

GUIA DE CURSO:

DATA WAREHOUSING BÁSICO CON PENTAHO

Versión 1.0

ELABORADO POR:

ING. LUIS TAMAYO

www.relative-engine.com



PARA:

CEC - EPN

FECHA DE PRESENTACIÓN:

Quito, 08 de Abril del 2013



Contenido

OBJETIVOS DEL CURSO.....	3
MÓDULO I. CONCEPTOS GENERALES DE INTELIGENCIA DE NEGOCIOS.....	4
CONCEPTOS GENERALES	4
PROCESO DE INTELIGENCIA DE NEGOCIOS.....	4
<i>INFORMACION OPERATIVA</i>	4
<i>ETL</i>	5
<i>ETERPRISE DATAWAREHOUSE (EDWH)</i>	5
<i>SOLUCIONES DE INTELIGENCIA DE NEGOCIOS</i>	12
MODULO 2 MODELADO DIMENSIONAL.....	13
ELEGIR EL PROCESO DE NEGOCIO	14
ESTABLECER EL NIVEL DE GRANULARIDAD	14
ELEGIR LAS DIMENSIONES	14
IDENTIFICAR LAS TABLAS DE HECHOS Y MEDIDAS.....	14
MODELO GRÁFICO DE ALTO NIVEL.....	15
IDENTIFICACIÓN DE ATRIBUTOS DE DIMENSIONES Y TABLAS DE HECHOS	16
IMPLEMENTAR EL MODELO DIMENSIONAL DETALLADO	17
PRUEBA DEL MODELO	17
REVISIÓN Y VALIDACIÓN DEL MODELO	17
DOCUMENTOS FINALES	17
DISEÑO FÍSICO.....	18
DISEÑO DEL SISTEMA DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETL).....	18
<i>LA TECNICA SCD</i>	18
ESPECIFICACIÓN Y DESARROLLO DE APLICACIONES DE BI.....	19
<i>INFORMES ESTANDAR</i>	19
<i>APLICACIONES ANALITICAS</i>	20
REFERENCIAS	21

OBJETIVOS DEL CURSO

- Comprender los conceptos generales de Inteligencia de Negocios y Data Warehousing
- Comprender y aplicar la metodología de Kimball para la construcción de un sistema de DWH/BI
- Diseñar modelos dimensionales de datos
- Conocer, instalar y configurar la plataforma de DWH/BI Pentaho
- Utilizar Pentaho Data Integration para realizar las actividades de integración de datos y procesos de extracción, transformación y carga (ETL)
- Administrar de forma básica la plataforma de DWH/BI Pentaho.

El curso se comprende de cuatro módulos:

- **Módulo I.** Conceptos Generales. Inteligencia de Negocios y Data Warehousing
- **Módulo II.** Técnicas de modelamiento dimensional
- **Módulo III.** Caso de Estudio a ser utilizado durante el transcurso de todo el curso
- **Módulo IV.** Diseño del Data Warehouse para el caso de estudio
- **Módulo V.** Plataforma de DWH/BI Pentaho CE v4.0
- **Módulo VI.** Integración de datos y actividades de extracción, transformación y carga (ETL) usando Pentaho Data Integration

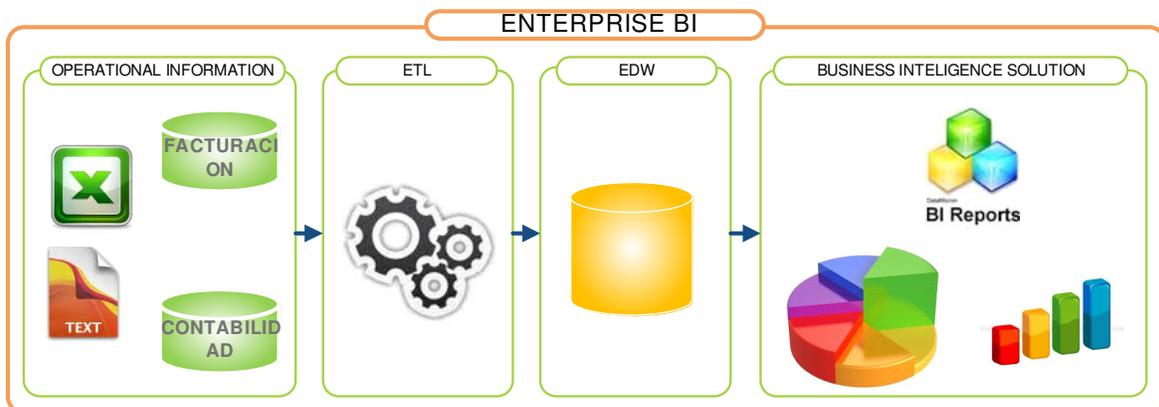
MÓDULO I. CONCEPTOS GENERALES DE INTELIGENCIA DE NEGOCIOS

CONCEPTOS GENERALES

- Que es la inteligencia de negocios (BI)?
- El término BI se refiere al uso de datos en una empresa y convertirlo en información analizada para facilitar la toma de decisiones.
- Abarca la comprensión del funcionamiento actual de la empresa, así como la anticipación de acontecimientos futuros, con el objetivo de ofrecer conocimientos para respaldar las decisiones empresariales.
- Apoya a la toma de decisiones estratégicas, tácticas y operativas de una compañía
- BI no es Data Warehousing.

PROCESO DE INTELIGENCIA DE NEGOCIOS

De acuerdo al nivel y a la complejidad de análisis, las actividades de inteligencia de negocios se resumen de la siguiente manera:



INFORMACION OPERATIVA

- Corresponde a la información que administra la empresa como parte de su giro de negocio.
- La misma puede provenir de distintas fuentes:
 - Archivos de texto
 - Archivos Excel
 - BDD empresariales

ETL

- Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizarla y apoyar un proceso de negocio
 - Extract (Extraer)
 - Extraer la información de las diferentes fuentes de datos.
 - Transform (Transformar)
 - Esta fase aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados.
 - Load (carga)
 - Es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino

ENTERPRISE DATAWAREHOUSE (EDWH)

- Un almacén de datos (DM o DWH) es la principal fuente de datos utilizada para la presentación de informes y análisis BI.
- Los datos almacenados se cargan desde las fuentes de datos a través de los ETL (por ejemplo, facturación, contabilidad).
- Usualmente guarda años de historia y es consultado para inteligencia de negocios u otras actividades analíticas
- Un Data Mart (DM) es un DWH para un tema o aplicación específica (Ventas, Reclamos, etc.)

Ventajas de un DWH

- Toda la información está un solo lugar (Una sólo fuente de la verdad)
- Información actualizada
- Acceso rápido
- No hay límites de espacio (Ej. Archivos XLS)
- Contiene toda la historia de la compañía
- Fácil de comprender (Modelada en términos del negocio)
- Contiene definiciones claras y uniformes
- Datos estandarizados

Desafíos de un DWH

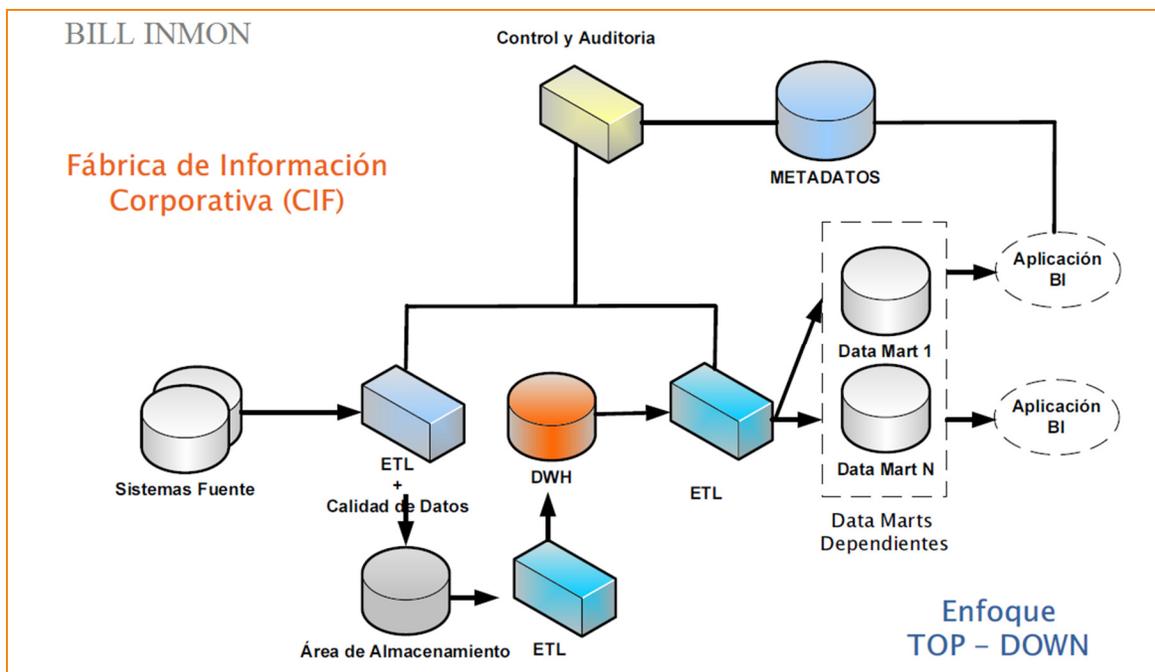
- Problemas de calidad de datos, entre ellos:
 - Datos Duplicados
 - Datos Incompletos
 - Datos Incorrectos
 - Datos Conflictivos
 - Datos Faltantes (Falla de Integridad Referencial)
 - Valores nulos

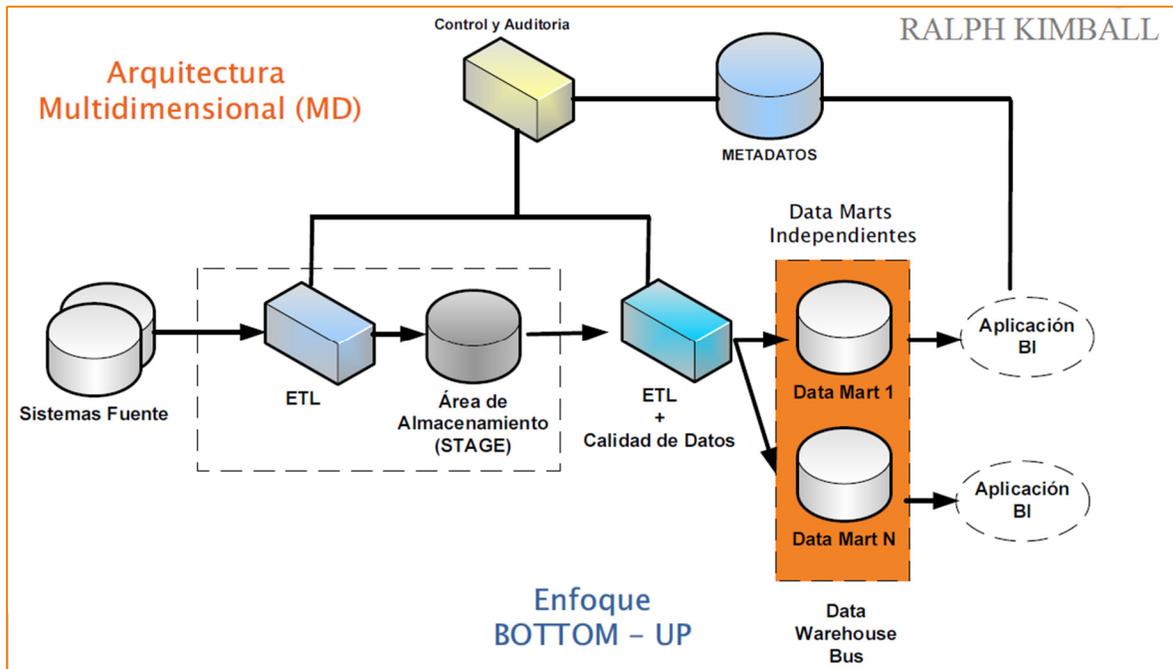
- Volumen de datos (Tb) y rendimiento (Índices, índices binarios, particiones, agregaciones, vistas materializadas).
- Captura de Cambios de Datos (CDC)
- Consiste en identificar qué datos han sido insertados, modificados o borrados
- Puede ser Intrusiva (Triggers) y No Intrusiva (PDI, Merge Rows)
- Requerimientos cambiantes (Un DWH Evoluciona).

Metodologías

METODOLOGÍAS ACTUALES

- Existen muchas metodologías de diseño y construcción de DW.
- Cada fabricante de software de inteligencia de negocios busca imponer una metodología con sus productos. Sin embargo, se imponen entre la mayoría dos metodologías, la de Kimball y la de Inmon. Para comprender la mayor diferencia entre estas dos metodologías, debemos explicar además de la noción de DW mencionando en la introducción, la idea de Data mart.
- Un Data mart es un repositorio de información, similar a un DW, pero orientado a un área o departamento específico de la organización (por ejemplo Compras, Ventas, RRHH, etc.), a diferencia del DW que cubre toda la organización, es decir la diferencia fundamental es su alcance.
- Desde el punto de vista arquitectónico, la mayor diferencia entre los dos autores es el sentido de la construcción del DW, esto es comenzando por los Data marts o ascendente (Bottom-up, Kimball) o comenzando con todo el DW desde el principio, o descendente (TopDown, Inmon).





- Por otra parte, la metodología de Inmon se basa en conceptos bien conocidos del diseño de bases de datos relacionales; la metodología para la construcción de un sistema de este tipo es la habitual para construir un sistema de información, utilizando las herramientas habituales, al contrario de la de Kimball, que se basa en un modelado dimensional (no normalizado)
- **¿Cuál metodología adoptar?**
 - Para el curso adoptaremos la metodología de Kimball, por cuanto proporciona un enfoque de menor a mayor, muy versátil, y una serie de herramientas prácticas que ayudan a la implementación de un DW. Es acorde a nuestras empresas porque se pueden implementar pequeños datamarts en áreas específicas de las mismas (compras, ventas, etc.), con pocos recursos y de poco irlos integrándolos en un gran almacén de datos. Por tanto, detallaremos esta metodología.

METODOLOGÍA KIMBALL EN DETALLE

- La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Este ciclo de vida del proyecto de DW, está basado en cuatro principios básicos:

Centrarse en el negocio

- Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.

Construir una infraestructura de información adecuada

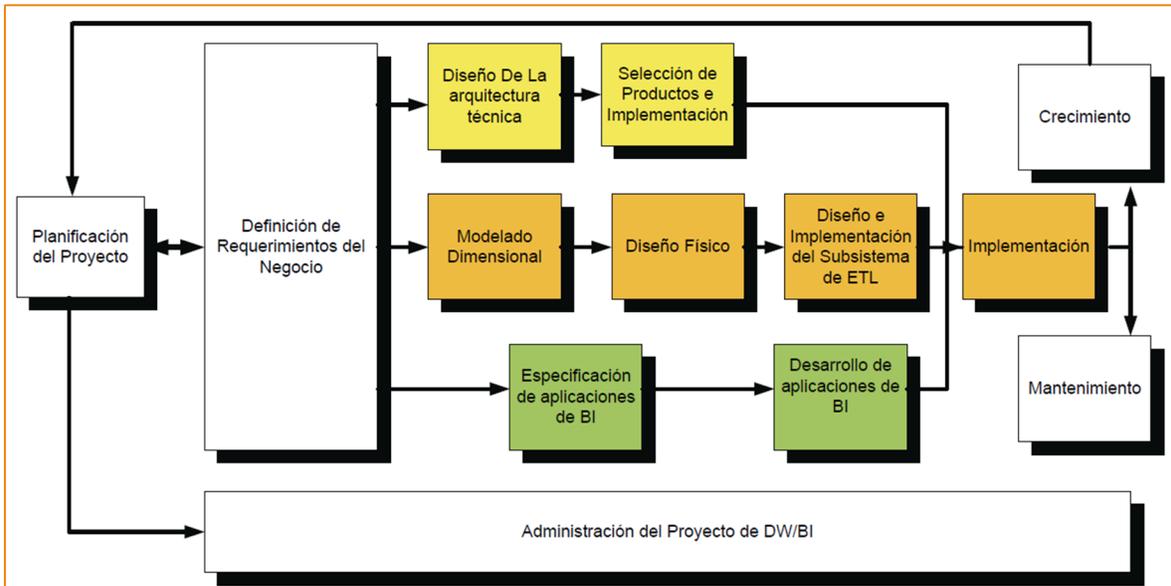
- Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.

Realizar entregas en incrementos significativos

- Crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.

Ofrecer la solución completa

- Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.
- La construcción de una solución de DW/BI (Datawarehouse/Business Intelligence) es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad. Las tareas de esta metodología (ciclo de vida) se muestran en la figura



- De la figura, podemos observar dos cuestiones.
- Primero, hay que resaltar el rol central de la tarea de definición de requerimientos.
- Los requerimientos del negocio son el soporte inicial de las tareas subsiguientes. También tiene influencia en el plan de proyecto (nótese la doble fecha entre la caja de definición de requerimientos y la de planificación).
- En segundo lugar podemos ver tres rutas o caminos que se enfocan en tres diferentes áreas:
 - Tecnología (Camino Superior). Implica tareas relacionadas con software específico, por ejemplo, Microsoft SQL Analysis Services.
 - Datos (Camino del medio). En la misma diseñaremos e implementaremos el modelo dimensional, y desarrollaremos el subsistema de Extracción, Transformación y Carga (Extract, Transformation, and Load - ETL) para cargar el DW.
 - Aplicaciones de Inteligencia de Negocios (Camino Inferior). En esta ruta se encuentran tareas en las que diseñamos y desarrollamos las aplicaciones de negocios para los usuarios finales.
- Estas rutas se combinan cuando se instala finalmente el sistema.
- En la parte de debajo de la figura se muestra la actividad general de administración del proyecto. A continuación describiremos cada una de las tareas.

Planificación

- En este proceso se determina el propósito del proyecto de DW/BI, sus objetivos específicos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información.
- En la visión de programas y proyectos de Kimball, Proyecto, se refiere a una iteración simple del KLC (Kimball Life Cycle), desde el lanzamiento hasta el despliegue.
- Esta tarea incluye las siguientes acciones típicas de un plan de proyecto:
 - Definir el alcance (entender los requerimientos del negocio).
 - Identificar las tareas
 - Programar las tareas
 - Planificar el uso de los recursos.
 - Asignar la carga de trabajo a los recursos
 - Elaboración de un documento final que representa un plan del proyecto.
- Además en esta parte definimos cómo realizar la administración o gestión de esta subfase que es todo un proyecto en sí mismo, con las siguientes actividades:
 - Monitoreo del estado de los procesos y actividades.
 - Rastreo de problemas
 - Desarrollo de un plan de comunicación comprensiva que dirija la empresa y las áreas de TI.

Análisis de requerimientos

- La definición de los requerimientos es en gran medida un proceso de entrevistar al personal de negocio y técnico, pero siempre conviene tener un poco de preparación previa. Se debe aprender tanto como se pueda sobre el negocio, los competidores, la industria y los clientes del mismo. Hay que leer todos los informes posibles de la organización; rastrear los documentos de estrategia interna; entrevistar a los empleados, analizar lo que se dice en la prensa acerca de la organización, la competencia y la industria. Se deben conocer los términos y la terminología del negocio.
- Parte del proceso de preparación es averiguar a quién se debe realmente entrevistar.
- Esto normalmente implica examinar cuidadosamente el organigrama de la organización. Hay básicamente cuatro grupos de personas con las que hablar desde el principio:
 - El directivo responsable de tomar las decisiones estratégicas
 - Los administradores intermedios y de negocio responsables de explorar alternativas estratégicas y aplicar decisiones

- Personal de sistemas, si existen, la gente que realmente sabe qué tipos de problemas informáticos y de datos existen.
- Por último, la gente que se necesita entrevistar por razones políticas.
- A partir de las entrevistas, podemos identificar temas analíticos y procesos de negocio. Los temas analíticos agrupan requerimientos comunes en un tema común.

Tema Analítico	Análisis o requerimiento inferido o pedido	Proceso de negocio de soporte	Comentarios
Planificación de ventas	Análisis histórico de ordenes de revendedores	Ordenes de compras	Por cliente, por país, por región de ventas
	Proyección de ventas	Ordenes de compras	La proyección es un proceso de negocio que usa las órdenes como entradas

- Por otra parte, a partir del análisis se puede construir una herramienta de la metodología denominada matriz de procesos/dimensiones (Bus Matrix en inglés).
- Una dimensión es una forma o vista o criterio por medio de cual se pueden sumariar, cruzar o cortar datos numéricos a analizar, datos que se denominan medidas (measures en inglés).
- Esta matriz tiene en sus filas los procesos de negocio identificado, y en las columnas, las dimensiones identificadas.
- Un ejemplo de esta matriz se puede observar en la siguiente tabla. Cada X en la intersección de las filas y columnas significa que en el proceso de negocio de la fila seleccionada se identifican las dimensiones propuestas.

Proceso de Negocio	Dimensiones					
	Tiempo	Producto	Empleados	Clientes (Revendedores)	Geografía de ventas	Importes
Proyección de ventas	X	X	X	X	X	X
Compras	X	X	X	X	X	X
Control de llamadas	X	X	X	X	X	
...						

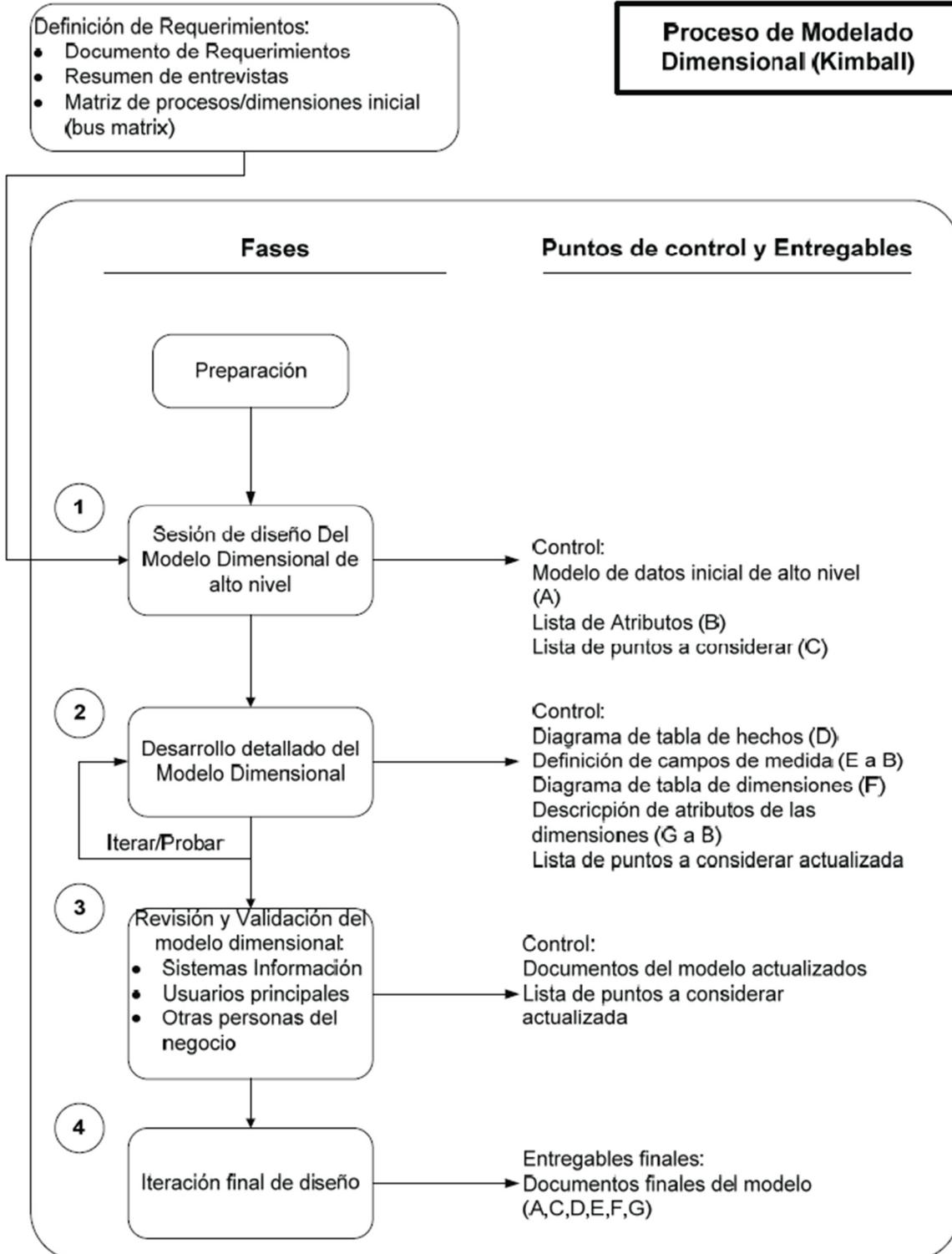
- Finalmente se busca priorizar los requerimientos o procesos de negocios más críticos.

SOLUCIONES DE INTELIGENCIA DE NEGOCIOS

- Este conjunto de herramientas y metodologías tienen en común las siguientes características:
 - Accesibilidad a la información.
 - Los datos son la fuente principal de este concepto.
 - Acceso de los usuarios a los datos con independencia de la procedencia de estos.
 - Apoyo en la toma de decisiones.
 - Orientación al usuario final.
 - Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.
- Para este curso se utilizara la Suite Pentaho como solución de BI.

MODULO 2 MODELADO DIMENSIONAL

La creación de un modelo dimensional es un proceso dinámico y altamente iterativo.



- El proceso de diseño comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados de la matriz descrita en el módulo anterior.
- El proceso iterativo consiste en cuatro pasos:
 - Elegir el proceso de negocio.
 - Establecer el nivel de granularidad.
 - Elegir las dimensiones.
 - Identificar medidas y las tablas de hechos.

Elegir el proceso de negocio

- El primer paso es elegir el área a modelar. Esta es una decisión de la dirección, y depende fundamentalmente del análisis de requerimientos y de los temas analíticos anotados en la etapa anterior.

Establecer el nivel de granularidad

- La granularidad significa especificar el nivel de detalle. La elección de la granularidad depende de los requerimientos del negocio y lo que es posible a partir de los datos actuales.
- La sugerencia general es comenzar a diseñar el DW al mayor nivel de detalle posible, ya que se podría luego realizar agrupamientos al nivel deseado. En caso contrario no sería posible abrir (drill-down) las sumalizaciones en caso de que el nivel de detalle no lo permita.
- Las agregaciones mejoran el rendimiento.

Elegir las dimensiones

- Las dimensiones surgen naturalmente de las discusiones del equipo, y facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensiones. Las tablas de dimensiones tienen un conjunto de atributos (generalmente textuales) que brindan una perspectiva o forma de análisis sobre una medida en una tabla hechos.
- Una forma de identificar las tablas de dimensiones es que sus atributos son posibles candidatos para ser encabezado en los informes, tablas pivot, cubos, o cualquier forma de visualización, unidimensional o multidimensional.

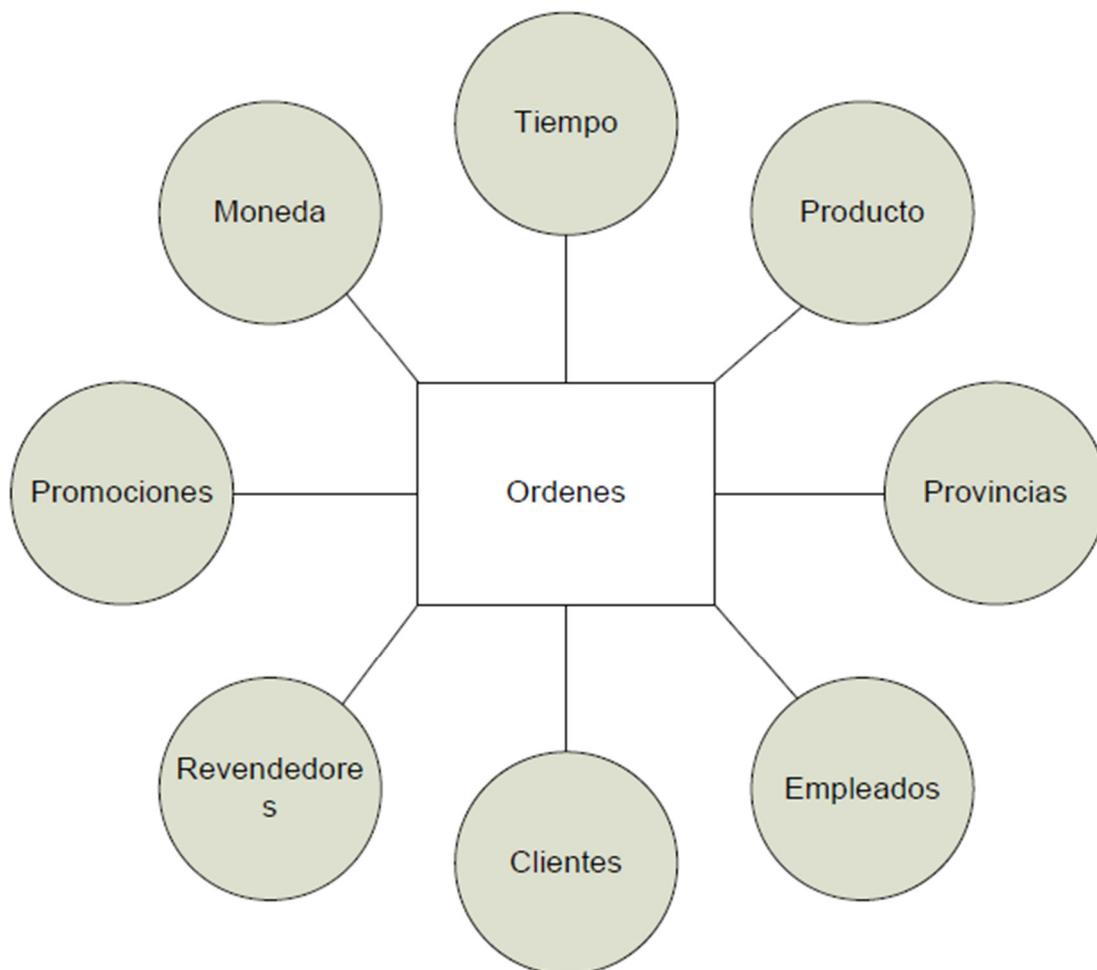
Identificar las tablas de hechos y medidas.

- El último paso consiste en identificar las medidas que surgen de los procesos de negocios.
- Una medida es un atributo (campo) de una tabla que se desea analizar, sumalizando o agrupando sus datos, usando los criterios de corte conocidos como dimensiones.
- Las medidas habitualmente se vinculan con el nivel de granularidad, y se encuentran en tablas que denominamos tablas de hechos (fact tables en inglés).

- Cada tabla de hechos tiene como atributos una o más medidas de un proceso organizacional, de acuerdo a los requerimientos. Un registro contiene una medida expresada en números, como ser cantidad, tiempo, dinero, etc., sobre la cual se desea realizar una operación de agregación en función de una o más dimensiones.
- Las cinco funciones básicas de agregación son:
 - Máximo: max()
 - Mínimo: min()
 - Conteo: count()
 - Suma: sum()
 - Promedio: avg()
- La granularidad es el nivel de detalle que posee cada registro de una tabla de hechos.

Modelo gráfico de alto nivel

- Para concluir con el proceso dimensional inicial se realiza un gráfico denominado modelo dimensional de alto nivel (o gráfico de burbujas, Bubble chart, en el léxico de Kimball), como ilustra la figura.



Identificación de atributos de dimensiones y tablas de hechos

- La segunda parte de la sesión inicial de diseño consiste en completar cada tabla con una lista de atributos bien formada. Una lista de este tipo se muestra en la figura. Esta lista o grilla se forma colocando en las filas los atributos de la tabla, y en las columnas la siguiente información:
 - Características relacionadas con la futura tabla dimensional del almacén de datos (target), por ejemplo tipo de datos, si es clave primaria, valores de ejemplo, etc.
 - El origen de los datos (source, por lo general atributos de las tablas transaccionales).
 - Reglas de conversión, transformación y carga (ETL rules), que nos dicen cómo transformar los datos de las tablas de origen a las del almacén de datos.

Table Name:	DimOrderInfo
Table Type:	Dimension
View Name:	OrderInfo
Description:	OrderInfo is the "junk" dimension that includes miscellaneous information about the Order transaction
Used in schemas:	Orders
Generate script?	Y

Column Name	Description	Target									
		Datatype	Size	Key?	FK To	NULL?	Default Value	Unknown Member	Example Values	SCD Type	Source System
Extended Property?	Y				Y				Y	Y	Y
OrderInfoKey	Surrogate primary key	smallint		PK ID		N		-1	1, 2, 3, 4...		ETL Process
BKSalesReasonID	Sales reason ID from source system	smallint				N		-1			AW
Channel	Sales channel	char	8					Unknown	Reseller, Internet	1	AW
SalesReason	Reason for the sale, as reported by the customer	varchar	30					Unknown		1	AW
SalesReasonType	Type of sales reason	char	10					Unknown	Marketing, Promotion, Other	1	AW
AuditKey	What process loaded this row?	int		FK	Audit Dim	N		-1		1	Derived

SCD Type	Source						ETL Rules	Comments
	Source System	Source Schema	Source Table	Source Field Name	Source Datatype			
Y	Y	Y	Y	Y	Y	Y	Y	Y
							Standard surrogate key	
	AW	Sales	SalesReason	SalesReasonID	int		Convert to char; left-pad with zero. R for reseller row.	We need to insert a single row for Reseller
1	AW	Sales	SalesReason	DerivedID			Internet! for real sales reasons. 'Reseller' for reseller row.	
1	AW	Sales	SalesReason	Name	nvarchar(30)		Convert to varchar; 'Reseller' for reseller row.	
1	AW	Sales	SalesReason	ReasonType	nvarchar(30)		Convert to varchar; 'Reseller' for reseller row	
1	Derived						Populated by ETL system using standard technique	

Implementar el modelo dimensional detallado

- Este proceso consiste simplemente en completar la información incompleta de los pasos anteriores.
- El objetivo en general es identificar todos los atributos útiles y sus ubicaciones, definiciones y reglas de negocios asociadas que especifican cómo se cargan estos datos. Para este cometido se usa la misma planilla del punto anterior.

Prueba del modelo

- Si el modelo ya está estable, lo que se hace habitualmente es probarlo contra los requerimientos del negocio. Haciendo la pregunta práctica de ¿Cómo podemos obtener esta información en particular del modelo? Para las pruebas podemos usar diseños de reportes estructurados, de usuarios actuales, diseños de cubos prospectivos, etc.

Revisión y validación del modelo

- Un vez que tenemos confianza plena en el modelo, ingresamos en esta etapa final (ver figura 2), lo cual implica revisar el modelo con diferentes audiencias, cada una con diferentes conocimientos técnicos y del negocio.
- En el área de sistemas deberían revisarlo los programadores y analistas de los sistemas, y el DBA si existe.
- También debería revisarse con usuarios y personas del negocio que tengan mucho conocimiento de los procesos y que quizás no hayan participado del diseño del modelo.
- Finalmente podemos hacer un documento que enuncie una serie de preguntas del negocio (tomadas a partir de los requerimientos), y las conteste por medio del modelo.

Documentos finales

- El producto final, como se puede ver en la Figura, son una serie de documentos (solo mencionamos los más importantes), a saber:
 - Modelo de datos inicial de alto nivel
 - Lista de atributos
 - Diagrama de tablas de hechos
 - Definición de campos de medida
 - Diagrama de tablas de dimensiones
 - Descripción de los atributos de las dimensiones
 - Matriz DW (o DW Bus Matrix) completa

Diseño Físico

- En esta parte, intentamos contestar las siguientes preguntas:
 - ¿Cómo puede determinar cuán grande será el sistema de DW/BI?
 - ¿Cuáles son los factores de uso que llevarán a una configuración más grande y más compleja?
 - ¿Cómo se debe configurar el sistema?
 - ¿Cuánta memoria y servidores se necesitan? ¿Qué tipo de almacenamiento y procesadores?
 - ¿Cómo instalar el software en los servidores de desarrollo, prueba y producción?
 - ¿Qué necesitan instalar los diferentes miembros del equipo de DW/BI en sus estaciones de trabajo?
 - ¿Cómo convertir el modelo de datos lógico en un modelo de datos físicos en la base de datos relacional?
 - ¿Cómo conseguir un plan de indexación inicial?
 - ¿Debe usarse la partición en las tablas relacionales?

Diseño del sistema de Extracción, Transformación y Carga (ETL).

- El sistema de Extracción, Transformación y Carga (ETL) es la base sobre la cual se alimenta el Datawarehouse.
- Si el sistema ETL se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar (grabar) la información en el DW en un formato acorde para la utilización por parte de las herramientas de análisis.

LA TECNICA SCD

- La técnica de Slowly Changing Dimensions permite identificar y guardar qué registros cambiaron.
- Existen los siguientes tipos:

Tipo	Descripción
SCD – Tipo 1	Sobrescribe el registro.
SCD – Tipo 2	Agrega un registro con columnas que indican su validez.
SCD – Tipo 3	Consiste en agregar una columna por cada una que se quiera registrar el cambio.

Especificación y desarrollo de aplicaciones de BI

- Una parte fundamental de todo proyecto de DW/BI está en proporcionarles a una gran comunidad de usuarios una forma más estructurada y por lo tanto, más fácil, de acceder al almacén de datos.
- Proporcionamos este acceso estructurado a través de lo que llamamos aplicaciones de inteligencia de negocios (Business Intelligence Applications).
- Las aplicaciones de BI son la cara visible de la inteligencia de negocios: los informes y aplicaciones de análisis proporcionan información útil a los usuarios. Las aplicaciones de BI incluyen un amplio espectro de tipos de informes y herramientas de análisis, que van desde informes simples de formato fijo a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio.
- Kimball divide a estas aplicaciones en dos categorías basadas en el nivel de sofisticación, y les llama informes estándar y aplicaciones analíticas.

INFORMES ESTANDAR

- Los informes estándar son la base del espectro de aplicaciones de BI. Por lo general son informes relativamente simples, de formato predefinido, y parámetros de consulta fijos.
- En el caso más simple, son informes estáticos pre-almacenados. Los informes estándar proporcionan a los usuarios un conjunto básico de información acerca de lo que está sucediendo en un área determinada de la empresa. Este tipo de aplicaciones son el caballo de batalla de la BI de la empresa.
- Son informes que los usuarios usan día a día. La mayor parte de lo que piden las personas durante el proceso de definición de requisitos se clasificaría como informes estándar. Por eso es conveniente desarrollar un conjunto de informes estándar en el ciclo de vida del proyecto.
- Algunos informes estándares típicos podrían ser:
 - Ventas del año actual frente a previsión de ventas por vendedor
 - Tasa de renovación mensual por plan de servicio
 - Tasa quinquenal de deserción por unidad académica
 - Tasas de respuestas de correo electrónico por promoción por producto (marketing)
 - Recuento de audiencia y porcentaje de la audiencia total por la red de televisión por día de la semana y hora del día (Sistema de marketing televisivo)
 - Reclamos del año actual hasta la fecha frente a previsión, por tipo de vehículo
 - Volumen de llamadas por producto como un porcentaje del total de ventas

APLICACIONES ANALITICAS

- Las aplicaciones analíticas son más complejas que los informes estándar.
- Normalmente se centran en un proceso de negocio específico y resumen cierta experiencia acerca de cómo analizar e interpretar ese proceso de negocio.
- Estas aplicaciones pueden ser muy avanzadas e incluir algoritmos y modelos de minería de datos, que ayudan a identificar oportunidades o cuestiones subyacentes en los datos.
- Otra característica avanzada en algunas aplicaciones analíticas es que el usuario puede pedir cambios en los sistemas transaccionales basándose en los conocimientos obtenidos del uso de la aplicación de BI.
- En el otro extremo del espectro, algunas aplicaciones analíticas se venden como soluciones cerradas o enlatados, y son independientes de las aplicaciones particulares de la empresa.
- Algunas aplicaciones analíticas comunes incluyen:
 - Análisis de la eficacia de la promociones
 - Análisis de rutas de acceso en un sitio Web
 - Análisis de afinidad de programas
 - Planificación del espacio en espacios comerciales
 - Detección de fraudes
 - Administración y manejo de categorías de productos

REFERENCIAS

- Imhoff & Gallemmo, Mastering Data Warehouse Design: Relational and Dimensional Techniques, Wiley Publishing, 2003 Inmon, Building the Data Warehouse, (Third Edition). John Wiley & Sons, 2002.
- La metodología de Kimball para el diseño de almacenes de datos (Data warehouses) Gustavo R. Rivadera Facultad de Ingeniería e Informática, UCASAL.

Páginas Útiles

- <http://www.kimballgroup.com> : Este sitio contiene mucha información y artículos sobre la metodología, y además una serie de planillas de Excel usadas en cada paso de la metodología.
- <http://www.bi-bestpractices.com/view-articles/4768>
- <http://kle.sisorg.com.mx/articulo04.html>
- <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptosdel-modelado-dimENSIONAL/>