

OLAP



Base de Datos II

UNIVERSIDAD
SIGLO 21

MIEMBRO DE LA RED
ILUMNO

» Administración del Modelo

“Las herramientas OLAP permiten a los usuarios finales tratar la información de forma multidimensional para explorarla desde distintas perspectivas y períodos de tiempo.”

(Cano, 2007, p. 132)

Herramientas OLAP

En la parte final de la lectura de la unidad 1 (módulo 1), pudimos analizar las principales características de una herramienta OLAP.

En términos generales, una herramienta OLAP puede ser Cliente/Servidor o Web Enabled:

Cliente / Servidor, lo que significa tener las instalaciones locales en los ordenadores de los usuarios.

Acceso Web: cliente, cliente ligero, o solo con el navegador. En este tipo de acceso el navegador se comunica con un servidor web, el cual habla con la aplicación del servidor, que es la que conecta con el Data Warehouse. En el caso de acceder con el navegador sin ningún tipo de cliente o con cliente ligero (por ejemplo JAVA), normalmente se descargan pequeñas aplicaciones para aumentar la funcionalidad.

Algunos autores también hablan del **Virtual o Desktop OLAP (DOLAP)**. En este caso creamos un cubo con las dimensiones que le interesan al usuario, lo cargamos en memoria en su ordenador, trabaja y, cuando acaba, lo eliminamos de la memoria. La ventaja

es que el usuario sólo recibe los hechos y las dimensiones en los que está interesado y los analiza en forma local.

(Cano, 2007, pp. 130-131)

FASMI

De acuerdo con Josep Lluís Cano (2007, p. 126), “el OLAP Council sumarizó las 12 reglas de Codd en lo que ellos llamaban el concepto FASMI y que los productos OLAP deben cumplir”.

FASMI es una sigla compuesta por los siguientes anglicismos:

- Fast
- Analysis
- Shared
- Multidimensional
- Information

A continuación, vemos lo que representa cada una de ellas:

FAST (Rápido): Debe ser rápido, necesitamos lanzar consultas y ver los resultados inmediatamente.

ANALYSIS (Análisis): Debe soportar la lógica de negocio y análisis estadísticos que sean necesarios para los usuarios.

SHARED (Compartido): Tiene que manejar múltiples actualizaciones de forma segura y rápida.

MULTIDIMENSIONAL (Multidimensional): Tiene que proveer de una visión conceptual de la información a través de distintas dimensiones.

INFORMATION (Información): Debe poder manejar toda la información relevante y la información derivada.

(Cano, 2007, pp. 126-127)

Almacenamiento OLAP

Existen diferentes tipos de herramientas OLAP según su grado de almacenamiento:

- ROLAP
- MOLAP
- HOLAP

Con las siguientes características principales:

ROLAP: Relational OLAP

Las capacidades OLAP acceden directamente a la base de datos relacional. Se accede por tanto a una base de datos relacional (RDBMS). Accede habitualmente sobre un modelo “estrella”. La principal ventaja es que no tiene limitaciones en cuanto al tamaño, pero es más lento que el MOLAP, aunque algunos productos comerciales nos permiten cargar cubos virtuales para acelerar los tiempos de acceso.

MOLAP: Multidimensional OLAP

La implementación OLAP accede directamente sobre una base de datos multidimensional (MDBMS). La ventaja principal de esta alternativa es que es muy rápida en los tiempos de respuesta y la principal desventaja es que, si queremos cambiar las dimensiones, debemos cargar de nuevo el cubo.

HOLAP: Hybrid OLAP

Accede a los datos de alto nivel en una base de datos multidimensional y a los atómicos directamente sobre la base de datos relacional. En esencia, utiliza las ventajas del ROLAP y del MOLAP.

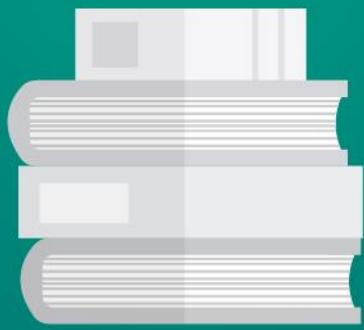
(Cano, 2007, p. 130)



Bibliografía de referencias

- **Cano, J. L.** (2007). *Business Intelligence: Competir con Información*. España: Banesto Fundación Cultural y ESADE.
- **Inmon, W.** (2005). *Building the Data Warehouse* (4º Edición). Estados Unidos: Wiley Publishing.
- **Kimball Ralph, R. M.** (2013). *The Data Warehouse Toolkit* (3º Edición). Estados Unidos: Wiley Publishing.
- **QlikTech** (2010). *QlikView architectural overview* (White Paper). Recuperado el 7 de Mayo de 2014: http://www.swbi.co.uk/pdf/QlikView_Architectural_Overview.pdf
- **QlikTech** (s.f.). *QlikView es pionero del BI en memoria*. Recuperado el 7 de Mayo de 2014: <http://www.qlik.com/es/explore/products/overview>

Proceso de diseño dimensional



Base de Datos II

UNIVERSIDAD
SIGLO 21

MIEMBRO DE LA RED
ILUMNO



Proceso de diseño dimensional

Proceso de Diseño Dimensional en 4 Pasos

De acuerdo con Ralph Kimball (2013), las cuatro decisiones clave que se toman durante el diseño de un modelo multidimensional incluyen los siguientes pasos:

- Seleccionar el proceso de negocio
- Declarar la granularidad
- Identificar las dimensiones
- Identificar los hechos.

Las respuestas a estos interrogantes se determinan considerando las necesidades del negocio junto con las realidades de los datos, fuente subyacentes durante las sesiones de modelado colaborativas.

Luego de los cuatro pasos, el equipo de diseño determina los nombres de tablas y columnas, valores de dominio de muestra y reglas de negocio. Representantes del negocio deben participar en esta actividad detallada para garantizar su compromiso.

1. Procesos de Negocio

Los **procesos de negocio** son las actividades operativas realizadas por una organización, como por ejemplo: tomar una orden o solicitud, procesar una queja o reclamo, registrar estudiantes en un curso, y otros.

Los eventos de un proceso de negocio generan o capturan métricas de desempeño que se traducen en hechos en una tabla de hechos. Muchas tablas de hechos se enfocan en los resultados de un proceso de negocio en particular.

Seleccionar el proceso es importante porque define un objetivo de diseño específico y permite declarar la granularidad, las dimensiones y los hechos.

2. Granularidad

Declarar la **granularidad** es el paso principal en un diseño dimensional. La granularidad establece exactamente qué representa una fila de una tabla de hechos, y debe declararse antes de seleccionar las dimensiones o hechos, ya que cada uno de estos debe ser consistente con la granularidad. Esta consistencia, fuerza a una uniformidad en todos los diseños dimensionales que es crítica para la performance y facilidad de uso de una aplicación BI.

La granularidad atómica se refiere al nivel más bajo, en el cual los datos son capturados por un proceso de negocio dado. Ralph Kimball (2013) recomienda iniciar enfocándose en datos de granularidad atómica porque permite resistir el surgimiento de consultas de usuario impredecibles.

La granularidad sumarizada es buena para la performance, pero presupone las consultas comunes de los usuarios. Cada granularidad de tabla de hechos propuesta resulta en una tabla física separada; diferentes niveles de granularidad no deben mezclarse en la misma tabla de hechos.

3. Dimensiones para la descripción del contexto

Las **dimensiones** proporcionan el contexto: **quién, qué, dónde, cuándo, por qué y cómo**, circundante a un evento de proceso de negocio.

Las tablas de dimensión contienen los atributos descriptivos usados por las aplicaciones BI para filtrar y agrupar los hechos.

Con la granularidad de una tabla de hechos en mente, todas las dimensiones posibles pueden ser identificadas. Cada vez que sea posible, una dimensión debería ser valuada cuando se asocia a una fila de un hecho dado.

Las tablas de dimensión son algunas veces llamadas el "alma" del Data Warehouse porque contienen los puntos de entrada y etiquetas descriptivas que le permiten a un sistema BI habilitar el análisis del negocio.

4. Hechos para medición

Los **hechos** son las mediciones que resultan de un evento de proceso de negocio y son casi siempre numéricos. Una fila de una tabla de hechos tiene una relación uno a uno con un evento de medición, en función de la granularidad de la tabla de hechos.

Por lo tanto, una tabla de hechos corresponde a un evento observable físico y no a la demanda de un reporte particular. Dentro de una tabla de hechos sólo son permitidos los consistentes con la granularidad detallada.

Técnicas básicas para Tablas de Hechos

A continuación, veremos un conjunto de técnicas básicas para las Tablas de Hechos, de acuerdo con los lineamientos de Ralph Kimball (2013).

Estructura de la Tabla de Hechos

Una Tabla de Hechos contiene las mediciones numéricas producidas por un evento de medición operacional en el mundo real.

En el más bajo nivel de granularidad, una fila de tabla de hechos corresponderá a un evento de medición y viceversa. Por lo tanto el diseño fundamental de una tabla de hechos está completamente basado en una actividad física y no está influenciado por los reportes eventuales que puedan producirse.

Además de las métricas numéricas, una Tabla de Hechos siempre contiene claves foráneas (FK) para cada una de sus dimensiones asociadas, así como opcionales claves de dimensiones degeneradas y columnas de fecha/hora.

Estas tablas son el objetivo primario de cálculos y summarizaciones dinámicas en las consultas.

Hechos Aditivos, Semi-Aditivos y No Aditivos

Las métricas numéricas en una tabla de hechos caen dentro de tres categorías:

- Los hechos más flexibles y útiles son totalmente **aditivos**. Las métricas aditivas pueden sumarse a lo largo de cualquiera de las dimensiones asociadas con la tabla de hechos.
- Las métricas **semi-aditivas** pueden sumarse a lo largo de algunas dimensiones, pero no todas; por ejemplo, las cantidades de un balance contable pueden sumarse para todas las dimensiones, excepto el tiempo.
- Finalmente, algunas métricas son completamente **no aditivas**, como por ejemplo, los ratios. Un buen enfoque para los hechos no aditivos es, donde sea posible, almacenar los componentes totalmente aditivos de las métricas no aditivas y sumar estos componentes en la respuesta final antes de calcular el hecho no aditivo. Este cálculo final frecuentemente se realiza en la aplicación BI o cubo OLAP.

Valores NULL en Tablas de Hechos

Las mediciones que tienen valores NULL se comportan correctamente en las Tablas de Hechos. Las funciones de agregación (SUM, COUNT, MIN, MAX y AVG) se comportan bien con hechos que tienen valores NULL.

Sin embargo, los NULL deben evitarse en las claves foráneas de las tablas de hechos porque los mismos causarían automáticamente una violación de la integridad referencial.

En lugar de una clave foránea NULL, las tablas de dimensión asociadas deben tener una fila por default que represente un valor desconocido o una condición no aplicable.

Hechos Ajustados

Si la misma medición aparece en Tablas de Hechos separadas, deben tomarse resguardos para asegurar que las definiciones técnicas de los hechos sean idénticas si son comparadas o computadas conjuntamente.

Si las definiciones de hechos separadas son consistentes, los hechos ajustados deberían ser nombrados idénticamente; pero si son incompatibles, deberían ser nombrados de manera diferente para alertar a los usuarios y aplicaciones BI.

Tablas de Hechos de transacción

Una fila en una Tabla de Hechos de transacción corresponde a un evento de medición en un punto en espacio y tiempo.

Las Tablas de Hechos cuyo nivel de granularidad es de transacción atómica/detallada son las tablas más dimensionales y expresivas. Esta dimensionalidad robusta permite el máximo nivel de *slicing* y *dicing* de datos de las transacciones.

Estas tablas pueden ser o no densas porque las filas solo existen si las mediciones toman lugar. Además, siempre contienen una clave foránea por cada dimensión asociada, y opcionalmente contienen *timestamps* precisos y claves de dimensiones degeneradas.

Los hechos numéricos medidos deben ser consistentes con la granularidad de la transacción.

Tablas de Hechos de Snapshots Periódicos

Una fila en una Tabla de Hechos de *snapshot* periódica sumariza muchos eventos de medición que ocurren sobre un período estándar, como por ejemplo: un día, una semana o un mes. La granularidad es el período, no la transacción individual.

Estas tablas frecuentemente contienen muchos hechos, ya que cualquier evento de medición consistente con la granularidad de la tabla de hechos es permisible. Además, son uniformemente densas en sus claves foráneas porque aún si ninguna actividad toma lugar durante el período, una fila es típicamente insertada en la tabla conteniendo un cero o NULL por cada hecho.

Tablas de Hechos de Snapshots de Acumulación

Una fila en una Tabla de Hechos de *Snapshots* de Acumulación sumaiza los eventos de medición que ocurren en pasos predecibles entre el comienzo y fin de un proceso.

Los procesos de *workflow*, que tienen un punto de inicio determinado, pasos intermedios estándares y punto final definido, pueden ser modelados con este tipo de tablas. Existe una clave foránea de tipo *Fecha* en la Tabla de Hechos por cada hito crítico en el proceso.

Una fila individual en una Tabla de Hechos de Snapshots de Acumulación corresponde, por ejemplo, a una línea en una solicitud o factura (initialmente se inserta cuando la línea se crea). Cada vez que hay un avance en el workflow, la fila de la tabla de hechos se actualiza.

Además de las claves foráneas de fechas asociadas con cada paso crítico del proceso, las Tablas de Hechos de Snapshots de Acumulación contienen claves foráneas para otras dimensiones, y opcionalmente contienen dimensiones degeneradas.

Es frecuente que este tipo de tablas incluya mediciones de retraso numérico consistentes con la granularidad, junto con contadores de completitud de los hitos.

Tablas de Hechos sin Hechos

Aunque muchos eventos de medición capturan resultados numéricos, es posible que el evento solamente registre un conjunto de entidades dimensionales que vienen juntas en un determinado momento.

Por ejemplo, un evento de un estudiante asistiendo a clases en un día dado puede que no tenga un hecho numérico registrado, pero que una fila de hechos con clave foránea para el día calendario, estudiante, profesor, ubicación y clase, esté bien definida. Asimismo, las comunicaciones con los clientes son eventos, pero podría no haber métricas asociadas.

Las tablas de hechos sin hechos pueden también ser usadas para analizar qué no sucedió. Estas consultas siempre tienen dos partes: una tabla sin hechos que contiene todas las posibilidades de eventos que podrían suceder y una tabla de actividades que contiene los eventos que sucedieron. Cuando la actividad se extrae de la primera, el resultado es el conjunto de eventos que no sucedió.

Tablas de Hechos Sumarizadas o Cubos OLAP

Las Tablas de Hechos Sumarizadas son simples *rollups* numéricos de Tablas de Hechos atómicas, construidas para acelerar la performance de las consultas. Estas tablas deberían estar disponibles en la capa BI al mismo tiempo que las Tablas de Hechos atómicas, de modo que las herramientas BI seleccionen el nivel de sumarización apropiado en el tiempo de consulta.

Este proceso, conocido como *aggregate navigation*, debe ser abierto de manera tal que cada herramienta de consulta, reporte y/o aplicación BI - entre otros- tenga los mismos beneficios de performance. Un conjunto apropiadamente diseñado de sumarizaciones debería comportarse como índice de bases de datos que acelere la performance de las consultas.

Las Tablas de Hechos Sumarizadas contienen claves foráneas a dimensiones ajustadas reducidas, así como hechos sumarizados creados sumando métricas de tablas de hechos más atómicas.

Finalmente, los cubos OLAP sumarizados con métricas sumarizadas son frecuentemente construidos de la misma manera que los relacionales, pero los usuarios pueden acceder directamente a los cubos OLAP.

Tablas de Hechos Consolidados

Es frecuentemente conveniente combinar hechos de múltiples procesos juntos en una simple tabla de hechos consolidada si pueden ser expresados a la misma granularidad. Por ejemplo, las ventas actuales pueden consolidarse con las predicciones de ventas en la misma Tabla de Hechos para permitir análisis más simples y rápidos en comparación con ensamblar una aplicación que use tablas separadas.

Estas tablas agregan procesamiento adicional a los ETL, pero también facilitan el análisis en la aplicación BI. Además, dichas tablas deben considerarse para métricas de procesos cruzados que frecuentemente se analizan conjuntamente.

Técnicas Básicas para Tablas de Dimensión

A continuación, veremos un conjunto de técnicas básicas para las tablas de dimensión, de acuerdo con los lineamientos de Ralph Kimball (2013).

Estructura de Tablas de Dimensión

Cada tabla de dimensión tiene una única columna PK. Esta clave primaria está embebida como una FK en cualquier tabla de hechos asociada, donde el contexto descriptivo de la fila de la dimensión es exactamente correcto para esa fila de la Tabla de Hechos.

Las tablas de dimensión generalmente son tablas de-normalizadas con muchos atributos de texto de baja cardinalidad. Estos atributos son el objetivo primario de restricciones y agrupamientos en las consultas y aplicaciones BI.

Claves Subrogadas de Dimensiones

Una tabla de dimensión se diseña con una columna que sirve como única PK; esta no puede ser la clave natural del sistema operacional porque habrá múltiples filas de dimensión para esa clave natural cuando los cambios se sigan en el tiempo. Además, las claves naturales para una dimensión pueden ser creadas por más de un sistema fuente y pueden ser incompatibles o pobemente administradas.

El sistema BI necesita tener control de las PK de todas las dimensiones en lugar de usar claves naturales con fechas agregadas; deberían crearse PK enteras anónimas para cada dimensión. Estas Claves Subrogadas de Dimensiones son números enteros, asignados en secuencia, iniciando con el valor 1.

La dimensión Tiempo está exenta de esta regla, ya que es una dimensión altamente predecible y estable y puede usar una PK más significativa.

Claves naturales, durables y sobrenaturales

Las claves naturales creadas por los sistemas operacionales están sujetas a reglas de negocio fuera del control del sistema BI.

Cuando el Data Warehouse desea tener una única clave, debe crearse una nueva clave durable, persistente y que no cambie. En ocasiones, se hace referencia a la misma como ‘clave sobrenatural durable’. Generalmente son secuencias numéricas iniciando en 1.

Mientras múltiples claves subrogadas pueden estar asociadas, por ejemplo, con un registro de un empleado a lo largo del tiempo, aunque sus datos cambien, la clave durable nunca cambia.

Drill Down

Drill Down es la forma de análisis fundamental. Significa agregar un encabezado de fila a una consulta existente en la que el nuevo encabezado es un atributo de dimensión añadido a la expresión GROUP BY en una consulta SQL. El atributo puede venir de cualquier dimensión adjunta a la Tabla de Hechos, no requiere la definición de jerarquías predeterminadas.

Dimensiones Degeneradas

En algunas ocasiones una dimensión se define porque no tiene contexto, excepto para su PK. Una dimensión degenerada es aquella que no tiene tabla de dimensión asociada. Por otro lado, las Dimensiones Degeneradas son más comunes en las tablas de hechos de snapshot acumulativas y transaccionales.

Dimensiones Aplanadas De-normalizadas

En general, los diseñadores dimensionales deben resistir a la normalización, de-normalizando dimensiones con dos objetivos: simplicidad y velocidad.

Múltiples Jerarquías en Dimensiones

Muchas dimensiones contienen más de una jerarquía natural. En estos casos, las jerarquías separadas pueden coexistir en la misma tabla de dimensión. Por ejemplo: en una dimensión *tiempo*, jerarquía día-semana-mes-año fiscal y otra de día-mes-año.

Banderas e Indicadores como Atributos Textuales

Las abreviaturas crípticas, banderas verdadero/falso y los indicadores operacionales deberían ser suplementados en las tablas de dimensión con palabras de texto completas que tengan significado cuando se visualicen independientemente.

Los códigos operacionales con significado embebido dentro del valor del código deberían ser fragmentados en cada parte del código, expandiéndose en distintos atributos descriptivos de dimensión separados.

Atributos NULL en las Dimensiones

Los atributos de dimensiones con valor NULL resultan cuando una fila de dimensión dada no ha sido totalmente llenada o cuando hay atributos que no son aplicables a todas las filas de la dimensión. En ambos casos, Ralph Kimball (2013) recomienda sustituir un *string* descriptivo, tal como por ejemplo: *Desconocido* o *No Aplica*, en lugar del valor NULL.

Los valores NULL en los atributos de dimensión deberían evitarse porque diferentes bases de datos manejan el agrupamiento o restricciones sobre los NULL de manera inconsistente.

Dimensiones de Fecha Calendario (Tiempo)

Las dimensiones de fecha calendario (Tiempo) están adjuntas virtualmente a cada Tabla de Hechos para permitir la navegación a través de fechas familiares, meses, períodos fiscales y días especiales del calendario.

La dimensión Tiempo generalmente tiene muchos atributos que describen características tales como nº de semana, nombre del mes, período fiscal o vacaciones nacionales.

Para facilitar el particionamiento, la PK de una dimensión *Fecha o Tiempo* puede ser más significativa, como un entero que representa AAAAMMDD, en lugar de dimensión Tiempo necesita una fila especial para representar fechas desconocidas o a ser determinadas. Cuando se necesita mayor precisión, un *timestamp* separado puede agregarse a la tabla de hechos. El timestamp no es una FK relacionada con una tabla de dimensión, sino que es una columna aislada.

Por otra parte, si los usuarios del negocio restringen o agrupan los datos en base a atributos de la hora del día, debería agregarse una FK separada con los rangos horarios a la tabla de hechos.

Dimensiones ‘Role-Playing’

Una única dimensión física puede ser referenciada múltiples veces en una tabla de hechos, con cada referencia enlazada a un rol lógicamente distinto para la dimensión.

Por ejemplo, una tabla de hechos puede tener distintas fechas, cada una de las cuales estará representada por una FK a la dimensión Tiempo.

Es esencial que cada FK se refiera a una vista separada de la dimensión Tiempo, de modo que las referencias sean independientes. Estas vistas de dimensión separadas son llamadas roles.

Dimensiones Inútiles

Los procesos de negocio transaccionales producen normalmente un número de banderas e indicadores misceláneos de baja cardinalidad. En lugar de hacer dimensiones separadas para cada bandera y atributo, puede crearse una única dimensión inútil combinándolas conjuntamente.

Esta dimensión, frecuentemente etiquetada como una dimensión de perfil transacción en un esquema, no necesita ser el producto cartesiano de todos los valores posibles del atributo, sino que debería solo contener la combinación de valores que actualmente están presentes en los datos fuente.

Dimensiones Copos de Nieve

Cuando una relación jerárquica en una tabla de dimensión es normalizada, los atributos de baja cardinalidad aparecen como tablas secundarias conectadas a la tabla de dimensión base por una clave de atributo.

Cuando este proceso es repetido con todas las jerarquías de la tabla de dimensión, una estructura multinivel característica es creada, la cual es llamada copo de nieve.

Aunque el copo de nieve representa los datos jerárquicos precisamente, deberían evitarse porque es difícil para un usuario comprenderlos y navegarlos; a su vez, pueden impactar negativamente en la performance de las consultas.

Una tabla de dimensión de-normalizada aplanada, contiene exactamente la misma información que una dimensión en copo de nieve.

Dimensiones de Refuerzo

Una dimensión puede contener una referencia a otra tabla de dimensión. Por ejemplo, una dimensión de Cuenta Bancaria puede referenciar una dimensión separada que represente la Fecha en la cual se abrió la cuenta.

Estas referencias a dimensiones secundarias se denominan Dimensiones de Refuerzo; si bien están permitidas, deberían usarse con moderación. En muchos casos, las correlaciones entre dimensiones deben llevarse a una tabla de hechos en donde las dimensiones se representan como FK separadas.



Bibliografía de referencias

- **Cano, J. L.** (2007). *Business Intelligence: Competir con Información*. España: Banesto Fundación Cultural y ESADE.
- **Inmon, W.** (2005). *Building the Data Warehouse* (4º Edición). Estados Unidos: Wiley Publishing.
- **Kimball Ralph, R. M.** (2013). *The Data Warehouse Toolkit* (3º Edición). Estados Unidos: Wiley Publishing.
- **QlikTech** (2010). *QlikView architectural overview* (White Paper). Recuperado el 7 de Mayo de 2014: http://www.swbi.co.uk/pdf/QlikView_Architectural_Overview.pdf
- **QlikTech** (s.f.). *QlikView es pionero del BI en memoria*. Recuperado el 7 de Mayo de 2014: <http://www.qlik.com/es/explore/products/overview>

Tipos de usuarios



Bases de Datos II

UNIVERSIDAD
SIGLO 21

MIEMBRO DE LA RED
ILUMNO

» **Tipos de usuarios**

“Existen cuatro tipos de usuarios de una solución de Business Intelligence (BI): Granjeros, Exploradores, Mineros y Turistas”.

(William Inmon, 2005)

Herramientas y Soluciones de BI

Clasificación de William Inmon

De acuerdo con William Inmon (2005), no existe un único tipo de usuario de una solución de BI, sino cuatro tipos distintos, cada uno con sus propias características:

- Granjeros
- Exploradores
- Mineros
- Turistas

Demográficamente, hay más granjeros que otros tipos de usuarios.

Sin embargo, debe aclararse que el mismo usuario puede variar en distintos tipos, comportándose según la ocasión o la tarea.

A continuación, analizaremos las principales características de cada uno de estos cuatro tipos de usuarios.

Granjeros

- Es el tipo de usuario más predominante.
- Es una persona predecible. Hace la misma actividad en forma rutinaria.
- Los tipos de consultas varían solo en el tipo de dato de la consulta. Repite el mismo tipo de consulta siempre.
- Las consultas tienden a ser cortas; dado que sabe lo que quiere, va directamente a donde están los datos.
- Sus consultas tienen un patrón de acceso similar (ejemplo: una vez por semana los lunes a la tarde).
- Tiene un alto ratio de probabilidad de encontrar lo que busca.
- Acceden a los datos activamente usados en el Data Warehouse.
- En el Corporate Information Factory generalmente acceden a Data Marts.

Exploradores

- Es una persona que no sabe lo que quiere. Opera en un modo de completa impredecibilidad. Quizás pasen 6 meses sin que haga una sola consulta y luego haga varias en pocos días.
- Tiene el hábito de navegar en grandes volúmenes de datos. No sabe lo que quiere antes de iniciar el proceso de exploración. Busca patrones en los datos que pueden o no existir.
- Opera en modo heurístico, es decir, no sabe el próximo paso de análisis hasta completar los resultados del paso actual.
- También opera en el marco de proyectos. Cuando se termina un proyecto, el proceso de exploración finaliza.
- Muchas veces busca algo y nunca lo encuentra, pero en otras ocasiones, encuentra un diamante del que nadie se había percatado.
- Acceden a todos los datos del Data Warehouse.

- En el Corporate Information Factory, generalmente acceden al Data Warehouse y al Exploration Warehouse.

Mineros

- Es el individuo que navega en pilas de datos y determina si los datos dicen algo o no. A partir de un supuesto, determina su validez a través de herramientas estadísticas.
- Opera en proyectos.
- Crea consultas enormes en tamaño.
- Opera de modo heurístico.
- Frecuentemente trabaja en cooperación con un Explorador. El Explorador crea los supuestos e hipótesis y el Minero determina su validez.
- Tiene habilidades especiales en matemática que lo diferencia de otros técnicos.
- En el Corporate Information Factory, generalmente acceden al Data Mining Warehouse o al Exploration Warehouse.

Turistas

- Es un individuo que sabe dónde encontrar cosas. Tiene una gran amplitud de conocimientos, pero no gran profundidad.
- Está familiarizado con los sistemas formales e informales.
- Sabe cómo usar Internet.
- Conoce sobre metadatos, sabe dónde encontrar índices y cómo usarlos.
- Conoce sobre datos estructurados y desestructurados.
- Conoce de código fuente, cómo leerlo e interpretarlo.
- En el 'Corporate Information Factory' generalmente accede al Repositorio de Metadatos.



Bibliografía de referencias

- **Cano, J. L.** (2007). *Business Intelligence: Competir con Información*. España: Banesto Fundación Cultural y ESADE.
- **Inmon, W.** (2005). *Building the Data Warehouse* (4º Edición). Estados Unidos: Wiley Publishing.
- **Gartner** (2013). Gartner Says Worldwide Business Intelligence, CPM and Analytic Applications/Performance Management Software Market Grew Seven Percent in 2012. Recuperado el 8 de Mayo de 2014: <http://www.gartner.com/newsroom/id/2507915>.
- **Kimball Ralph, R. M.** (2013). *The Data Warehouse Toolkit* (3º Edición). Estados Unidos: Wiley Publishing.
- **Kimball Ralph, R. M.; Thornthwaite, W. y Mundy Joy, B. B.** (2008). *The Data Warehouse Lifecycle Toolkit* (2º Edición). Estados Unidos: Wiley Publishing.

Tipos de soluciones



Bases de Datos II

UNIVERSIDAD

SIGLO 21

MIEMBRO DE LA RED
ILUMNO



Tipos de soluciones

Clasificación de Wayne Eckerson y Cindi Howson

Por otra parte, de acuerdo con Josep Lluís Cano (2007), en base a la clasificación de Wayne Eckerson y Cindi Howson (2005), podemos identificar dos grandes tipos de usuarios:

- Los productores de información
- Los consumidores de información

Para cada uno de los cuales podemos también identificar qué **tipo de soluciones** son las más apropiadas:

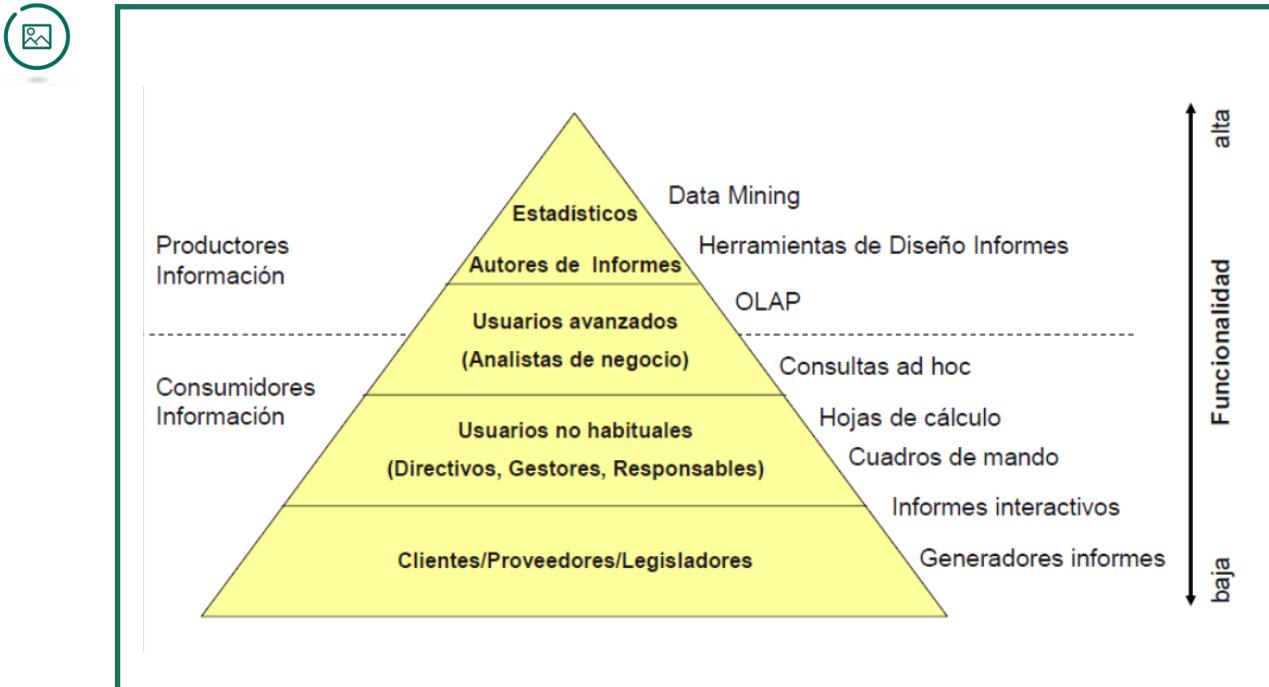
Los productores de información: Normalmente se trata del 20% de los usuarios y utilizan herramientas desktop para crear informes o modelos. [...] se trata de estadísticos que utilizan herramientas Data Mining o autores de informes que utilizan herramientas de diseño o de programación para crear informes específicos. Habitualmente los autores de informes son: técnicos de sistemas de información no usuarios de negocio avanzados que son capaces de entender la información y la informática. Los usuarios avanzados pueden crear o utilizar informes, por lo que en el gráfico están a medio camino entre los productores y los consumidores de información. Usualmente utilizan hojas de cálculo, herramientas de consulta y de informes para acceder y analizar la información.

Los consumidores de información: La mayoría de los consumidores de información son usuarios no habituales que regularmente consultan informes para la toma de decisiones, pero no acceden a los números o hacen análisis detallados diariamente. Los usuarios no habituales son directivos, gestores, responsables, colaboradores y usuarios externos. Este numeroso grupo está bien servido con cuadros de mando con análisis guiados, informes interactivos (por ejemplo: OLAP, informes parametrizados, vinculados) e informes de gestión estandarizados. La mayoría de estas herramientas proveen ahora acceso vía web para promover

el acceso desde cualquier lugar y facilitar el uso y minimizar los costes de administración y mantenimiento.

(Cano, 2007, p. 140)

Imagen 1: Tipos de Usuarios y Soluciones según Eckerson y Howson



Fuente: Josep Lluís Cano (2007, p. 141)

Es importante señalar que se diferencian los roles, no los individuos.

La mayoría de los usuarios no se adecuan absolutamente a una de las dos categorías. Un ejecutivo podría querer ver información financiera utilizando un informe estático publicado semanalmente, pero ver información operacional en tiempo real utilizando un cuadro de mando.

De esta manera, es crítico conocer la información a la que acceden los usuarios y las funcionalidades que necesitan en función de los roles.

(Cano, 2007, p. 141)



Bibliografía de referencias

- **Cano, J. L.** (2007). *Business Intelligence: Competir con Información*. España: Banesto Fundación Cultural y ESADE.
- **Inmon, W.** (2005). *Building the Data Warehouse* (4º Edición). Estados Unidos: Wiley Publishing.
- **Gartner** (2013). Gartner Says Worldwide Business Intelligence, CPM and Analytic Applications/Performance Management Software Market Grew Seven Percent in 2012. Recuperado el 8 de Mayo de 2014: <http://www.gartner.com/newsroom/id/2507915>.
- **Kimball Ralph, R. M.** (2013). *The Data Warehouse Toolkit* (3º Edición). Estados Unidos: Wiley Publishing.
- **Kimball Ralph, R. M.; Thornthwaite, W. y Mundy Joy, B. B.** (2008). *The Data Warehouse Lifecycle Toolkit* (2º Edición). Estados Unidos: Wiley Publishing.