

# Desarrollo de Herramientas para Warehousing en el Municipio de Junín

Mariana Adó, María Cecilia Rastelli, Ana Smail<sup>1</sup>, Rodolfo Bertone<sup>1,2</sup>

<sup>1</sup> Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Junín, Argentina

<sup>2</sup> Facultad de Informática, UNLP, La Plata, Argentina

{[mariana.ado@itt.unnoba.edu.ar](mailto:mariana.ado@itt.unnoba.edu.ar), [cecirastelli@hotmail.com](mailto:cecirastelli@hotmail.com), [anasmail@unnoba.edu.ar](mailto:anasmail@unnoba.edu.ar),  
[pbertone@lidi.info.unlp.edu.ar](mailto:pbertone@lidi.info.unlp.edu.ar)}

**Abstract.** Desde hace algunos años, el Gobierno Local de Junín promueve iniciativas de modernización para lograr una gestión más simple, transparente y eficiente.

Uno de los pilares de este proceso de modernización ha sido el desarrollo basado en E-Government. Esto implica la incorporación de Tecnologías de Información y Comunicación (TICs) en organizaciones del Estado, con el objetivo de mejorar no solo la gestión interna, sino también la interacción con los proveedores y los ciudadanos.

Estos motivos, llevaron a plantear el desarrollo del Data Warehouse (DW) para el municipio, que permita centralizar y administrar de una manera más eficiente la información de la comuna. El objetivo primario perseguido es buscar que el conocimiento generado sirva de apoyo a la toma de decisiones en la implementación de políticas públicas que permitan una mejora continua en los servicios que se brindan a la comunidad.

Este trabajo consistió, en una primera etapa, en estudiar el estado del arte de DW aplicado a E-Gobierno. Se evaluaron diferentes alternativas con el fin de definir una metodología de trabajo consistente y adecuada para la implantación en el municipio. Posteriormente, y a partir del trabajo realizado, se utilizó una herramienta Open Source (OS) con la que se desarrolló el primer Data Mart aplicado al área de Recursos Humanos.

**Keywords:** E-Government, Data Warehousing, Data Mart

## 1 Introducción

La gestión de información en las empresas es, actualmente, una herramienta clave para sobrevivir en un mercado cambiante, dinámico y global. Aprender a competir con esta información es fundamental para la toma de decisiones, el crecimiento y la gestión de las organizaciones. La disciplina denominada como Business Intelligence (BI) permite el acercamiento a los sistemas de información que ayudan a la toma de decisiones en la organización. [1]

Una *decisión* es una elección que se hace entre varias alternativas [2] y optar por aquella que se considera más correcta, es un proceso esencial en el contexto social globalizado. La mayor parte de las decisiones gerenciales carecen de estructura y conllevan riesgo, incertidumbre y conflicto, por lo que es necesario contar con información de cada una de las alternativas de solución y sus posibles consecuencias según los resultados que se esperan obtener. La importancia de la información en la toma de decisiones queda clara en la definición de decisión propuesta por Forrester, entendiendo por esta “el proceso de transformación de la información en acción” [3].

BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un data warehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. El proceso de BI incluye la comunicación de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores, productos, servicios y competidores. [4] BI es un término paraguas que abarca los procesos, las herramientas, y las tecnologías para convertir datos en información, información en conocimiento y planes para conducir de forma eficaz las actividades de los negocios. BI abarca las tecnologías de data warehousing, los procesos en el ‘back end’, consultas, informes, análisis y las herramientas para presentar información y los procesos en el ‘front end’. [5]

### **1.1 Data Warehouse**

Un Data Warehouse (DW) es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia. [6] Un DW es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis. [7]

Una Arquitectura Data Warehousing (DWA) es una forma de representar la estructura total de los datos, la comunicación, el procesamiento y la presentación de los mismos, ante los usuarios finales que disponen de una computadora dentro de la empresa. [8]

Los tipos de tecnología que se pueden aplicar son:

- OLTP: La tecnología OLTP está conformada por todos aquellos medios que soportan las operaciones diarias de la organización. No ha sido diseñado para proporcionar funciones potentes de síntesis, análisis y consolidación de los datos. (Bases de datos relacionales, archivos de texto, XML, Word, Excel, etc.)
- OLAP: La tecnología OLAP facilita el análisis de datos en línea en un DW, proporcionando respuestas rápidas a consultas analíticas complejas.

El área de datos puede ser definida como:

- ODS: (Operational Data Store) es la información de los sistemas transaccionales, con la cual se actualizará el DW. Se admite un pequeño desfase o retraso sobre los datos operacionales.
- Staging Área: sistema que permanece entre la fuente de datos OLTP y el DW. Facilita la extracción de datos desde las fuentes de origen. Minimiza la afectación a los sistemas origen, ya que los datos de los sistemas

transaccionales pueden ser extraídos desde su almacenamiento al SA, en horarios donde no haya actividad, para no comprometer los sistemas OLPT, y ser usados desde el SA en otro momento.

- Data Mart: Un Data Mart es un subconjunto de los datos del DW con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica.

## **1.2 Metodología Ralph Kimball**

Este enfoque también se lo denomina Bottom-Up, pues el DW final, consiste en la unión de los diferentes Data Marts. Esta característica hace que sea más flexible y sencillo de implementar, pues se puede generar un Data Mart como primer elemento del sistema de análisis, y luego incorporar otros que comparten las dimensiones ya definidas e incluyen otras nuevas. Con esta solución, los procesos ETL (Extract, Transform, Load) extraen la información de los sistemas operacionales, los procesan y transforman en la Staging Area, para que los datos tengan la estructura del DW; y posteriormente se realiza la carga de cada uno de los Data Mart de una forma individual, aunque siempre respetando la estandarización de las dimensiones conformadas.

Los datos transaccionales son estructurados de una forma especial para el análisis, de acuerdo al Modelo Dimensional (no normalizado), que incluye las dimensiones de análisis y sus atributos, su organización jerárquica, así como los diferentes hechos de negocio y las medidas involucradas.

Las componentes de un modelo multidimensional son las que se detallan a continuación:

- Áreas temáticas: son las distintas áreas de interés que se quieren analizar en una organización.
- Hechos: es un patrón de interés o evento dentro de la empresa que se desea analizar.
- Medidas: son cuantificadores del desempeño de un hecho.
- Dimensiones: son una colección de entidades del mismo tipo, y mediante las cuales el usuario podrá filtrar y manipular la información almacenada, más precisamente, podrá analizar los hechos.

## **2 Investigación y Comparación de herramientas OSBI**

Se investigaron diferentes herramientas OSBI. Las mismas fueron comparadas mediante el Método de Multicriterio por Scoring [9], y se optó por aquella que, a criterio de los autores, tiene las mejores características para implantar la solución al problema.

La selección de herramientas para la comparación se basó en el pronóstico publicado por la consultora Gartner en 2014, donde plantea como las tres mejores soluciones OSBI del mercado, a las herramientas: Pentaho, Jasper Sotf y Palo.

A partir de este estudio, se compararon dichas herramientas, teniendo en cuenta los siguientes criterios del BI:

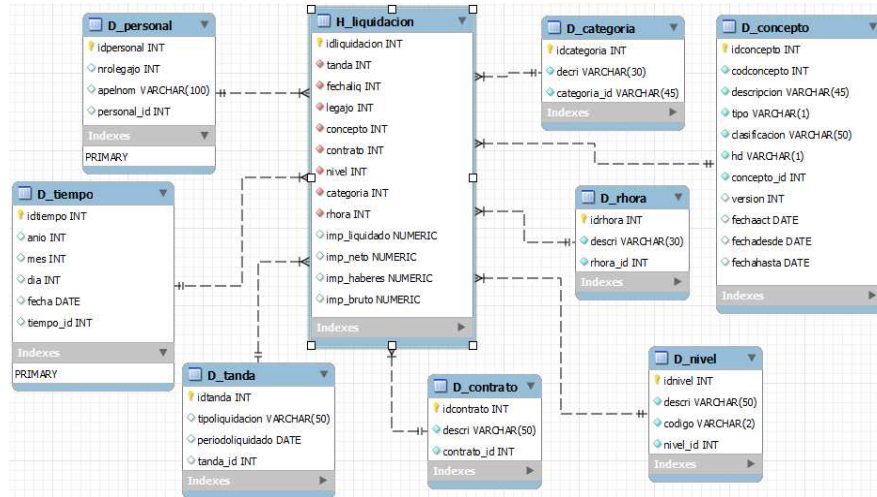
1. Calidad de las herramientas de Integración de Datos.
2. Calidad de las herramientas de Análisis OLAP de los datos
3. Calidad de las herramientas de Reportes.
4. Calidad de las herramientas de Cuadros de Mandos.
5. Calidad de las herramientas de Minería de Datos.

El estudio realizado por el Método de Multicriterio por Scoring, arrojó como resultado, que la mejor herramienta OSBI del mercado es la Suite de Pentaho BI.

### **3 Data Mart de Liquidación de sueldos**

Como parte del trabajo de desarrollo donde se implanten las técnicas analizadas, se realizó la construcción de un Data Mart de Liquidación de Sueldos para la Municipalidad de Junín. Si bien el objetivo final es el desarrollo de un DW que contemple todas las actividades del municipio se decidió comenzar por esta actividad a fin de analizar en campo el estudio realizado y proveer a las autoridades de una herramienta, al menos parcial, para la toma de decisiones. Por el ámbito en que se desarrolló el DM y los recursos destinados, se utilizó el enfoque de Ralph Kimball.

Para el caso práctico, se identifica un hecho: la liquidación. Cada transacción se traduce en una liquidación. El modelo multidimensional está compuesto por la tabla de hechos H\_liquidación y las dimensiones Tiempo, Personal, Tanda, Nivel, Categoría, Régimen Horario, Contrato y Concepto. La figura 1 representa el diagrama del modelo de datos.



**Figura 1.** Modelo Multidimensional del Data Mart de Liquidación de Sueldos.

En la tabla 1 se describen las dimensiones del modelo realizado y la Tabla 2 describe las medidas del modelo.

Nombre	Descripción
Tiempo	Año y mes correspondiente a la liquidación.
Personal	Legajo al cual se le realizó la liquidación.
Concepto	Concepto liquidado.
Tanda	Tipo de liquidación que se realizó.
Nivel	Dependencia a la que pertenece el legajo.
Contrato	Tipo de contrato del legajo.
Categoría	Tipo de categoría del legajo.
Régimen Horario	Tipo de régimen horario del legajo.

**Tabla 1.** Dimensiones del modelo.

Nombre	Descripción
Importe Liquidado	Importe liquidado por concepto.
Importe Neto	Esta medida se calcula a partir del importe y del tipo de concepto. Corresponde al Importe Neto a cobrar de bolsillo.
Importe de Haberes	Esta medida se calcula a partir del importe y del tipo de concepto. Se suman todos los importes de los conceptos que correspondan al tipo H (Haberes).
Importe Bruto	Esta medida se calcula a partir del importe y del tipo de concepto. Se suman todos los importes de los conceptos de tipo H (Haberes) y se restan los conceptos de tipo D (Descuento) y clasificación RET (Retenciones).

**Tabla 2.** Medidas del Modelo.

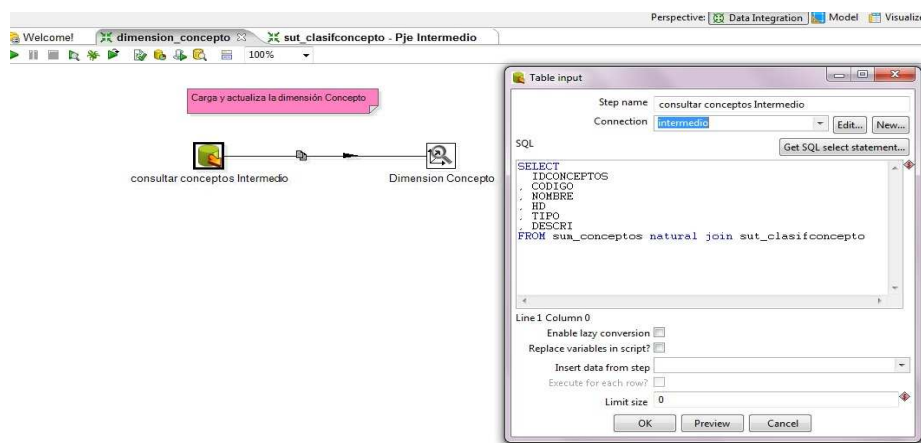
Para los procesos ETL se utiliza la herramienta PDI (Pentaho Data Integration), también conocida como Kettle.

En ella, se realizan todas las tareas de extracción de datos de la base de datos municipal, que se encuentra en Oracle, a una Staging Area (SA) en MySQL Server 5.5. En la SA se procesan y transforman los datos extraídos, para darles un formato acorde a la estructura del Data Mart es decir, se depuran, se crean las tablas dimensiones y la tabla de hechos.

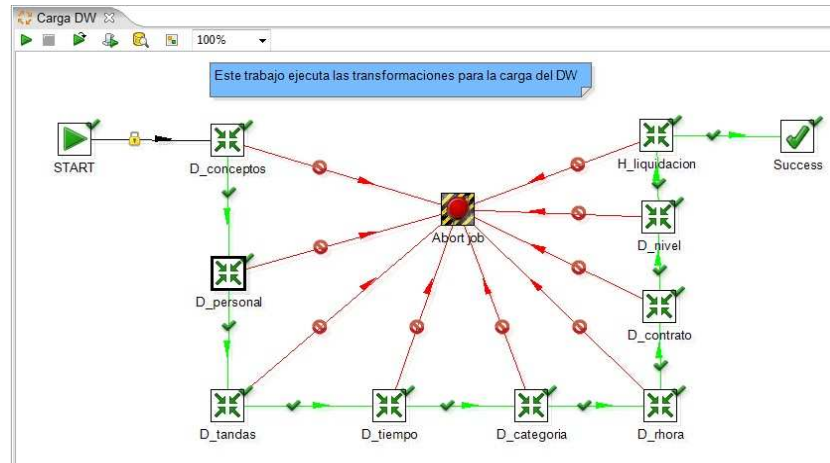
Por último, se copian las tablas de dimensiones y la tabla de hechos en el repositorio del Mart, que también se encuentra en una base de datos en MySQL Server 5.5.

La figura 2, muestra la tarea de transformación para la creación de tabla de dimensiones de conceptos en la SA. La figura 3, en tanto, muestra el trabajo final, donde se ejecutan todas las transformaciones para cargar en el repositorio del Mart, las tablas de dimensiones y la tabla de hechos.

Cuando ya se tienen las tablas componentes del Data Mart de Liquidación de Sueldos en el repositorio, se genera y se publica el esquema del mismo en el servidor de Pentaho, usando la herramienta PSW (Pentaho Schema WorkBench).



**Figura 2.** Tarea de transformación para la creación de tabla de dimensiones de conceptos en la SA.



**Figura 3.** Trabajo para la carga de las tablas de dimensiones y la tabla de hechos en el DW.

La figura 4 muestra el esquema del Data Mart de Liquidación de Sueldos en PSW.

#### 4 Resultados Obtenidos

Una vez desarrollado el Data Mart de Liquidación de Sueldos, para la construcción de vistas de análisis, se utiliza una herramienta OLAP destinada a usuarios finales de Pentaho, Saiku. Esta herramienta es una mejora de la interfaz gráfica del portal web Pentaho, que reemplaza a JPivot. Suministra a los usuarios un sistema avanzado de análisis de información. El usuario puede navegar por los datos, ajustando la visión de los mismos, aplicando filtros de visualización, añadiendo o quitando campos de agregación.

Se presentan algunos resultados obtenidos a partir del análisis de datos del DM. La figura 5 presenta la vista de la evolución de los totales por contratos del personal durante los años 2007 al 2013, en primer instancia se muestran los valores concretos y luego una representación gráfica de los mismo.

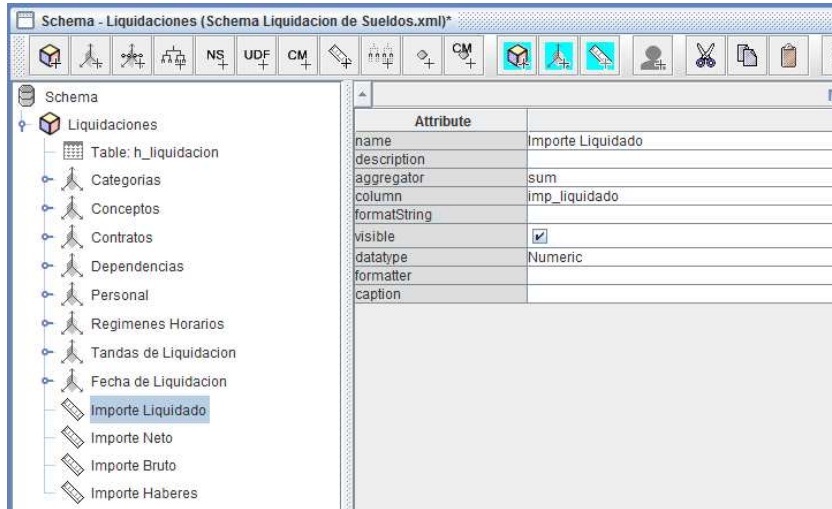


Figura 4. Esquema del Data Mart de Liquidación de Sueldos en PSW.

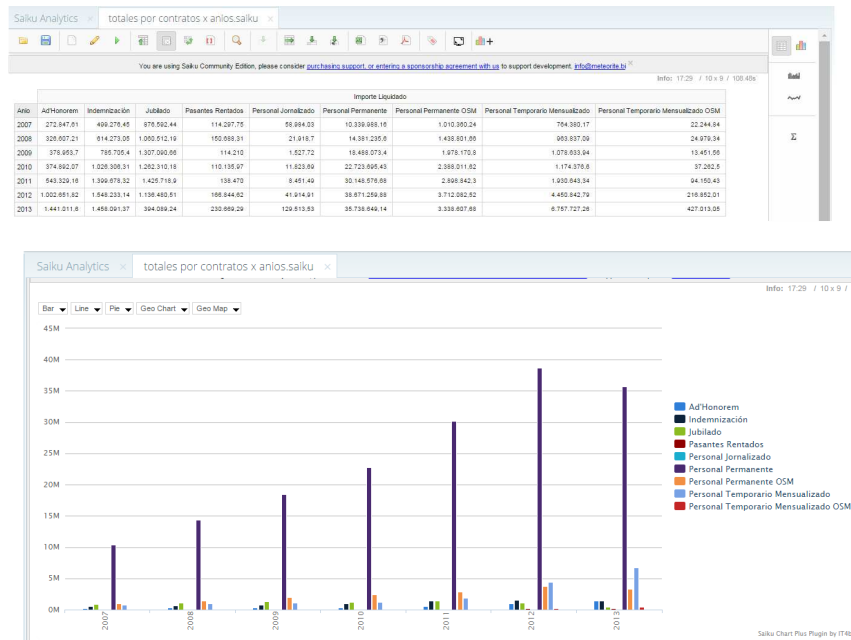


Figura 5. Representación Gráfica de resultados de evolución de contratos de empleados

La figura 6 muestra la vista de la evolución de los gastos de liquidación por categoría de la dependencia de Desarrollo de Sistemas mes a mes durante el año 2010.



		Importe Liquidado			
		DESARROLLO SISTEMAS			
Anio	Mes	JERARQ. II PROF.	JERARQUICO III	PROFESIONAL II	TECNICO I
2010	1	6.756,61	1.326	5.661,29	2.131,03
	2	7.066,61	1.333,88	5.661,29	2.131,03
	3	7.164,04	1.333,88	5.660,61	2.131,03
	4	6.912,04	1.333,88	5.702,66	2.099,06
	5	7.444,25	1.417,38	6.148,08	2.307,6
	6	10.354,84	2.811,91	11.188,16	3.311,4
	7	7.244,25	1.974,61	7.726,3	2.307,6
	8	7.980,25	1.974,61	7.743,37	2.307,6
	9	6.644,25	1.974,61	7.743,37	2.307,6
	10	6.661,56	1.974,61	7.743,37	2.307,6
	11	7.085,69	2.122,58	8.370,63	2.488,23
	12	10.531,24	3.033,87	12.158,75	3.582,34

Info: 19:01 / 6 x 15 / 6.55s

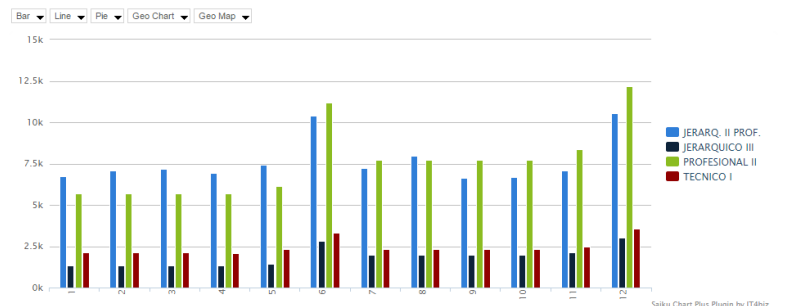


Figura 6. Gastos de liquidación de sueldos por categoría de empleados

La figura 7 representa la vista de la evolución anual por conceptos de la dependencia de Desarrollo de Sistemas desde el año 2007 al 2013.

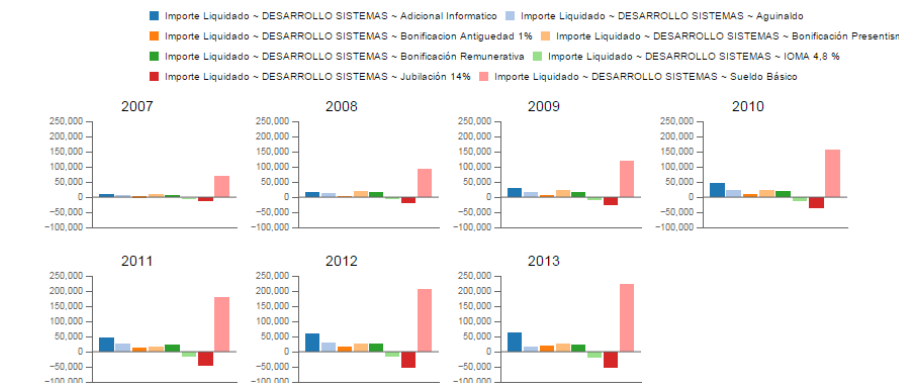


Figura 7. Evolución anual de conceptos de la Dependencia de Desarrollo de Sistemas

Los datos para este trabajo fueron extraídos del sitio <http://www.junin.gob.ar>. A modo de ejemplo se presentaron algunos resultados obtenidos, con los que el área de gestión de personal del municipio puede trabajar y realizar diversos cruces de información a fin de tomar las decisiones que políticamente se consideren oportunas.

## 5 Conclusiones

El trabajo realizado tuvo como objetivo evaluar distintas alternativas OSBI, que permitan seleccionar una herramienta y definir una metodología, que permitan desarrollar el futuro DW de la Municipalidad de Junín.

Se investigaron y compararon diferentes tecnologías, y se llegó a la conclusión de implementar el primer Data Mart con la herramienta OSBI PENTAHO.

Se implementó un Data Mart de Liquidaciones de Sueldos el cual sirve para demostrar las virtudes y ventajas del uso de estas tecnologías.

Con el desarrollo realizado, se brindó al personal jerárquico de la Municipalidad de Junín una potente herramienta para el proceso de BI, con bajos costos.

Esto brinda una alternativa hasta ahora no disponible en el proceso de toma de decisiones, con información más precisa y oportuna para articular nuevas políticas, que permitan realizar una mejor gestión y prestar mejores servicios a los ciudadanos.

## Referencias

1. Ana Smail - Taller Pentaho UNNOBA 2012. Anexo 1
2. Leon Blanck Buris. Administración en las Organizaciones: un enfoque estratégico. Centro Editorial Universidad del Valle. Facultad de Ciencias de la Administración. Cali 1990.
3. J Forrester, Dinámica Industrial. Editorial El Ateneo Buenos Aires, 1972
4. Gartner, [www.gartner.com](http://www.gartner.com), enero 2006.
5. Data Warehousing Institute. El Data Warehousing Institute ofrece educación, capacitación, certificación, noticias, e investigación y tecnologías de la información para ejecutivos. <http://tdwi.org/>
6. Bill Inmon. Considerado por muchos el padre del Data Warehouse.
7. Ralph Kimball. Considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos.
8. Nociones de Data Warehousing. <http://tdwi.org/>
9. Libro "Evaluación y Decisión Multicriterio - reflexiones y experiencias". EDUARDO MARTÍNEZ Y MAURICIO ESCUDEY. Editorial Universidad de Santiago/UNESCO, Santiago de Chile, 1998.