la tabla de hechos fact\_sensados.



Puede ver la dimensión de índice, tiene una jerarquía de índice de dimensión con una clave principal para el nombre y pk\_estacion, también tiene el nivel nombre que está relacionado con la dimensión estación para obtener la columna nombre, y al final una tabla dim\_estacion.



Schema - pregunta\_1 (Schema1.xml)

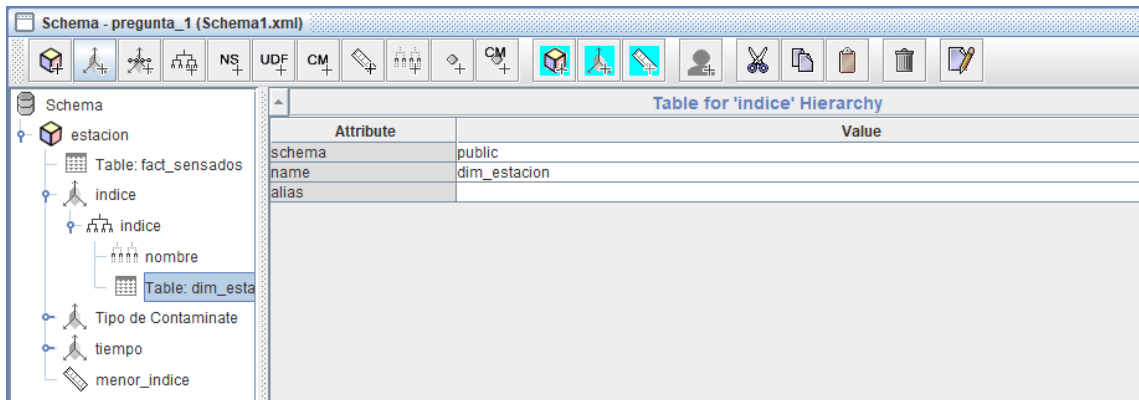
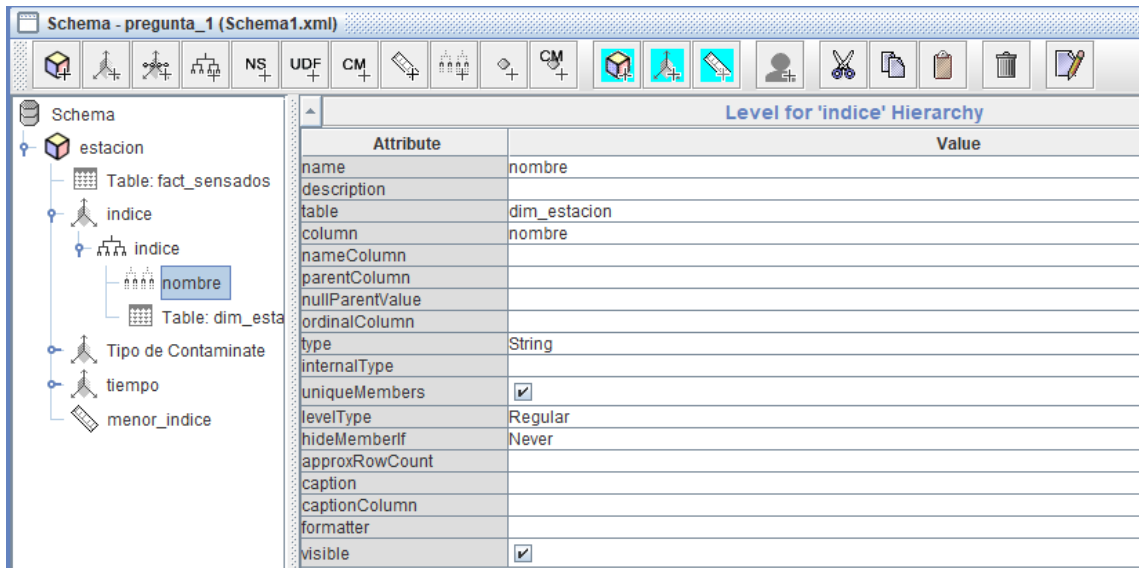
Dimension for 'estacion' Cube

Attribute	Value
name	indice
description	
foreignKey	sk_estacion
type	StandardDimension
usagePrefix	
caption	
visible	<input checked="" type="checkbox"/>

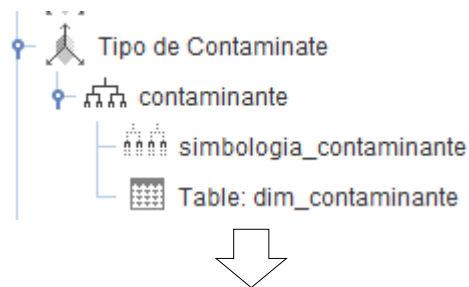
Schema - pregunta\_1 (Schema1.xml)

Hierarchy for 'indice' Dimension

Attribute	Value
name	indice
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	
allMemberCaption	
allLevelName	
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	pk_estacion
caption	
visible	<input checked="" type="checkbox"/>



Puede ver la dimensión de Tiempo de Contaminante, tiene una jerarquía de contaminante de dimensión con una clave principal para el nombre y pk\_contaminante, también tiene el nivel simbología\_contaminante que está relacionado con la dimensión simbologia para obtener la columna simbologia, y al final una tabla dim\_contaminante.



Schema - pregunta\_1 (Schema1.xml)

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminante
  - contaminante
    - simbologia\_contaminante
    - Table: dim\_contaminante
- tiempo
- menor\_indice

Attribute	Value
name	Tipo de Contaminante
description	
foreignKey	sk_sensor
type	StandardDimension
usagePrefix	
caption	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminante
  - contaminante
    - simbologia\_contaminante
    - Table: dim\_contaminante
- tiempo
- menor\_indice

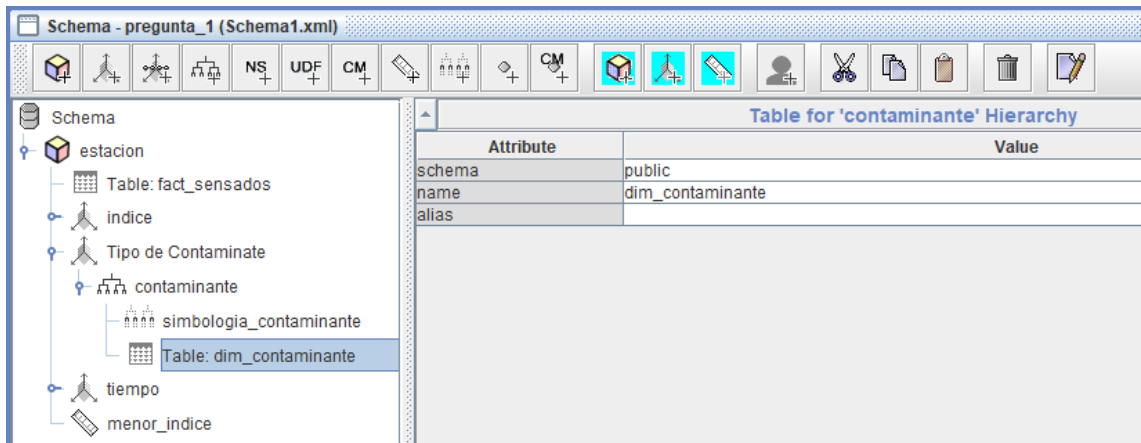
Attribute	Value
name	contaminante
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	
allMemberCaption	
allLevelName	
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	sk_contaminante
caption	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminante
  - contaminante
    - simbologia\_contaminante
    - Table: dim\_contaminante
- tiempo
- menor\_indice

Attribute	Value
name	simbologia_contaminante
description	
table	
column	simbologia
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	String
internalType	
uniqueMembers	<input type="checkbox"/>
levelType	Regular
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>



Puede ver la dimensión de Tiempo, tiene una jerarquía de tiempo de dimensión con una clave principal para el nombre y pk\_tiempo, también tiene el nivel año, día, trimestre, mes, semana que está relacionado con la dimensión tiempo para obtener las columnas año, día, trimestre, mes, semana y a la tabla dim\_tiempo. Para finalizar la parte de medida menor\_indice tiene la función de obtener el mínimo valor en forma de numero con la columna valor\_sensado.



Schema - pregunta\_1 (Schema1.xml)

Dimension for 'estacion' Cube

Attribute	Value
name	tiempo
description	
foreignKey	sk_fecha
type	TimeDimension
usagePrefix	
caption	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)

Hierarchy for 'tiempo' Dimension

Attribute	Value
name	tiempo
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	
allMemberCaption	
allLevelName	
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	sk_fecha
caption	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)

Level for 'tiempo' Hierarchy

Attribute	Value
name	anio
description	
table	dim_tiempo
column	anio
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	Integer
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	TimeYears
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminate
- tiempo
  - tiempo
    - anio
    - dia**
    - trimestre
    - mes
    - semana
  - Table: dim\_tiem
- menor\_indice

Attribute	Value
name	dia
description	
table	dim_tiempo
column	dia
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	Integer
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	TimeDays
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminate
- tiempo
  - tiempo
    - anio
    - dia
    - trimestre**
    - mes
    - semana
  - Table: dim\_tiem
- menor\_indice

Attribute	Value
name	trimestre
description	
table	dim_tiempo
column	trimestre
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	Integer
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	TimeQuarters
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminate
- tiempo
  - tiempo
    - anio
    - dia
    - trimestre
    - mes**
    - semana
  - Table: dim\_tiem
- menor\_indice

Attribute	Value
name	mes
description	
table	dim_tiempo
column	mes
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	Integer
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	TimeMonths
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminate
- tiempo
  - tiempo
    - anio
    - dia
    - trimestre
    - mes
    - semana
  - Table: dim\_tiem
- menor\_indice

Level for 'tiempo' Hierarchy	
Attribute	Value
name	semana
description	
table	dim_tiem
column	semestre
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	String
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	TimeWeeks
hideMemberif	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Schema - pregunta\_1 (Schema1.xml)\*

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminate
- tiempo
  - tiempo
    - anio
    - dia
    - trimestre
    - mes
    - semana
  - Table: dim\_tiem
- menor\_indice

Table for 'tiempo' Hierarchy	
Attribute	Value
schema	public
name	dim_tiem
alias	

Schema - pregunta\_1 (Schema1.xml)\*

estacion

- Table: fact\_sensados
- indice
- Tipo de Contaminate
- tiempo
  - tiempo
    - anio
    - dia
    - trimestre
    - mes
    - semana
  - Table: dim\_tiem
- menor\_indice

Measure for 'estacion' Cube	
Attribute	Value
name	menor_indice
description	
aggregator	min
column	valor_sensado
formatString	
datatype	Numeric
formatter	
caption	
visible	<input checked="" type="checkbox"/>

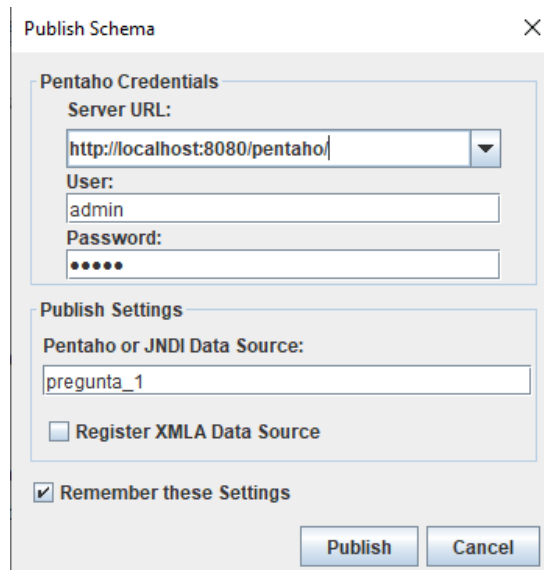
La segunda forma de obtener un cubo es en formato XML, que es un lenguaje de marcado que almacena datos de forma legible.

```

<Cube name="estacion" visible="true" cache="true" enabled="true">
  <Table name="fact_sensados" schema="public">
  </Table>
  <Dimension type="StandardDimension" visible="true" foreignKey="sk_estacion" highCardinality="false"
name="indice">
  <Hierarchy name="indice" visible="true" hasAll="true" primaryKey="pk_estacion">
    <Table name="dim_estacion" schema="public">
    </Table>
    <Level name="nombre" visible="true" table="dim_estacion" column="nombre" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>
  <Dimension type="StandardDimension" visible="true" foreignKey="sk_sensor" highCardinality="false"
name="Tipo de Contaminate">
  <Hierarchy name="contaminante" visible="true" hasAll="true" primaryKey="sk_contaminante">
    <Table name="dim_contaminante" schema="public">
    </Table>
    <Level name="simbologia_contaminante" visible="true" column="simbologia" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>
  <Dimension type="TimeDimension" visible="true" foreignKey="sk_fecha" highCardinality="false"
name="tiempo">
  <Hierarchy name="tiempo" visible="true" hasAll="true" primaryKey="sk_fecha">
    <Table name="dim_tiempo" schema="public">
    </Table>
    <Level name="anio" visible="true" table="dim_tiempo" column="anio" type="Integer"
uniqueMembers="true" levelType="TimeYears" hideMemberIf="Never">
    </Level>
    <Level name="dia" visible="true" table="dim_tiempo" column="dia" type="Integer" uniqueMembers="true"
levelType="TimeDays" hideMemberIf="Never">
    </Level>
    <Level name="trimestre" visible="true" table="dim_tiempo" column="trimestre" type="Integer"
uniqueMembers="true" levelType="TimeQuarters" hideMemberIf="Never">
    </Level>
    <Level name="mes" visible="true" table="dim_tiempo" column="mes" type="Integer"
uniqueMembers="true" levelType="TimeMonths" hideMemberIf="Never">
    </Level>
    <Level name="semana" visible="true" table="dim_tiempo" column="semestre" type="String"
uniqueMembers="false" levelType="TimeWeeks" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>
  <Measure name="menor_indice" column="valor_sensado" datatype="Numeric" aggregator="min" visible="true">
  </Measure>
</Cube>

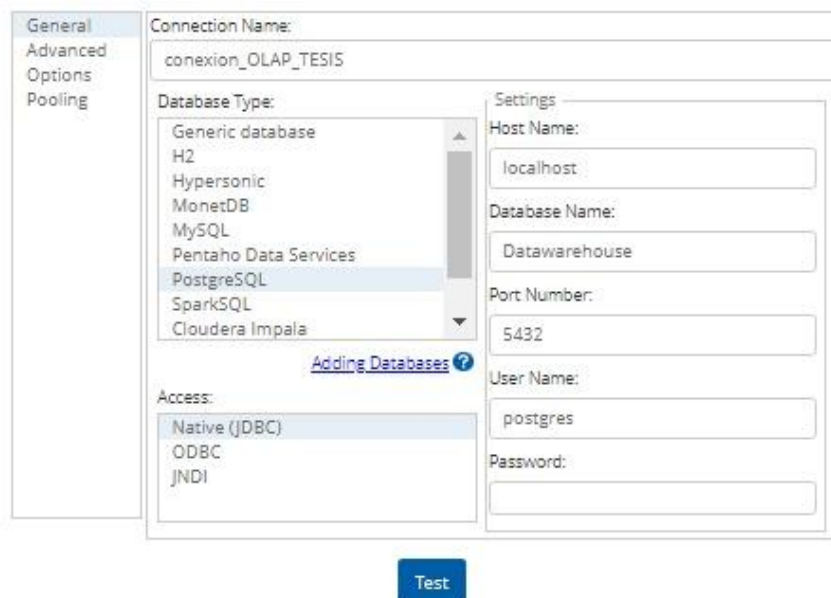
```

La conexión al servidor pentaho se la hace en el apartado que dice publicación de esa forma se puede publicar los cubos de cada pregunta de negocio.

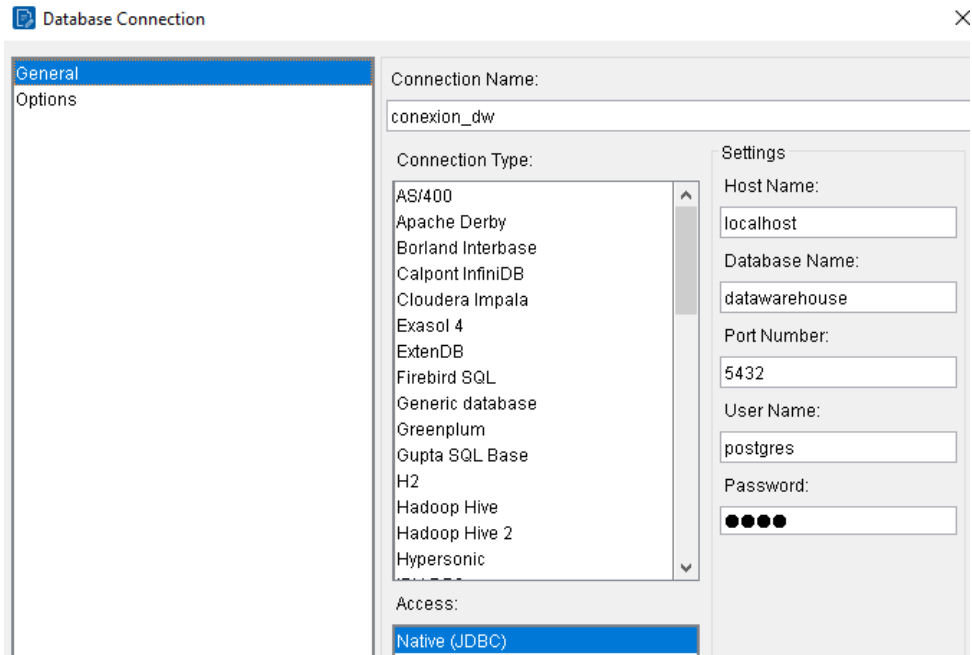


Finalizando la parte de publicación se realiza una nueva conexión de base de datos, pero esta vez sería del servidor de pentaho hacia la base de datos Datawarehouse para poder realizar el despliegue de los cubos y Dashboard esta parte de aquí lo puede ver en los resultados.

## Database Connection



Para crear y diseñar los reportes se realiza una conexión de la herramienta de Report hacia la base de datos Datawarehouse.



Para diseñar los reportes se debe realizar una consulta por cada reporte. y se puede visualizar como finalizaría el diseño del reporte con diferentes graficas como pastel, barra, lineal entre otro.

```

SELECT
dest.nombre,
MIN(fsensados.valor_sensado) AS indice, tiem.anio
FROM fact_sensados as fsensados
INNER JOIN dim_estacion as dest ON
dest.pk_estacion=fsensados.sk_estacion
INNER JOIN dim_sensor as sen ON
sen.pk_sensor=fsensados.sk_sensor
INNER JOIN dim_parametro as para ON
para.pk_parametro=sen.fk_parametro
INNER JOIN dim_contaminante as con ON
con.pk_contaminante=para.fk_contaminante
INNER JOIN dim_tiempo as tiem ON
tiem.sk_fecha=fsensados.sk_fecha
GROUP BY
dest.nombre, fsensados.valor_sensado, tiem.anio

```

