

BUSINESS INTELLIGENCE

ANEXO 2

Marzo 2021



BUSINESS INTELLIGENCE

Anexo 2

Bogotá D.C
Marzo de 2021

Derechos reservados ©
Oficina de las Naciones Unidas Contra la Droga y el Delito –UNODC-

Las opiniones expresadas en el documento no comprometen ni vinculan a ninguna de las instituciones que hacen parte del proyecto, y sus contenidos son puntos de vista expresados por sus autores que no tienen otro fin que servir a manera de recomendaciones, por lo cual no se entiende de ninguna forma que sean vinculantes.

Las denominaciones empleadas en este documento y la forma en que aparecen presentados los datos no implican, de parte de la Oficina de las Naciones Unidas contra la Droga y el Delito – UNODC -, juicio alguno sobre las acciones internas del Estado colombiano.

La información sobre los localizadores uniformes de recursos y enlaces a sitios de Internet contenidos en el documento se consigna para facilitar la consulta y es exacta al tiempo de la elaboración. Las Naciones Unidas no asumen ninguna responsabilidad por la exactitud de dicha información en el futuro ni por el contenido de sitios web externos. Además, la inclusión de referencias, vínculos o enlaces a otros documentos o sitios web no tiene la intención de asignar importancia a esos autores o sitios, y la información contenida en ellos, no tiene por objeto respaldar, recomendar o favorecer a las opiniones expresadas, o productos comerciales o servicios ofrecidos en estos sitios externos, o las organizaciones que patrocinan los sitios, por nombre comercial, marca comercial, fabricación, o de otra manera.

2

Los puntos de vista y opiniones expresados en este documento no representan ni reflejan necesariamente las de las Naciones Unidas o de la UNODC, y no serán utilizados con fines publicitarios o de patrocinio de personas o productos.

La información que se relaciona a continuación es confidencial, está sujeta a reserva hasta que las partes realicen acuerdos explícitos sobre aquello que se hace público y la manera de hacerlo.

Documento elaborado en el marco del Proyecto “Abordaje de los motores criminales de la deforestación en las áreas más afectadas por el fenómeno en Colombia” apoyado por la Embajada Británica en Colombia y desarrollado por la Oficina de Naciones Unidas contra la Droga y el Delito – UNODC- Colombia (2019-2021).





BUSINESS INTELLIGENCE

Anexo 2

Oficina de las Naciones Unidas contra la Droga y el Delito

Elaborado por:

Rubén Darío Aristizábal Ordóñez
Asesor Especializado de Información

Revisado y aprobado por:

Diego Ricardo Quintero Martínez
Coordinador del Pilar I de Seguridad, Reforma Penitenciaria, Prevención y Justicia Ambiental

Anamaría Talero Pilonieta
Coordinadora del Programa de Prevención y Justicia Ambiental

Marzo 2021
Bogotá D.C. – Colombia



Contenido

SISTEMA GERENCIAL DEL CONALDEF EN EL MARCO DE CONCEPTO DE INTELIGENCIA EMPRESARIAL (Business Intelligence)	8
1 CONCEPTOS SOBRE BUSINESS INTELLIGENCE	9
1.1 ¿Qué es Business Intelligence?.....	9
1.2 Datawarehouse.....	12
1.3 Datamart.	15
1.4 Cubo.	16
1.5 Minería de datos. Datamining.	16
1.6 Extracción, Transformación y Carga (ETL).	17



SISTEMA GERENCIAL DEL CONALDEF EN EL MARCO DE CONCEPTO DE INTELIGENCIA EMPRESARIAL (*Business Intelligence*)¹

La necesidad de contar con información clave para tomar decisiones se ha vuelto importante para las diferentes instituciones, sin embargo, muchas veces éstas no cuentan con la información estructurada, homologada, estandarizada, entre otros; o ésta se encuentra dispersa y en diferentes formatos, sean ellos: bases de datos, archivos de texto, hojas de cálculo, etc. lo que impide tomar decisiones oportunas y técnicamente sustentadas.

Por otra parte, los sistemas de información de procesamiento de transacciones en línea (OLTP - OnLine Transaction Processing, por su sigla en inglés) con que cuentan muchas instituciones no son lo suficientemente flexibles para poder contar con información precisa y de forma rápida para tomar decisiones, aunque su diseño permita el manejo de los datos, no permite obtener la información de éstos, y mucho menos inferir el conocimiento almacenado en el día a día de las bases de datos. Las principales características que limitan estos sistemas son:

1. Gran rigidez a la hora de extraer datos, de manera que el usuario tiene que ceñirse a los informes predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.
2. Necesidad de conocimientos técnicos. Para la generación de nuevos informes o métricas suele resultar ineludible acudir al departamento técnico, solicitando una consulta adecuada para interrogar la base de datos.
3. Largos tiempos de respuesta, ya que las consultas complejas de datos suelen implicar la unión de tablas operacionales de gran tamaño, lo que se traduce en una incómoda espera que dificulta la fluidez del trabajo.
4. Deterioro en el rendimiento del Sistema de Información. Cuando la base de datos consultada, para generar informes o ratios de negocio, es la misma que la que soporta el operativo de la empresa, el funcionamiento del sistema puede degradarse hasta afectar y paralizar a todos los usuarios conectados.
5. Falta de integración que implica islas de datos. Muchas organizaciones disponen de múltiples sistemas de información, incorporados en momentos distintos, para resolver problemáticas diferentes. Sus bases de datos no suelen estar integradas, lo que implica la existencia de islas de información.
6. Datos erróneos, obsoletos o incompletos. El tema de la calidad de los datos siempre es considerado

¹ Business Intelligence: sistema de información sobre indicadores de medición para la valoración de modelos de optimización productiva de la Universitat Politècnica de València. Es la capacidad de transformar los datos recolectados de las diferentes fuentes en información, y se puede definir como el conjunto de metodologías, aplicaciones y tecnologías que permiten recolectar, estandarizar, homologar, categorizar y transformar datos de los sistemas transaccionales e información no estructurada, tales como: archivos de texto, hojas de cálculo, entre otros en información estructurada, para su uso a través de reportes, alertas, procesamiento analítico en línea, OLAP, por sus siglas en inglés, procesamiento analítico espacial en línea, SOLAP por sus siglas en inglés, entre otros para su análisis y generación de conocimiento garantizando así un soporte técnico y oportuno a la toma de decisiones.

como algo importante, pero esta labor nunca se lleva al extremo de garantizar la fiabilidad de la información aportada.

7. Problemas para adecuar la información al cargo del usuario. No se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga acceso a la información que necesita para que su trabajo sea lo más eficiente posible.
8. Ausencia de información histórica. Los datos almacenados en los sistemas operacionales están diseñados para llevar la empresa al día, pero no permiten contrastar la situación actual con una situación retrospectiva de años atrás.

Para superar todas las limitaciones mencionadas anteriormente, se plantea el uso de la tecnología de Business Intelligence (BI) para soportar el módulo de batería de indicadores del Sistema de Gestión Gerencial de CONALDEF, ya que el BI se apoya de un conjunto de herramientas que facilitan la extracción, la depuración, el análisis y el almacenamiento de los datos generados en las instituciones que hacen parte del CONALDEF, con la velocidad adecuada para generar conocimiento y apoyar la toma de decisiones de los tomadores de decisiones y de los demás usuarios de la información. Así mismo, el BI es un proceso interactivo para explorar y analizar información estructurada sobre un área o temática específica (normalmente almacenada en un almacén de datos o datawarehouse), para descubrir tendencias o patrones, correlaciones variables que permitan predecir situaciones, realizar análisis multivariados a partir de los cuales realizar recomendaciones y conclusiones para generar acciones o definir o reorientar políticas públicas.

La solución planteada no busca reemplazar los sistemas existentes en las instituciones, sino que sea complementario, con el fin de optimizar el valor de los sistemas de información.

1 CONCEPTOS SOBRE BUSINESS INTELLIGENCE²

1.1 ¿Qué es Business Intelligence?

Es la capacidad de transformar los datos recolectados de las diferentes fuentes en información, y se puede definir como el conjunto de metodologías, aplicaciones y tecnologías que permiten recolectar, estandarizar, homologar, categorizar y transformar datos de los sistemas transaccionales e información no estructurada, tales como: archivos de texto, hojas de cálculo, entre otros en información estructurada, para su uso a través de reportes, alertas, procesamiento analítico en línea, OLAP, por sus siglas en inglés, procesamiento analítico espacial en línea, SOLAP por sus siglas en inglés, entre otros para su análisis y generación de conocimiento garantizando así un soporte técnico y oportuno a la toma de decisiones.

² Sistema de información sobre indicadores de medición para la valoración de modelos de optimización productiva. Universitat Politècnica de València.

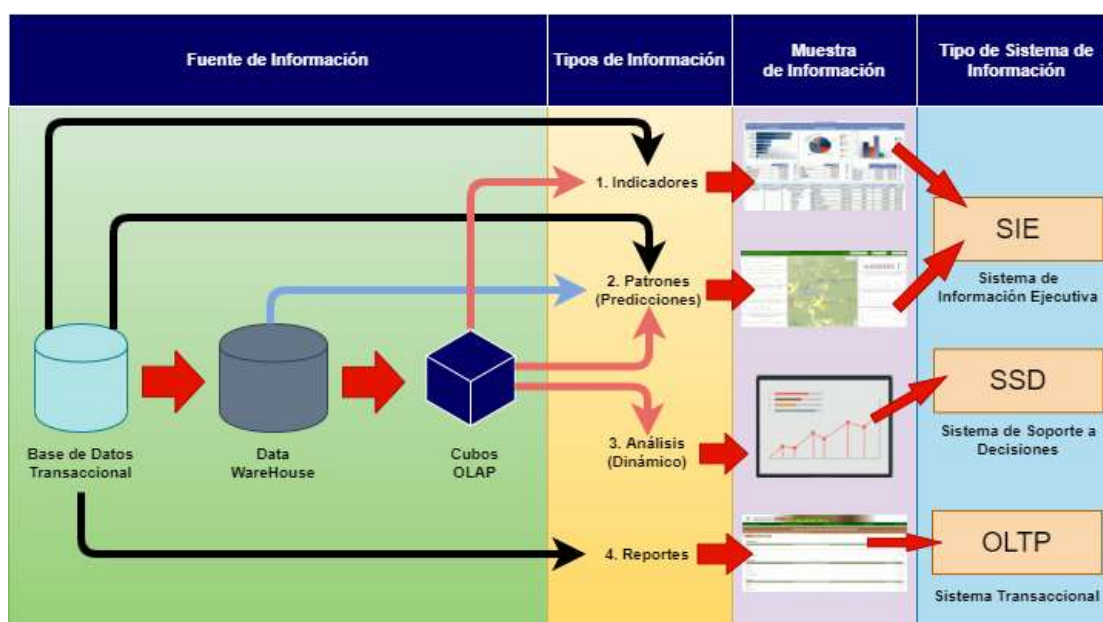


Ilustración 1 Bloques del Sistema de Gestión Gerencial. Fuente: <http://rimenri.blogspot.com/2009/05/caso-de-estudio-inteligencia-de.html>. Elaboración propia

BI soporta entre otros el nivel estratégico o de la institución que lo implemente permitiendo contar con información oportuna para responder a los problemas que se presenten.

Adicional a lo anterior, permite contar en un solo sitio con los datos integrados a través de datamarts o cubos.

Los principales productos del BI son:

- **Cuadros de mandos integral (CMI).** También conocido como Balanced Scorecard (BSC) o dashboard, es una herramienta de control empresarial que permite establecer y monitorizar los objetivos de una empresa y de sus diferentes áreas o unidades.

También se puede considerar como una aplicación que ayuda a una compañía a expresar los objetivos e iniciativas necesarias para cumplir con su estrategia, mostrando de forma continuada cuándo la empresa y los empleados alcanzan los resultados definidos en su plan estratégico. Los beneficios de un CMI son:

1. La fuerza de explicitar un modelo de negocio y traducirlo en indicadores facilita el consenso en toda la empresa, no sólo de la dirección, sino también de cómo alcanzarlo.
2. Clarifica cómo las acciones del día a día afectan no sólo al corto plazo, sino también al largo plazo.
3. Una vez el CMI está en marcha, se puede utilizar para comunicar los planes de la empresa, aunar los esfuerzos en una sola dirección y evitar la dispersión. En este caso, el CMI actúa como un sistema de control por excepción.
4. Permite detectar de forma automática desviaciones en el plan estratégico u operativo, e incluso indagar en los datos operativos de la compañía hasta descubrir la causa original que dio lugar a esas desviaciones.

- **Sistemas de soporte a la Decisión (DSS).** Es una herramienta de BI enfocada al análisis de los datos de una organización.

En principio, puede parecer que el análisis de datos es un proceso sencillo, y fácil de conseguir mediante una aplicación hecha a medida o un ERP sofisticado. Sin embargo, no es así: estas aplicaciones suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas, etc.

El DSS es una de las herramientas más emblemáticas del BI ya que, entre otras propiedades, permiten resolver gran parte de las limitaciones de los programas de gestión. Estas son algunas de sus características principales:

1. Informes dinámicos, flexibles e interactivos, de manera que el usuario no tenga que ceñirse a los listados predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.
 2. No requiere conocimientos técnicos. Un usuario no técnico puede crear nuevos gráficos e informes y navegar entre ellos, haciendo drag&drop o drill through. Por tanto, para examinar la información disponible o crear nuevas métricas no es imprescindible buscar auxilio en el departamento de informática.
 3. Rapidez en el tiempo de respuesta, ya que la base de datos subyacente suele ser un datawarehouse corporativo o un datamart, con modelos de datos en estrella o copo de nieve. Este tipo de bases de datos están optimizadas para el análisis de grandes volúmenes de información.
 4. Integración entre todos los sistemas/departamentos de la compañía. El proceso de extracción, transformación y carga (ETL, por su sigla en inglés) previo a la implantación de un Sistema de Soporte a la Decisión garantiza la calidad y la integración de los datos entre las diferentes unidades de la empresa.
 5. Cada usuario dispone de información adecuada a su perfil. No se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga acceso a la información que necesita para que su trabajo sea lo más eficiente posible.
 6. Disponibilidad de información histórica. En estos sistemas está a la orden del día comparar los datos actuales con información de otros periodos históricos de la compañía, con el fin de analizar tendencias, fijar la evolución de parámetros de negocio, etc.
- **Sistemas de Información Ejecutiva (EIS).** Un Sistema de Información para Ejecutivos o Sistema de Información Ejecutiva es una herramienta software, basada en un DSS, que provee a los gerentes de un acceso sencillo a información interna y externa de su compañía, y que es relevante para sus factores clave de éxito.

La finalidad principal es que el ejecutivo tenga a su disposición un panorama completo del estado de los indicadores de negocio que le afectan al instante, manteniendo también la posibilidad de analizar con detalle aquellos que no estén cumpliendo con las expectativas establecidas, para determinar el plan de acción más adecuado.

De forma más pragmática, se puede definir un EIS como una aplicación informática que muestra

informes y listados de las diferentes áreas de negocio, de forma consolidada, para facilitar la monitorización de la empresa o de una unidad de la misma.

El EIS se caracteriza por ofrecer al ejecutivo un acceso rápido y efectivo a la información compartida, utilizando interfaces gráficas visuales e intuitivas. Suele incluir alertas e informes basados en excepción, así como históricos y análisis de tendencias. También es frecuente que permita la domiciliación por correo de los informes más relevantes.

A través de esta solución se puede contar con un resumen del comportamiento de una organización o área específica, y poder compararla a través del tiempo. Es posible, además, ajustar la visión de la información a la teoría de Balanced Scorecard o Cuadro de Mando Integral impulsada por Norton y Kaplan, o bien a cualquier modelo estratégico de indicadores que maneje la compañía.

1.2 Datawarehouse.

Los sistemas y componentes del BI se diferencian de los sistemas operacionales en que están optimizados para preguntar y divulgar sobre datos. Esto significa que, en un datawarehouse, los datos están desnormalizados para apoyar consultas de alto rendimiento, mientras que en los sistemas operacionales suelen encontrarse normalizados para apoyar operaciones continuas de inserción, modificación y borrado de datos. En este sentido, los procesos ETL (extracción, transformación y carga), que nutren los sistemas BI, tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén completamente integrados.

La solución BI propuesta permite:

- Observar ¿qué está ocurriendo?
- Comprender ¿por qué ocurre?
- Predecir ¿qué ocurriría?
- Colaborar ¿qué debería hacer el equipo?
- Decidir ¿qué camino se debe seguir?

• Modelo Dimensional

Según Loshin (2003), generalmente los almacenes de datos son diseñados y contruidos basándose en un modelo dimensional. Dicho modelo dimensional es considerado una variación del modelo relacional, pero no necesariamente involucra bases de datos relacionales.

Por su parte Kimball (1997), considera que un modelo dimensional es una técnica de diseño lógico que busca presentar los datos de manera estándar e intuitiva, y que, además, permita un acceso de alto rendimiento.

Así, podemos concluir que un modelo dimensional es una variación del modelo relacional, que proporciona al usuario final una forma intuitiva para construir consultas y agilizar tiempos de respuestas.

• Jerarquías

El término jerarquía tiene vital importancia cuando se habla de cubos de información, ya que representan la unión de varios atributos de una dimensión como una serie de relaciones padre-hijo, es decir, de tipo

jerárquico. Por lo tanto, se puede decir que cada elemento en una dimensión se puede resumir utilizando una jerarquía. Por ejemplo: ciudad -> estado -> país -> continente. (Kimball, R. & Ross, M., 2002)

- **Granularidad**

La granularidad se refiere al nivel de detalle con el que se almacenan los datos en una tabla de hechos o dimensiones. Existe una estrecha relación entre jerarquías y granularidad, ya que cuando se habla de jerarquías se está definiendo la granularidad de cada dimensión. El nivel superior de la jerarquía tiene una granularidad alta (grano grueso) y el nivel más bajo de la jerarquía cuenta con una granularidad baja (grano fino), donde se encuentran los datos en detalle (Kimball, R. & Ross, M., 2002).

- **Formas de Modelado.**

Esquema Estrella: Dicho esquema deriva su nombre del hecho que su diagrama forma una estrella, con puntos radiales desde el centro. El centro de la estrella consiste en una o más tablas de hechos y las puntas de la estrella están representadas por las tablas dimensiones (Wolff, 2002).

En el esquema estrella la tabla de hecho es enlazada por cada tabla dimensión por medio de referencia de una clave foránea. Cada una de estas tablas tiene un identificador único, el cual corresponde a la clave foránea en la tabla de hechos.

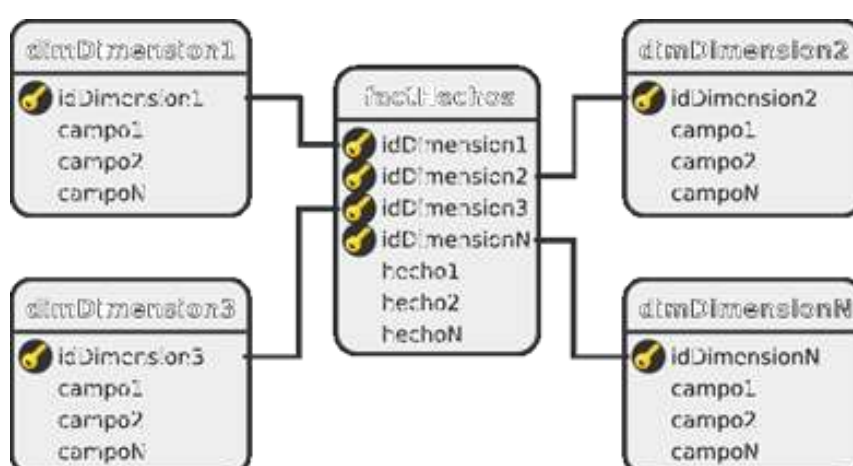


Ilustración 2 Esquema Estrella. Fuente: Dataprix.com, 2009

Esquema Copo de Nieve: La diferencia del esquema copo de nieve con el esquema estrella, se localiza en la estructura de las tablas dimensiones. Las mismas en este esquema se encuentran normalizadas y contienen solo el nivel que es clave primaria en la tabla y la clave foránea de su padre del nivel más cercano (Wolff, 2002).

Es decir, en este esquema las tablas dimensiones, además de estar enlazadas a la tabla de hechos, también se enlazan con otras tablas dimensiones. Así se presenta un copo de nieve.

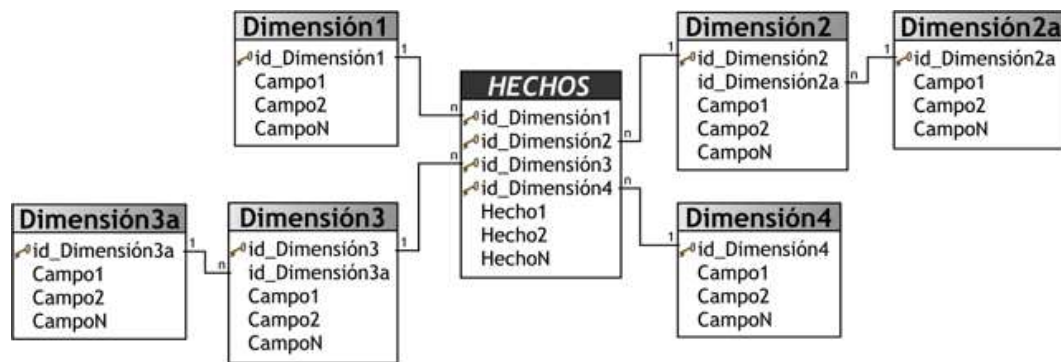


Ilustración 3 Esquema Copo de Nieve. Fuente: Dataprix.com, 2009

Cuando se diseñan bases de datos utilizando el modelo dimensional se definen:

- **Tabla Dimensiones**

Una dimensión se puede definir como una entidad independiente de un modelo dimensional, que agrupa información por categorías y se utiliza como mecanismo de acceso a dicha información. Generalmente, las dimensiones representan el contexto de un evento o hecho y responde a preguntas como qué, quién, dónde y cuándo (Rasmussen, N., Goldy, P., & Solli, P., 2002).

Asimismo, se dice que las dimensiones son cada uno de los ejes en un espacio dimensional. Generalmente esta información es estática en el tiempo y representan las distintas formas de ver la información, en el modelado dimensional se conoce como tablas de dimensiones (Kimball, R. & Ross, M., 2002).

En los almacenes de datos, las dimensiones permiten ver y analizar la información desde diferentes formas y se caracterizan por proporcionar filtrado, agrupación y etiquetado de datos.

Las tablas dimensiones son aquellas que se conectan a la tabla de hechos y permiten visualizar la información desde diferentes perspectivas.

Habitualmente, las tablas dimensiones contiene una clave primaria que la identifica, también conocida como clave subrogada, que es la clave que se usa en el almacén para enlazar con la tabla de hechos, además, contiene atributos descriptivos que se utilizan como filtros de las consultas o para agrupar datos bajo ciertos criterios

Atributos de Dimensiones: Cuando se habla de atributos de dimensiones, se refiere a los atributos de tipo cualitativo que suministra el contexto en el que se obtienen las medidas (Kimball, R. & Ross, M., 2002), como, por ejemplo: ciudad, estado, país, continente, etc.

- **Tabla de Hechos**

Las tablas de hechos son consideradas las tablas centrales dentro de un modelo dimensional, y, además, contienen los valores de las medidas del negocio que se desean analizar. Comúnmente, las tablas de hechos contienen un conjunto de claves foráneas asociadas a las dimensiones, es decir, se expresa una cardinalidad de 'n' a 'n' entre la tabla de hechos y las dimensiones (Wolff, 2002).

Adicionalmente, las tablas de hechos contienen una clave primaria que lo identifica y un conjunto de

indicadores, usados para representar los eventos a modelar, es decir, lo que se busca con la tabla de hechos es representar un proceso de negocio. Además, en ella se almacena información transaccional que varía en el tiempo (Kimball, R. & Ross, M., 2002).

La granularidad es uno de los aspectos más importante al momento de construir una tabla de hechos, ya que permite a los indicadores comparar valores específicos, o bien generalizarlos.

El hecho, o también llamado Fact, es una medida o indicador que representa un evento dentro de una organización. Generalmente, el hecho viene representado en un valor numérico o agregado (Rasmussen, N., Goldy, P., & Solli, P., 2002). Cuando se habla de un valor agregado, se refiere a aquellos hechos que derivan de funciones como la suma, el promedio, el máximo, el mínimo, etc. Por ejemplo, se puede tener un hecho de valor agregado que defina el sueldo máximo de los empleados de la organización.

1.3 Datamart.

Un Datamart es una base de datos sobre una temática particular, especializada en el almacenamiento de los datos de un área temática específica, para el caso: ambiente, control, seguridad o justicia. Los Datamart se caracterizan por contar con una estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho ámbito. Un datamart puede ser alimentado desde los datos de una datawarehouse, o la integración de datos desde distintas fuentes de información.

Por tanto, para crear el datamart de un tema particular es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, por ejemplo, la base de datos del IDEAM, como el propio datawarehouse generado para soportar la toma de decisiones de CONALDEF, o sobre una base de datos OLAP o SOLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada temática.

15

La solución planteada contará con tres (3) tipos de datamarts:

a. **Datamart OLAP.**

Se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o temática, las dimensiones y los indicadores necesarios de cada cubo relacional.

b. **Datamart OLTP**

Pueden basarse en un simple extracto del datawarehouse, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área temática.

c. **Datamart SOLAP**

Se basan en los populares cubos OLAP, pero se integrará el componente espacial a los procesos de agregación de la información, según los requisitos de cada área o temática, las dimensiones y los indicadores de cada cubo relacional.

El componente espacial para el caso serán las coordenadas de una alerta, un área geográfica, tales como:

veredas, municipios, departamentos o una división territorial como una corporación autónoma regional, un área protegidas o zonas estratégicas.

1.4 Cubo.

Son estructuras que permiten analizar rápidamente los datos, en ellos se permite manipular y detallar los datos desde diferentes perspectivas que varían dependiendo de los criterios construidos para su consulta. Principalmente, los cubos están compuestos por hechos, que representan algún evento dentro de la organización, y las dimensiones (Kimball, R. & Ross, M., 2002).

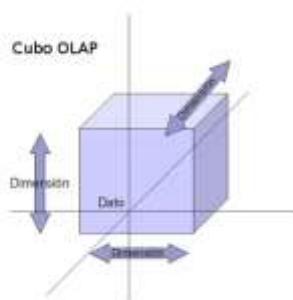


Ilustración 4 Cubo. Fuente: <https://www.tecnologia-informatica.com/sistemas-olap-cubos/>.

1.5 Minería de datos. Datamining.

El datamining (minería de datos), es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Básicamente, el datamining surge para intentar ayudar a comprender el contenido de un repositorio de datos. Con este fin, hace uso de prácticas estadísticas y, en algunos casos, de algoritmos de búsqueda próximos a la Inteligencia Artificial y a las redes neuronales.

Aunque en datamining cada caso concreto puede ser radicalmente distinto al anterior, el proceso común a todos ellos se suele componer de cuatro etapas principales:

- **Determinación de los objetivos.** Trata de la delimitación de los objetivos que el cliente desea bajo la orientación del especialista en datamining.
- **Pre-procesamiento de los datos.** Se refiere a la selección, la limpieza, el enriquecimiento, la reducción y la transformación de las bases de datos. Esta etapa consume generalmente alrededor del setenta por ciento del tiempo total de un proyecto de datamining.
- **Determinación del modelo.** Se comienza realizando unos análisis estadísticos de los datos, y después se lleva a cabo una visualización gráfica de los mismos para tener una primera aproximación. Según los objetivos planteados y la tarea que debe llevarse a cabo, pueden utilizarse algoritmos desarrollados en diferentes áreas de la Inteligencia Artificial.
- **Análisis de los resultados.** Verifica si los resultados obtenidos son coherentes y los coteja con los

obtenidos por los análisis estadísticos y de visualización gráfica. El cliente determina si son novedosos y si le aportan un nuevo conocimiento que le permita considerar sus decisiones.

1.6 Extracción, Transformación y Carga (ETL).

Según el Portal de ETL y Data Warehousing (ETLTools.org, 2009), la extracción, transformación y carga es un proceso usado en las áreas de base de datos, especialmente en los almacenes de datos, con el fin de migrar de un punto a otro. Este proceso consta de tres fases, que se combinan en una sola herramienta, facilitando el transporte de los datos, desde la fuente hasta el almacén. Las tres fases de este proceso se describen a continuación:

Extracción: Es la primera fase del proceso ETL, en esta fase se obtienen los datos provenientes de las diferentes fuentes, sean externas o internas, que serán utilizados por el almacén de datos. Esta fase busca estandarizar los datos, de tal forma que los mismos sean entendibles en la fase de transformación. Además, se incluye un filtrado de datos para eliminar la redundancia o los datos de poco interés.

Transformación: En esta fase se aplican una serie de reglas y funciones a los datos extraídos y que serán cargados posteriormente en el almacén. En algunos casos, se aplican a una serie de transformaciones, como, por ejemplo: ordenación de los datos, estandarización, georreferenciación, unión de los datos desde diferentes fuentes, validación, etc.

Carga: Representa la última fase del proceso ETL, y ésta se encarga de guardar los datos transformados en el almacén. Debido a que en esta fase existe una interacción directa con la base de datos, al momento de la inserción de datos, generalmente se activan los disparadores, o los llamados triggers, y las restricciones existentes.

17

Para poder realizar la transformación estructural se realiza la fase de extracción, transformación y carga de datos (ETL, por su sigla en inglés). En esta etapa se realiza un almacenamiento intermedio, que permite el intercambio entre los sistemas fuente y los sistemas destino (generalmente un datawarehouse), y cuyo principal objetivo es descongestionar los servidores que soportan los sistemas de información de la institución.

La información resultante, ya estandarizada, homologada, georreferenciada, unificada, depurada y consolidada, se almacena en un datawarehouse corporativa, que puede servir como base para la construcción de distintos datamarts. Estos datamarts se caracterizan por poseer la estructura óptima para el análisis de los datos de la temática a abordar, tales como: ambiente, seguridad, supervisión, control y justicia, ya sea mediante bases de datos transaccionales (OLTP) o mediante bases de datos analíticas (OLAP).

A cualquier empresa u organización le beneficia poner en marcha un proceso ETL para mover y transformar los datos que maneja por los siguientes motivos:

- Poder crear un repositorio central estandarizado (Master Data Management, en inglés) de todos los datos de la organización, unificando la información de las diferentes bases de datos en un único repositorio de datos.
- Facilitar a los directivos la toma de decisiones estratégicas basadas en el análisis de los datos cargados en las bases nuevas y actualizadas (datamart o datawarehouse).

- Integrar sistemas. Las organizaciones crecen de forma orgánica y cada vez se van agregando más fuentes de datos. Esto provoca que comience a surgir nuevas necesidades de integración que deben ser atendidas.
- Poder tener una visión global de todos los datos consolidados en un datawarehouse.