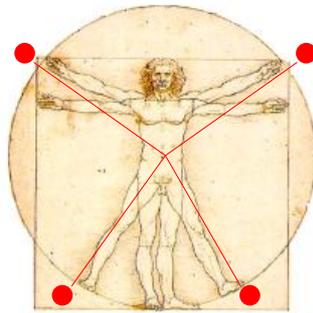


TECNOLOGÍ@ y DESARROLLO

Revista de Ciencia, Tecnología y Medio Ambiente

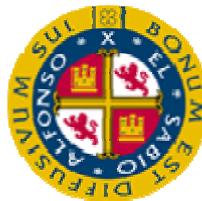
VOLUMEN XII. AÑO 2014

SEPARATA



SOLUCIÓN BUSINESS INTELLIGENCE (BI) CORPORATIVA Transformando Datos en Conocimiento

Nerea Sevilla-Marchena, Antonio J. Reinoso, Jesús Ordoñez-Orzaiz, Juan Ortega-Valiente



UNIVERSIDAD ALFONSO X EL SABIO
Escuela Politécnica Superior
Villanueva de la Cañada (Madrid)

© Del texto: Nerea Sevilla-Marchena, Antonio J. Reinoso, Jesús Ordoñez-Orzaiz, Juan Ortega-Valiente.
Mayo, 2014.

<http://www.uax.es/publicacion/solucion-business-intelligence-bi-corporativa.pdf>

© De la edición: *Revista Tecnol@ y desarrollo*

Escuela Politécnica Superior.

Universidad Alfonso X el Sabio.

28691, Villanueva de la Cañada (Madrid).

ISSN: 1696-8085

Editor: Javier Morales Pérez – tecnologia@uax.es

No está permitida la reproducción total o parcial de este artículo, ni su almacenamiento o transmisión ya sea electrónico, químico, mecánico, por fotocopia u otros métodos, sin permiso previo por escrito de la revista.

SOLUCIÓN BUSINESS INTELLIGENCE (BI) CORPORATIVA

Transformando Datos en Conocimiento

Nerea Sevilla-Marchena^a, Antonio J. Reinoso^b, Jesús Ordoñez-Orzaiz^c, Juan Ortega-Valiente^d

^{a)} Graduada en Ingeniería de Sistemas de Información, Universidad Alfonso X el Sabio. Avda. De la Universidad nº1, Villanueva de la Cañada, 28691, Madrid. España. nsevimar@myuax.com

^{b)} Doctor por la Universidad Alfonso X el Sabio, Ingeniero Informático. Adjunto a la Jefatura de Estudios Departamento de Ingenierías TIC, Escuela Politécnica Superior, Universidad Alfonso X el Sabio.

Avda. De la Universidad nº1, Villanueva de la Cañada, 28691, Madrid. España. areinpei@myuax.com

^{c)} Licenciado en Informática, Responsable de Inserción Laboral de Lanbide. Servicio Vasco de Empleo. Jose Atxotegi, 1, 01009 Vitoria-Gasteiz. España. jesus.orzaiz@lanbide.net

^{d)} Doctor en Ingeniería Informática. Departamento de Ingenierías TIC, Escuela Politécnica Superior, Universidad Alfonso X el Sabio. Avda. De la Universidad nº1, Villanueva de la Cañada, 28691, Madrid. España. jvalient@myuax.com

RESUMEN: A medida que los sistemas de información de las organizaciones han evolucionado, el volumen y la ineficiencia en la gestión de la información han aumentado. En las diferentes áreas funcionales que componen una organización a menudo se utilizan criterios “proprios” para la generación de informes de gestión, que demandan información diferente según el criterio de definición y de cálculo utilizados. Así mismo, también crece la dependencia de los directores y responsables de las organizaciones del área TIC, ya que sólo ellos conocen bien los modelos de datos de los sistemas de información. En la coyuntura económica actual, donde no hay margen de error para la toma de decisiones en las organizaciones, la información de sus sistemas tiene que ayudar, no entorpecer y es en este momento donde las soluciones BI toman protagonismo. Estas soluciones nos permitirán integrar todo los datos almacenados en los distintos sistemas de información. El BI de la organización será el punto de referencia para la generación de informes. Permitirá a los directores y responsables ser autónomos y utilizar el BI corporativo como complemento a la toma de decisiones.

PALABRAS CLAVE: Business Intelligence, BI, conocimiento, cuadros de mando, inteligencia de negocio, Pentaho, metodología Kimball

ABSTRACT: *The evolution of organizational information systems has brought a considerable increase in the amount of information to process as well as in the inefficiencies found in the management procedures and techniques. Considering that even within a given organization, its different functional areas tend to use their own criteria to produce management reports, very different kind of data are requested to meet all these particular definitions and calculation systems. Likewise, the interdependence between directors and responsables for TIC areas in the organizations has been also increased, as latter are only the ones that have a good knowledge of the data models in which information systems have been built upon. In the current economic situation, where organizations have no margin of error in making their decisions, the information from supporting systems must help, not hinder, and it is at this moment where BI solutions will play a fundamental role. These solutions will allow to integrate all the data stored in the different information systems deployed to support organizations' internal procedures. The BI of the organization will be the basis for building reports. This will allow managers to be autonomous and use the corporate BI for making decisions.*

KEY-WORDS: Business Intelligence, BI, knowledge, Dashboard, Pentaho, Kimball methodology

1. Introducción.

La definición del concepto de Business Intelligence que ofrece la edición española de Wikipedia es:¹

Se denomina inteligencia empresarial, inteligencia de negocios o BI (del inglés business intelligence) al conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa.

Esta definición deja claro el propósito. La inteligencia de negocio persigue crear conocimiento a partir de datos. Las herramientas Business Intelligence ayudan a las organizaciones a analizar los datos para la toma de decisiones que influyan en las estrategias y en la actividad principal (core) de sus negocios.

La inteligencia en un proceso de negocio, consiste en extraer el conocimiento inherente tanto de los datos internos generados por los sistemas de información de la organización (CRM, ERP, Aplicaciones de Gestión, etc...) como de los datos externos que se relacionan con su actividad (Social Media, Estadísticas, Open Data, etc...). Toda esta información permitirá definir planes de acción que resulten de aplicación tanto en áreas internas como en el descubrimiento de nuevas oportunidades para tomar ventaja sobre los competidores directos.

Las herramientas tecnológicas más importantes que permiten la extracción, integración, representación y análisis de datos y que se incluyen generalmente en las soluciones de *Business Intelligence (BI)* son las siguientes:

- **Integración de datos. Procesos ETL**

Software que realiza la integración de datos y su consolidación en una base de datos DataWareHouse (DWH) mediante procesos que los extraen, transforman y cargan de las diferentes fuentes de información.

- **Modelado de datos**

SW que permite construir el modelo BI, que es la representación o abstracción del modelo físico.

- **Reporting**

SW para generación de informes utilizando los indicadores y dimensiones.

- **Cuadros de Mando**

Visualización rápida de los indicadores más importantes.

¹ Definición de BI en wikipedia: (http://es.wikipedia.org/wiki/Inteligencia_empresarial)

- **OLAP: Procesamiento analítico en línea**

Software para el análisis multidimensional de los datos que permite tener una visión más rápida e interactiva de los mismos.

- **Minería de Datos o Data Mining**

Análisis de la información de interés para la predicción de tendencias, comportamiento e identificación de patrones ocultos.

2. Descripción de la arquitectura BI.

Una solución BI es una arquitectura completa, con muchos factores y procesos importantes en el desarrollo que van desde la extracción e integración de los datos, a la construcción del modelo que utilizará el usuario final para que pueda realizar informes y análisis.

Las soluciones BI son un conjunto de tecnologías, aunque habitualmente se las relaciona sólo con herramientas de *Reporting*. Esto se debe a que estas herramientas son las que utiliza el usuario final para la generación de informes. Sin embargo, para poder realizar tales análisis es necesario construir un almacén integral de datos (Datawarehouse, DWH) de la organización. Este sistema será donde se almacene la información integrada de todas las fuentes de datos internas y externas.

Los procesos ETL se encargarán de realizar la extracción de todas las fuentes de datos. Así mismo, también serán responsables de su integración, transformación al modelo BI y carga definitiva en el Datawarehouse. A partir de este DataWarehouse corporativo se construye el modelo que utilizará el BI para la generación de informes y cuadros de mando en la parte de usuario final.

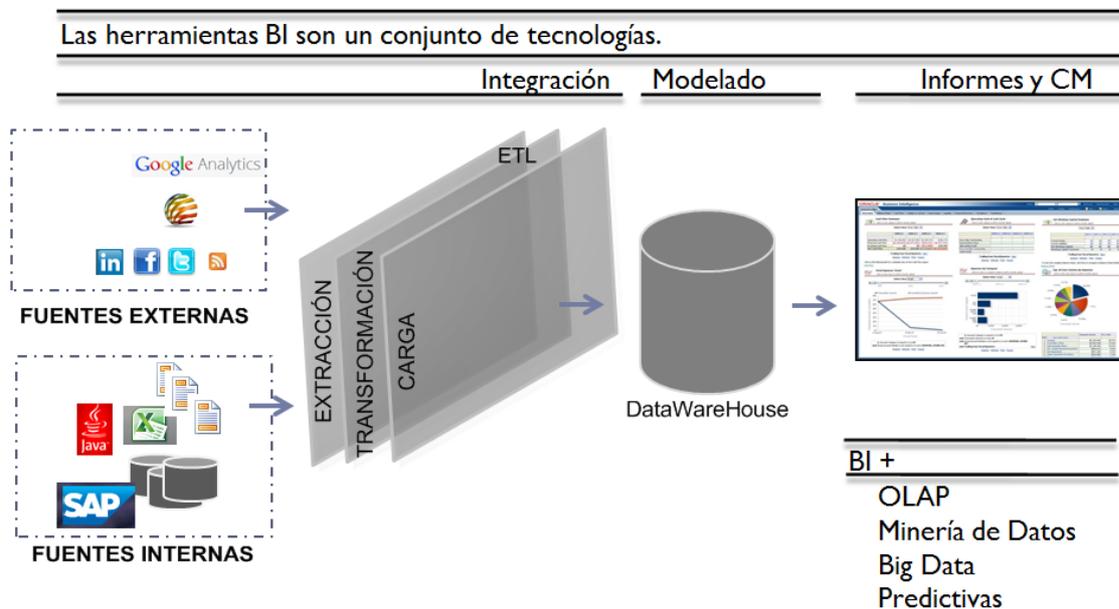


Fig.2.1 Tecnologías que componen una solución BI (Elaboración propia)

Otras herramientas que pueden dar valor añadido a un proyecto BI son los desarrollos OLAP, la Minería de Datos, Big Data (análisis de datos no estructurados) y herramientas de predicción (¿Qué pasaría si...?). Todas estas herramientas forman la familia completa de una solución BI.

3. Revisión del estado del arte.

Una de las clasificaciones para la evaluación de las soluciones BI del mercado más importantes y seguidas tanto por los medios digitales especializados como los profesionales del mercado BI es el Cuadrante Mágico de soluciones Business Intelligence realizada todos los años por la consultoría Gartner. [1]

Gartner elabora sus informes, principalmente, a través de encuestas a los clientes de las distintas propuestas BI que hay en el mundo.

El informe utiliza dos métricas a modo de ejes:

El eje X, *completeness of visión*, representa el conocimiento de los proveedores sobre cómo se puede aprovechar el momento actual del mercado para generar valor tanto para sus clientes como para ellos mismos.

El eje Y, *ability to execute*, mide la habilidad de los proveedores para ejecutar con éxito su visión del mercado.



Fig 3.1. Gartner. Cuadrante Mágico Business Intelligence 2014. [1]

Los dos ejes dividen el cuadrante en cuatro sectores:

- Líderes (Leaders): Proveedores con una oferta amplia, completa y madura de productos BI, que evoluciona según demanda el mercado.
- Visionarios (Visionaries): Empresas con una fuerte (y acertada) visión del mercado actual en BI. Sin embargo, puede que no tengan la capacidad para realizar implantaciones, bien sea por su tamaño o por otras circunstancias. En 2013 no hay proveedores en esta categoría.
- Contendientes (Challengers): Este es el caso contrario al de los visionarios, se trata de proveedores bien posicionados y con altas posibilidades de éxito en implantaciones. No obstante, suelen ofrecer poca variedad de productos, o directamente centrarse en un único aspecto de lo que demanda el mercado. Por ejemplo, un déficit en su canal de ventas o presencia geográfica.

- Participantes eventuales (Niche Players): La última categoría es en principio la más desfavorable. Son proveedores que no llegan a puntuar lo suficiente en ninguna categoría como para alcanzar uno de los otros cuadrantes. No obstante, no significa que por ello sus soluciones no tengan calidad.

La clasificación de Gartner [1] también permite identificar nuevas oportunidades profesionales para soluciones BI que comienzan a mejorar su presencia, como es el caso de Pentaho que ofrece una plataforma completa de BI Open Source para la integración y el análisis de datos y está actualmente reforzando su posición en el mercado nacional.

Para nuestra propuesta, se ha elegido Pentaho por su bajo coste de licencia, funcionalidad, integración, facilidad de uso para los usuarios y acceso a los datos. En general, en cuanto a las capacidades de BI que se espera de una plataforma, las valoraciones de Pentaho son ligeramente superiores a la media de las analizadas como parte de este estudio y también obtiene buena valoración respecto a las herramientas de desarrollo. Tiene que mejorar en la valoración de funcionalidades de la última versión del producto, 5.0. También debe mejorar la calidad y el soporte del producto si quiere escalar posiciones en la clasificación en años posteriores.

Los autores Chandrasekhar, Vellore, Reddy y Rath hacen un estudio comparativo [2] entre las soluciones BI de código abierto y grandes marcas. En dicho estudio se analizan un gran número de plataformas diferentes para análisis de datos, comparándolas con base en el medio ambiente, la cantidad de datos que pueden ser procesados, las capacidades de toma de decisiones, la facilidad de uso, la energía y el tiempo consumido, y los precios. Junto con la clasificación Gartner es una herramienta que las organizaciones pueden utilizar antes de decidir que herramienta tecnológica utilizarán para el desarrollo de sus proyectos Business Intelligence.

Las soluciones Open Source como Pentaho, han permitido a las organizaciones socialización y uso de estas herramientas determinantes para la planificación y seguimiento de las estrategias que antes estaban sólo al alcance de las grandes organizaciones con grandes recursos económicos como seguros y banca.

En relación a los beneficios de incorporar soluciones BI en las organizaciones hay numerosas investigaciones con artículos publicados en prestigiosas revistas que analizan las ventajas de contar con herramientas de apoyo a la toma de decisiones.

La investigación de Neil Foshay y Craig Kuziemyky en su publicación [3] demuestran la importancia de la implementación de soluciones Business Intelligence en las organizaciones de ámbito de la salud como apoyo a la toma de decisiones basadas en el análisis de la información generada en este tipo de organizaciones. La investigación se desarrolla en Canadian Health Authority y analiza el impacto negativo de no contar con herramientas de apoyo a la toma de decisiones.

Así mismo, el artículo «Business Intelligence and Analytics: From Big Data to Big Impact» [4] refleja la magnitud y el impacto de los problemas relacionados con los datos que hay que resolver en las organizaciones empresariales contemporáneas. La inteligencia empresarial y análisis (BI&A) se ha convertido en un área importante de estudio para los profesionales y los investigadores para poder resolver dichos problemas. En la investigación se analiza los retos y oportunidades que se abren en esta área a incorporar en las organizaciones.

Los beneficios más importantes generados con el desarrollo de una solución BI en las organizaciones serían:

- **Autonomía de las TIC's.** Dotar de autonomía a los órganos de gobierno de las organizaciones para la ayuda a toma de decisiones utilizando los informes y cuadros de mando para el análisis y explotación de información.
- **Integración de datos.** Reducir tareas administrativas en labores de recopilación de datos para la preparación de informes corporativos y reporte de datos, centrándose en el análisis y toma de decisiones frente a la generación de informes.
- **Coherencia de datos.** Minimizar errores en la generación de informes.
- **Consistencia de datos.** Unificar criterios y conceptos en toda la organización.
- **Automatización.** Optimización del tiempo en la generación de informes y difusión y publicación.

4. Metodología Kimball para diseño de Datawarehouse/Business Intelligence (DWH/BI).

Para el desarrollo de la solución Business Intelligence se utilizará el enfoque más utilizado, la metodología Kimball². Existen muchas otras ya que cada proveedor intenta imponer la suya para que se utilicen sus productos. Sin embargo, en opinión de los autores, la metodología Kimball es la más genérica y versátil, ya que se puede adaptar en función de la solución BI finalmente elegida. Esta metodología se basa en lo que su creador denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle. [5] [6]

El ciclo de vida Kimball, está basado en cuatro principios básicos:

- **Proyecto BI centrado en el negocio:** Poniendo especial atención en la identificación de los requerimientos del negocio y su valor asociado, enfocar todos los esfuerzos técnicos y humanos en desarrollar relaciones sólidas del proyecto con el negocio,

² El Grupo Kimball (<http://www.kimballgroup.com/>) ayuda a las organizaciones a aprovechar la información para tomar mejores decisiones.

enfocándolo en dar respuesta a las necesidades de explotación de información y resultados analíticos.

- **Arquitectura adecuada.** Diseñar e implementar una base de datos única, un DataWareHouse corporativo que permita integrar toda la información en un solo almacén fácil de usar y con un alto rendimiento de procesamiento. Con ello se dispondrá de una herramienta de gran utilidad para dar respuesta a los requerimientos identificados en el negocio.
- **Cargas incrementales significativas.** Realizar cargas incrementales al DWH significativas en función de los requerimientos del negocio. Para ello se estudiará la frecuencia necesaria que aporte valor al proyecto.
- **Ofrecer una solución completa.** Proporcionar todos los elementos necesarios para ofrecer valor a los usuarios de la solución BI en el negocio. Comenzando por la construcción del DWH corporativo de calidad. Y aportando a la solución final consultas dinámicas, informes, cuadros de mando y todas las herramientas necesarias para el análisis que incorporen un valor añadido.

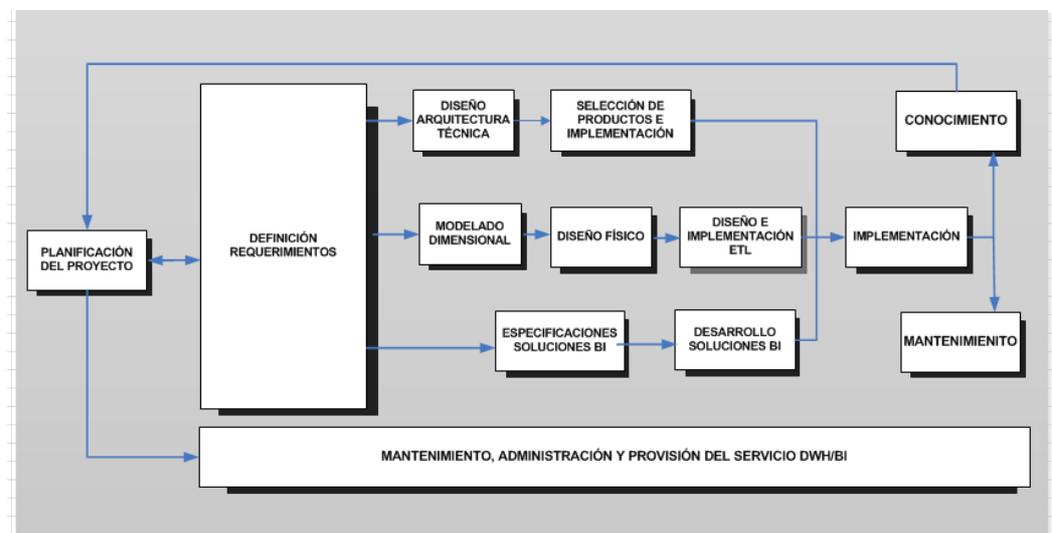


Fig 4.1. Ciclo de vida Kimball dimensional (Elaboración propia) [4] [5]

5. Diseño de DWH/BI.

En este apartado se aplicará la metodología Kimball aplicada al desarrollo de una solución BI en el Servicio Público de Empleo, cuya misión es relacionar a personas con un perfil profesional concreto con empresas que necesitan cubrir puestos de trabajo cuyas características coinciden con dicho perfil profesional.

Cada comunidad autonómica, con motivo de la transferencia de las políticas activas de empleo, gestiona de forma independiente el Servicio de Empleo.

La puesta en marcha de cada Servicio Autonómico de Empleo supone un nuevo modelo de gestión organizativo al que hay que dotarle de herramientas de inteligencia de negocios para análisis de datos y ayuda a la toma de decisiones tanto al equipo de dirección y como al personal técnico encargado de dirigir y gestionar el empleo y así poder ofrecer un mejor servicio a los demandantes de empleo, empresas y ciudadanía en general.

Todo esto se traduce en la necesidad de la creación de un Sistema de Business Intelligence que permita el análisis de la información obtenida a partir de las aplicaciones de gestión propias (ofertas de empleo, demandas de los usuarios, contrataciones y quejas) así como otras fuentes externas que se integrarán al sistema.

Los objetivos que se alcanzarán con el desarrollo del BI serán:

Primero

- Definir el catálogo de indicadores de cada área y servicio que componen la organización.
- Definir el catálogo de dimensiones generales (tiempo, edad, ámbito geográfico, sexo, nivel académico, titulaciones) y de dimensiones específicas a cada área y servicio que componen la organización.

Segundo

- Diseñar el catálogo de informes y cuadros de mando utilizando el catálogo de indicadores y dimensiones disponible para ofrecer a la organización una plataforma para el análisis y explotación de esta información corporativa.

Tercero

- Diseñar el un Cuadro de Mando Integral sólo para la dirección de la organización que incorpore los indicadores transversales más relevantes.

5.1. Modelo Dimensional.

A continuación, se explican los modelos que gobiernan el diseño del sistema BI propuesto para el Servicio Público de Empleo, siguiendo la metodología Kimball.

La representación más importante del modelo dimensional es el **esquema de estrella**. En el modelo dimensional encontramos hechos y dimensiones. Todo objeto de análisis es un hecho. Este hecho se representa en el modelo dimensional en forma de **tabla de hechos**. Los hechos son analizados a su vez, a través de las dimensiones, tantas como dimensiones participen en la descripción del hecho, que se representan en el modelo dimensional a partir de las **tablas de dimensiones**.

La representación de este esquema, tabla de hechos en el centro y tablas de dimensiones alrededor, parece dibujar una forma de estrella, origen del nombre. Las tablas de hechos tienen columnas de datos denominadas **métricas** y las de dimensiones tienen columnas que representan **atributos** y **niveles de jerarquías**. Las métricas son **indicadores** que nos permiten cuantificar los hechos.

En el proceso dimensional inicial se realiza un gráfico denominado modelo dimensional de alto nivel o gráfico de burbujas (*Bubble chart*) según Kimball

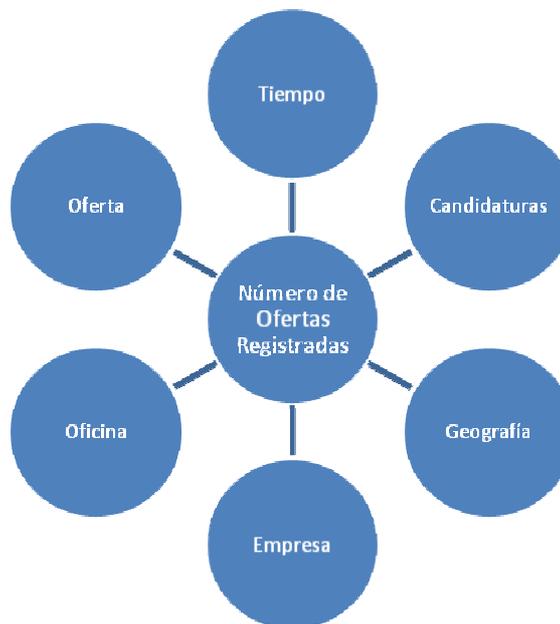


Fig 5.1. – Esquema Estrella. (Elaboración propia)

En esta fase también se realizará la identificación de atributos de las dimensiones y las tablas de hechos, consistente en completar una tabla con una lista de atributos bien formada.

Para realizar esta lista se utilizarán las plantillas que recomienda la metodología Kimball [7] para ello se colocará en las filas los atributos de la tabla y en las columnas la siguiente información:

- Características relacionadas con la futura tabla dimensional de la base de datos DWH (Target), por ejemplo tipo de datos, si es clave primaria, valores de ejemplo, etc.
- Orígenes de datos (Source, atributos de las tablas transaccionales de los operacionales).

- 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.

A efectos ilustrativos se acompaña un ejemplo de la dimensión Tiempo.

Table Name		DimTiempo										
Table Type		Dimension										
View Name		DimTiempo										
Display Name		DimTiempo										
Description		Dimensión Tiempo										
Used in schemas		DWH										
Generate script?		Y										

Schemas Page

Target												
Column Name	Display Name	Description	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Unknown Member	Example Values	SCD Type
id_Mes	id_Mes	Surrogate primary key	int				PK ID	N		-1	1, 2, 3...	
Num_Mes	Num_Mes	Número de Mes	datetime					N				
Nombre_Mes	Nombre_Mes	Nombre de Mes	datetime					N			ENE	
Año	Año	Descripción Año	datetime					N			2005	
RowIsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	char		1					Y	Y, N	
RowStartDate	Row Start Date	When did this row become valid for this member?	datetime								19/03/2004	
RowEndDate	Row End Date	When did this row become invalid? (12/31/9999 if current row)	datetime						31/12/9999	31/12/9999	1/14/1998, 12/31/9999	
RowChangeReason	Row Change Reason	Why did the row change last?	varchar		200					N/A		
AuditKey	Audit Key	What process loaded this row?	int								-1	

Fig 5.2. Tarjet. Dimensión Tiempo. [6]

Source						
Source System	Source Schema	Source Table	Source Field Name	Source Datatype	ETL Rules	Comments
Derived						
AR	GL	gl_periods	period_num	Number(15)		Restar 1, empieza en periodo 2
AR	GL	gl_periods	ENTERED_PERIOD_NAME	Varchar(15)	where period_set_name='ORONA' and ADJUSTMENT_PERIOD_FLAG='N'	
AR	GL	gl_periods	QUARTER_NUM	Number(15)		
AR	GL	gl_periods	PERIOD_YEAR	Number(15)		
Derived in ETL					Standard SCD-2	
Derived in ETL					Standard SCD-2	
Derived in ETL					Standard SCD-2	
Derived in ETL					Standard SCD-2	
Derived in ETL					Standard SCD-2	

Fig 5.3. Source y ETL. Dimensión Tiempo [6]

5.2. Modelo Físico.

La construcción del modelo físico consiste en convertir el modelo de datos lógico en un modelo de datos físico. Así, en base a la documentación del diseño dimensional realizado en los puntos anteriores se construirá el modelo físico del proyecto BI. [8]

A continuación se muestra algunos ejemplos de dimensiones y tablas de hechos y su modelo estrella, que se deberán implementar en la base de datos del DWH que permitirán construir el modelo del proyecto BI.



Fig 5.4. Modelo Físico. Dimensiones (Elaboración propia)

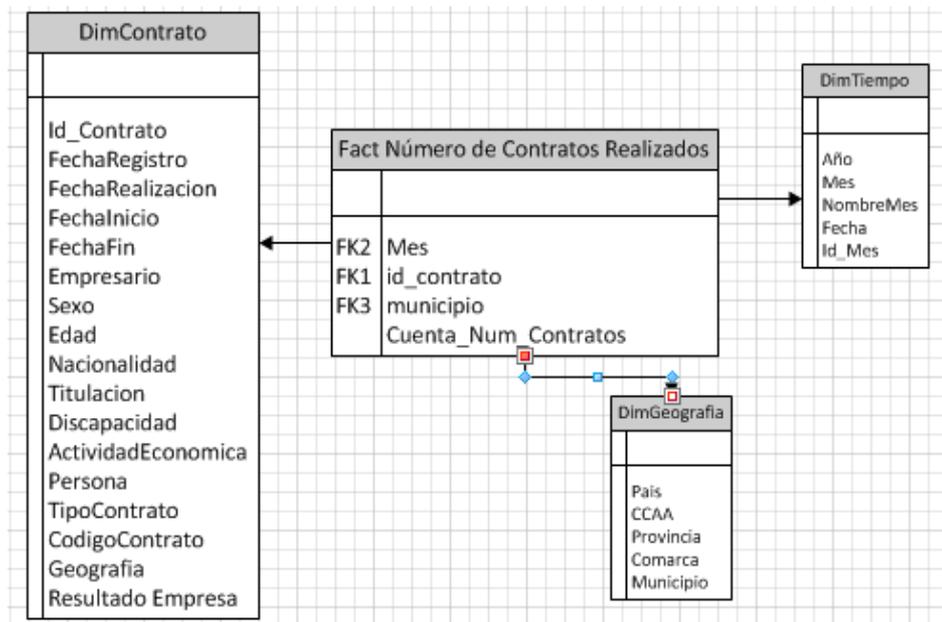


Fig 5.5. Modelo Físico. Estrella Dimensional (Elaboración propia)

6. Implementación de DWH/BI.

Con la arquitectura tecnológica de la solución BI disponible y el diseño del modelo físico realizado llega la fase de implementación de la solución BI de la organización, que comienza con la definición de los procesos ETL's.

6.1. Procesos 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 la información en el DWH.

6.2. Arquitectura Multicapa con Pentaho, Solución BI Open Source.

Pentaho³ es una Plataforma Business Intelligence de software libre para el análisis y explotación estadística que puede ser factible. Dentro del panorama de soluciones BI hay que analizar bien los pros y contras de las diferentes soluciones, ello no quiere decir que ante determinadas necesidades una solución de SW libre testada y con una visión profesional no pueda ser una alternativa válida.

En muchas ocasiones se ha puesto en duda la capacidad de un software libre como Pentaho de servir como plataforma BI de sistemas empresariales. Esto por diferentes motivos, generalmente el más extendido es su consideración como “software libre”. Esta categoría de sistemas software a veces se asocia a “incompleto”, “problemas con el soporte”, “sin garantías”, o, sencillamente, “no apropiado para entornos empresariales”

Se ha querido demostrar la aplicabilidad de Pentaho, para ello se ha utilizado la versión “Community” de Pentaho como plataforma BI empresarial. Su utilización para cualquier otro proyecto dependería de los requisitos específicos a satisfacer así como de los recursos con los que se cuenta.

Conforme a la documentación de Pentaho consultada [9], la arquitectura multicapa quedaría articulada con los siguientes elementos software:

- En la Capa de Datos: MySQL,
- En la Capa Middleware o Servidor de aplicaciones: Tomcat
- En la Capa Web: Apache

³ Pentaho Community: <http://www.pentaho.com>

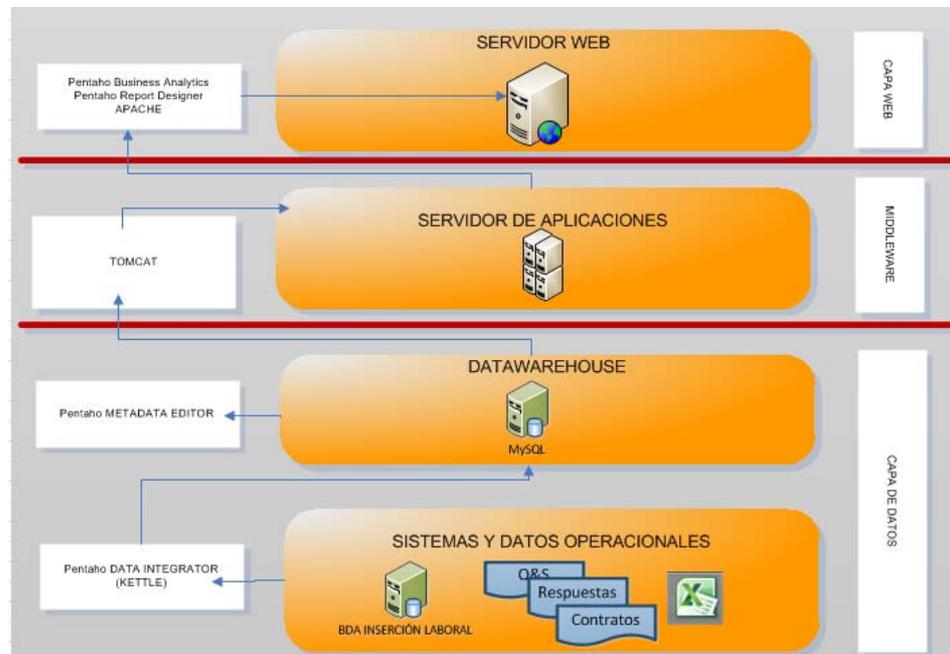


Fig 6.1. Arquitectura Multicapa con Pentaho (Elaboración propia)

Pentaho ofrece todas las tecnologías que deben componer una solución completa de BI. [9] [10]

- Integración de datos: Pentaho Data Integration [11]
- Modelado: Pentaho Metadata Editor, para la creación del modelo lógico.
- Informes y Cuadros de Mando: Pentaho Business Analytics
- Informes Entregables: Pentaho Report Designer

Además de estas herramientas fundamentales para el análisis, también cuenta con herramientas de procesamiento analítico en tiempo real, sistemas OLAP (Schema Workbench) y análisis específico de datos Data Mining y lo más actual, análisis de datos no estructurados (Pentaho Big Data Analytics).

Algunas ventajas de utilizar Pentaho como posible solución BI.

- Inversión mínima en costes de Licencias Software.
- El sistema debe implantarse de forma rápida a través de recursos propios, sin depender de terceros.

7. Resultados.

Con Pentaho como solución tecnológica, y mediante Pentaho Business Analytics (Pentaho BA) software para la construcción de informes y cuadros de mando de análisis se ha realizado la implementación del catálogo BI de contratos. Para ello se han utilizado como fuente de información, los datos públicos que de contratos del primer trimestre de 2013 obraban en poder del citado servicio de la administración pública, dentro del proyecto [OpenData](#)⁴.

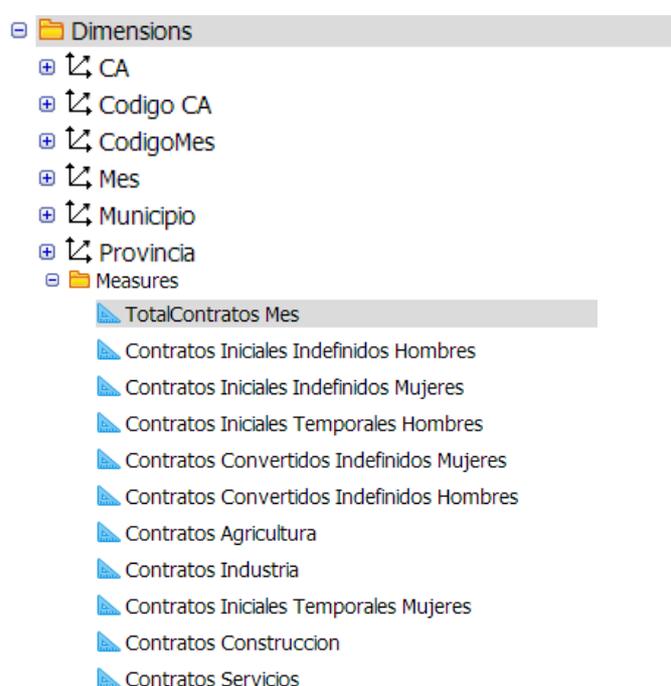


Fig 7.1. Catálogo de Indicadores y Dimensiones de Contratos 2013 (Elaboración propia)

7.1. Informes Estándar.

Los informes estándar son la parte más básica, pero imprescindible de las soluciones BI. Por lo general son informes relativamente simples, de formato predefinido, y parámetros de consulta fijos. En el caso más simple, son informes estáticos predefinidos. Los informes estándar proporcionan a los usuarios un conjunto básico de información acerca de lo que está sucediendo en un área funcional determinada de la empresa.

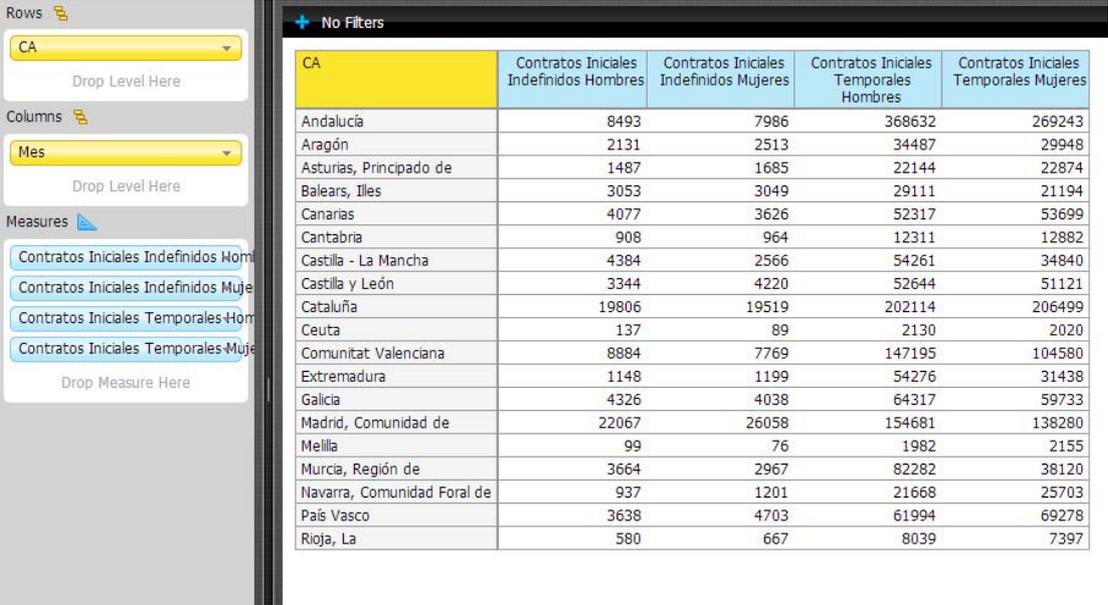
Estos reportes constituyen el catálogo de informes y cuadro de mando de la organización. Son informes que los usuarios usan día a día. La mayor parte de lo

⁴ Proyecto Opendata o Datos.gob.es (<http://datos.gob.es/datos/>) es el portal de carácter nacional que organiza y gestiona el Catálogo de Información Pública del Sector Público.

solicitado en las entrevistas durante el proceso de definición de requisitos se clasificaría como informes estándar. Por eso es conveniente desarrollar un catálogo de informes estándar en el ciclo de vida del proyecto.

A continuación, se presenta un ejemplo de un informe estándar realizado con Pentaho Business Analytics, utilizando el catálogo de dimensiones, jerarquías e indicadores de los contratos.

Este informe presenta los contratos indefinidos del primer trimestre de 2013 por sexo y por comunidad autónoma. Como puede observarse, se aprecia claramente que los contratos temporales superan a los indefinidos.



The screenshot shows the Pentaho Business Analytics interface. On the left, the 'Rows' field contains 'CA' and the 'Columns' field contains 'Mes'. The 'Measures' field contains four items: 'Contratos Iniciales Indefinidos Hom', 'Contratos Iniciales Indefinidos Mue', 'Contratos Iniciales Temporales Hom', and 'Contratos Iniciales Temporales Mue'. The main area displays a pivot table with the following data:

CA	Contratos Iniciales Indefinidos Hombres	Contratos Iniciales Indefinidos Mujeres	Contratos Iniciales Temporales Hombres	Contratos Iniciales Temporales Mujeres
Andalucía	8493	7986	368632	269243
Aragón	2131	2513	34487	29948
Asturias, Principado de	1487	1685	22144	22874
Baleares, Illes	3053	3049	29111	21194
Canarias	4077	3626	52317	53699
Cantabria	908	964	12311	12882
Castilla - La Mancha	4384	2566	54261	34840
Castilla y León	3344	4220	52644	51121
Cataluña	19806	19519	202114	206499
Ceuta	137	89	2130	2020
Comunitat Valenciana	8884	7769	147195	104580
Extremadura	1148	1199	54276	31438
Galicia	4326	4038	64317	59733
Madrid, Comunidad de	22067	26058	154681	138280
Melilla	99	76	1982	2155
Murcia, Región de	3664	2967	82282	38120
Navarra, Comunidad Foral de	937	1201	21668	25703
País Vasco	3638	4703	61994	69278
Rioja, La	580	667	8039	7397

Fig 7.2. Informe Estándar. Pentaho Business Analytics (Elaboración propia).

7.2. Informes Entregables o Predefinidos.

Se entiende por informes entregables, aquellos que son predefinidos y cuyos objetivos son:

- Generación automática de informes predefinidos a usuarios BI mediante planificación.
- Distribución de informes predefinidos a usuarios sin acceso a las herramientas BI.

Para el diseño y planificación para la generación automática de informes entregables con la herramienta Pentaho Business Analytics se utilizará la utilidad Pentaho Report Designer.

La distribución de los análisis realizados en los informes entregables, se podrán realizar en varios formatos: PDF, EXCEL o formato CSV. A continuación, se presenta un ejemplo del informe predefinido de Contratos Primer Trimestre 2013 por provincias con exportación a Excel listo para su distribución.

mayo 14, 2013 @ 11:14

Contratos Primer Trimestre 2013

Provincia	Suma Contratos
Albacete	21295
Alicante/Alacant	95198
Almería	29115
Araba/Álava	23095
Asturias	49883

Fig 7.3. Informe Predefinido. Pentaho Report Designer (Elaboración propia)

7.3. Cuadro de Mando Integral.

El Cuadro de Mando Integral diseñado en un BI es la representación gráfica de los indicadores más relevantes de una organización. A continuación se presenta un ejemplo donde se visualiza:

- Una tabla de las contrataciones del primer trimestre por sector.
- Un gráfico de tarta con las contrataciones realizadas en los tres primeros meses de 2013.
- Un gráfico de barras de las contrataciones por Comunidad Autónoma.

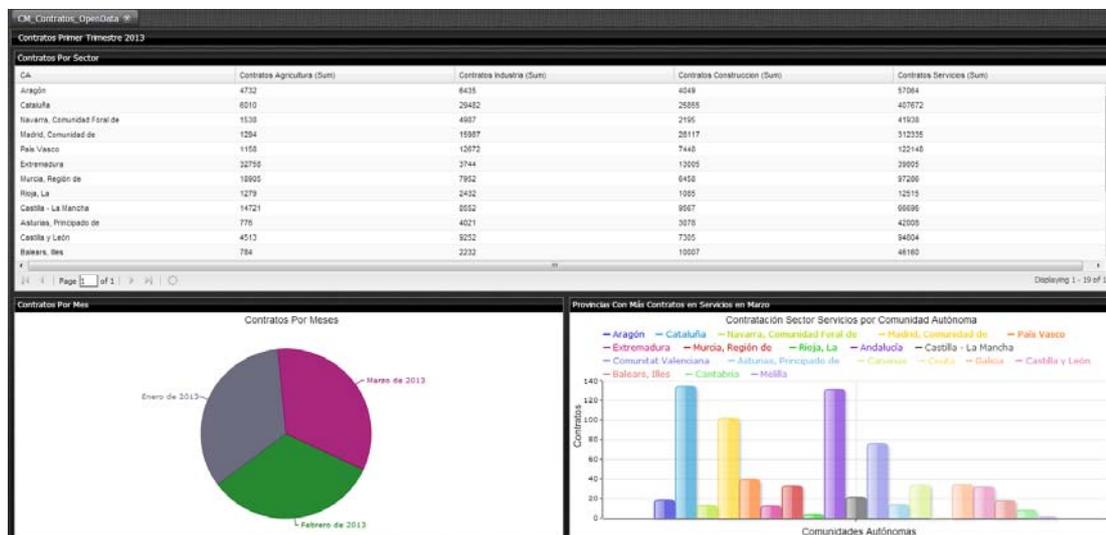


Fig 7.4. Cuadro de Mando. Pentaho Business Analytics (Elaboración propia)

8. Conclusiones.

Los cuadros de mando son la herramienta BI que sirve para representar de una forma directa y muy visual los indicadores más relevantes del catálogo definido para una organización dada.

En la implementación de una solución BI, los responsables de las diferentes áreas funcionales utilizarán los cuadros de mando para analizar y ayudar en la toma de decisiones que influyan en los planes de gestión de sus respectivas áreas.

El objetivo final del desarrollo de una solución BI en una organización se habrá conseguido cuando en el catálogo de informes de la organización exista el Cuadro de Mando Integral. Es decir, una visión gráfica y visual de todos los indicadores considerados como fundamentales para el negocio o actividad desarrollados por la organización.

Los directores y responsables utilizarán el cuadro de mando integral implementado en el BI para conocer en todo momento la situación y realizar un seguimiento cercano de los objetivos definidos en la estrategia. De esta forma, se podrá determinar si se están consiguiendo los objetivos planteados, o, por el contrario, hay desviaciones significativas que habrá que corregir.

En definitiva, se utilizará la solución BI como soporte para la toma de decisiones. Cuando esto sea posible, se habrá conseguido el objetivo final del desarrollo de una solución Business Intelligence, **transformar los datos en conocimiento**. Tal conocimiento será el utilizado para valorar las distintas alternativas involucradas en cualquier decisión o estrategia a seguir.

9. Bibliografía.

- [1] Rita L. Sallam, Joao Tapadinhas, Josh Parenteau, Daniel Yuen, Bill Hostmann.: Magic Quadrant for Business Intelligence and Analytics Platforms. Gartner, February 2014. <http://www.gartner.com/technology/reprints.do?id=1-1QLGACN&ct=140210&st=sb>
- [2] Foshay, Neil; Kuziemy, Craig. «Towards an implementation framework for business intelligence in healthcare», en INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT Volumen: 34 Número: 1 Páginas: 20-27 Fecha de publicación: FEB 2014
- [3] Chandrasekhar, U. Sch. of Inf. Technol. & Eng., VIT Univ., Vellore, India Reddy, A.; Rath, R.. « A comparative study of enterprise and open source big data analytical tools », en Information & Communication Technologies (ICT), 2013 IEEE Conference on Páginas: 372 - 377 Print ISBN: 978-1-4673-5759-3 Fecha de publicación: 11-12 April 2013
- [4] Chen, Hsinchun; Chiang, Roger H. L; Storey, Veda C. «Business Intelligence and Analytics: From Big Data to Big Impact», en MIS QUARTERLY Volumen: 36 Número: 4 Páginas: 1165-1188 Fecha de publicación: DEC 2012
- [5] Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite.: The Data Warehouse Lifecycle Toolkit. First Edition 1998.
- [6] Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker: The Data Warehouse Lifecycle Toolkit, 2nd Edition. 2008.
- [7] Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker. :The Data Warehouse Lifecycle Toolkit, 2nd Edition: Practical Techniques for Building Data Warehouse and Business Intelligence Systems. Chapter 10. Date dimension spreadsheet.
- [8] Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker.: The Kimball Group Reader; Relentlessly Practical Tools for Data Warehousing and Business Intelligence, Indianapolis, 2010.
- [9] Roland Bouman, Jos van Dongen.: Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL.
- [10] Pentaho BI Platform Community Edition (CE). Basic BI Platform Documentation. [<http://wiki.pentaho.com/display/ServerDoc2x/BI+Server+2.x-3.x+Community+Documentation>]
- [11] Matt Casters, Roland Bouman, Jos van Dongen.: Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration