



## **PROYECTO “ALMACÉN” DE LA CONSEJERÍA DE OBRAS PÚBLICAS Y TRANSPORTES DE LA JUNTA DE ANDALUCÍA, COMO PLATAFORMA PARA LA INTEGRACIÓN Y EL ANÁLISIS DE LOS DATOS**

### **Miguel Ángel De La Cruz Carrasco**

Asesor Técnico de Bases de Datos

Consejería de Obras Públicas y Transportes de la Junta de Andalucía

### **Manuel Sánchez Galey**

Coordinador de la Oficina Presupuestaria

Consejería de Obras Públicas y Transportes de la Junta de Andalucía

### **Manuel Perera Domínguez**

Jefe Desarrollo de Sistemas de Información

Consejería de Obras Públicas y Transportes de la Junta de Andalucía

---

## Palabras clave

*Integración, análisis, Data Warehouse, almacén de datos, Discoverer, ETL.*

## Resumen de su Comunicación

*El objeto de esta comunicación es dar a conocer a otras instituciones y empresas la experiencia obtenida por la Consejería de Obras Públicas y Transportes en la Implantación y Explotación del Proyecto 'ALMACÉN', tanto en lo que se refiere a la integración de datos en el Almacén de Datos, como en la explotación y análisis de los datos por los usuarios a través de la herramienta Discoverer.*

*Se pretende describir el almacén de datos, resaltando objetivos y características que lo convierten en un instrumento 'estratégico' para la consejería.*

*Se relaciona los servicios/funciones de explotación y análisis de datos que contribuyen la mejora de los servicios públicos, fundamentalmente en materia de integración de datos de distintos S.I., y divulgación de información y servicios, tanto para la propia administración (para cada departamento y/o unidad orgánica), como para los ciudadanos, empresas y otras administraciones.*

*Se termina haciendo mención a la importancia de la participación de los usuarios finales en el proyecto, y se exponen estadísticas que demuestran la activa participación y aceptación de los usuarios, tanto como usuarios de consulta, como de edición de informes.*

---

# PROYECTO “ALMACÉN” DE LA CONSEJERÍA DE OBRAS PÚBLICAS Y TRANSPORTES DE LA JUNTA DE ANDALUCÍA, COMO PLATAFORMA PARA LA INTEGRACIÓN Y EL ANÁLISIS DE LOS DATOS

## 1. Introducción y objetivos

La Consejería de Obras Públicas y Transportes de la Junta de Andalucía (en adelante COPT) es un ejemplo típico de organización en el que la informatización progresiva de los procesos de negocio ha generado una miscelánea de sistemas de información de ámbito departamental y/o propósito sectorial, con poco grado de integración entre sí. A medida que se consolidan estos sistemas de información se hace cada vez más importante comunicar y/o integrar los diversos sistemas de información sectoriales (en adelante Orígenes de Datos) para poder dar respuesta a consultas y/o problemas de ámbito más general, es decir, de toda la organización. Además, la tendencia actual a proporcionar cada vez más información y servicio al ciudadano y/o a otras administraciones, requiere dotarse de una infraestructura que permita el acceso a la información con unos niveles suficientes de:

- **Transparencia:** aislar a los procedimientos de consulta y/o análisis de los datos, de la complejidad y restricciones propias de los S.I. Transaccionales que gestionan los datos en el origen.
- **Seguridad:** proporcionar mecanismos fiables y eficaces, que garanticen la seguridad en el acceso a los datos, según criterios homogéneos de identificación, ámbito de responsabilidades y/o tipología de la información.
- **Disponibilidad:** Asegurar la disponibilidad del acceso a los datos, independientemente de los detalles de rendimiento y del tipo de soporte de cada uno de los orígenes de datos.
- **Rendimiento:** Asegurar unos niveles de rendimiento optimizados y proporcionales al volumen y el detalle de la información que se solicita.
- **Flexibilidad:** garantizar la adaptación rápida y segura de los procedimientos de consulta y análisis de datos, a los cambios producidos tanto en los orígenes de datos (nuevos orígenes de datos, cambios en los ya existentes, etc.) como en las reglas de negocio inherentes a la información (cambios normativos, etc.).
- **Facilidad de acceso a los datos:** proporcionar herramientas de front-end intuitivas y fáciles de usar para el acceso a la información (cuadros de mandos, informes dinámicos, portales web, etc.).

Atendiendo a estas necesidades, la COPT ha promovido la Implantación de un Almacén de Datos Corporativo (Data Warehouse), y la construcción de diversas ‘Áreas de Negocio’ en la herramienta Discoverer, para la explotación y análisis de la información.

La elección del Almacén de Datos (Data Warehouse) se justificó fundamentalmente en la necesidad de ‘integrar’ múltiples orígenes de datos heterogéneos, tanto en la estructura, contenido y detalle de la información, como en la procedencia y soporte de los datos. La idea básica consiste en que el almacén de datos se convierta en una ‘plataforma’ fiable, escalable, y fácil de mantener, que permita obtener los siguientes objetivos:

- Integración de datos de diversas fuentes:
  - S.I. propios (TRECE, DOMO, etc.) y S.I. externos (Hacienda, etc.).
  - Información no recogida en S.I. (datos estadísticos, etc.)

- 
- Proporcionar a la Dirección una visión global y homogénea de la información:
    - Datos completos (de todos los órganos/ departamentos/S.I.).
    - Información consolidada (sin ambigüedades, incoherencias y/o lagunas de datos).
    - Información homogénea a distintos niveles de detalle/agregación.
  - Permitir la toma de decisiones:
    - Conocer la evolución temporal/histórica de la información.
    - Contrastar las previsiones con la realidad (seguimiento de planes, programas, etc).
    - Detectar situaciones de riesgo en la consecución de objetivos.
  - Mejorar los procedimientos de consulta y análisis de información:
    - Fomentar el uso de herramientas de 'query and reporting' y OLAP.
    - Complementar a los S.I. con nuevos informes y consultas más útiles, flexibles y precisas.
    - Integración de Sistemas de Información:
    - Permitiendo un crecimiento integrado de nuevos S.I.
    - Completando y potenciando la utilidad de otros S.I. existentes.
    - Albergando y proporcionando información de interés común a todos los S.I. (ejecución contable, etc.)
  - Permitir la auditoria y mejora de los procesos de negocio en los distintos S.I.:
    - Detectar carencias y problemas de información en los S.I. fuentes de datos.
    - Facilitar el uso de 'Data Mining' para detectar nuevas relaciones y reglas de negocio, que permitan mejorar la completitud y calidad de los datos.

## 2. Estructura y contenidos del almacén de datos

### Contenidos principales

El Almacén de Datos se creó inicialmente con información relativa a los distintos tipos de Actuaciones de Inversiones y Gastos de la COPT. Se pretendía hacer un seguimiento y control de las inversiones y los gastos de la Consejería y sus Empresas Públicas, teniendo en cuenta distintas versiones de la información:

- La información relativa a la Planificación de las inversiones (planes, programas, etc).
- La información relativa a la Programación de Actuaciones y Asignación Presupuestaria.
- La información relativa a los Expedientes de Gastos (contratos, subvenciones, etc).
- La información relativa a la Contabilidad Pública (expedientes y documentos contables, aplicaciones presupuestarias, fuentes de financiación, etc.).

Toda esta información debe analizarse en función de distintos criterios de organización/clasificación de los datos, como son:

- Ámbito Geográfico que se ve afectado por la inversión y/o gasto (hasta nivel municipal).
- Clasificación Orgánica responsable del control y la gestión del gasto.
- Objeto de la inversión/gasto: los datos descriptivos de las actuaciones que justifican la conveniencia de la inversión y/o gasto.
- Tercero al que se adjudica la ejecución de la inversión y/o beneficiario del gasto.
- Clasificaciones estrictamente presupuestarias y/o contables.
- Tiempo: momento en el que se planifica, programa, adjudica y se ejecuta realmente la inversión y/o gasto.

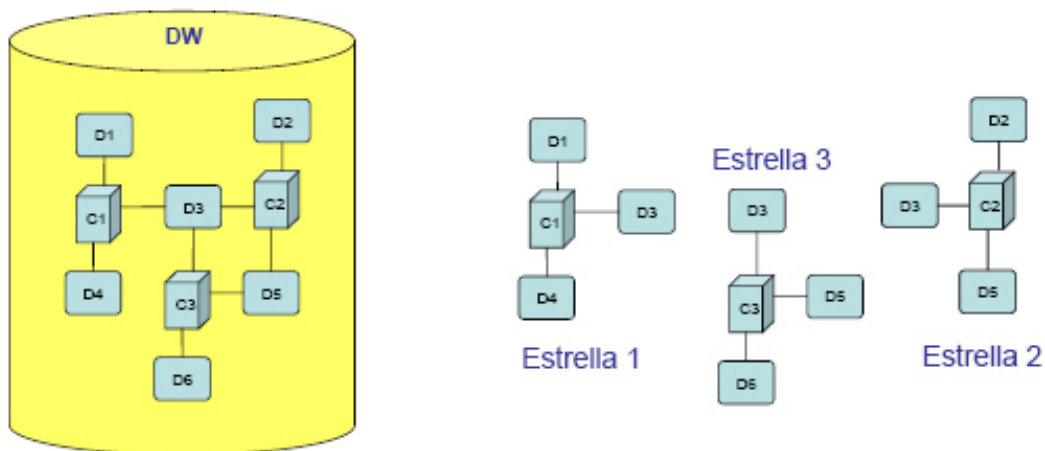


Figura 1: El almacén de datos se compone de un conjunto de 'Estrellas'

El Almacén de Datos se estructura con un modelo lógico de 'Estrellas', compuesto de varias 'Estrellas' o 'Áreas de Negocio' (para usar una terminología más conceptual y característica de las herramientas OLAP). Cada área de negocio se compone de varias 'Dimensiones' y un único 'Cubo' (o tabla de Hechos/ Medidas). Las dimensiones suelen ser comunes a varias áreas de negocio y representan los distintos criterios de organización de los datos. Los cubos almacenan las variables de medida (importes, fechas, etc.) y la relación entre las distintas dimensiones, a un nivel de detalle homogéneo y completo. La existencia de varias áreas de negocio responde a que las relaciones entre los datos se producen a distintos niveles de detalle y/o con un conjunto distinto de dimensiones.

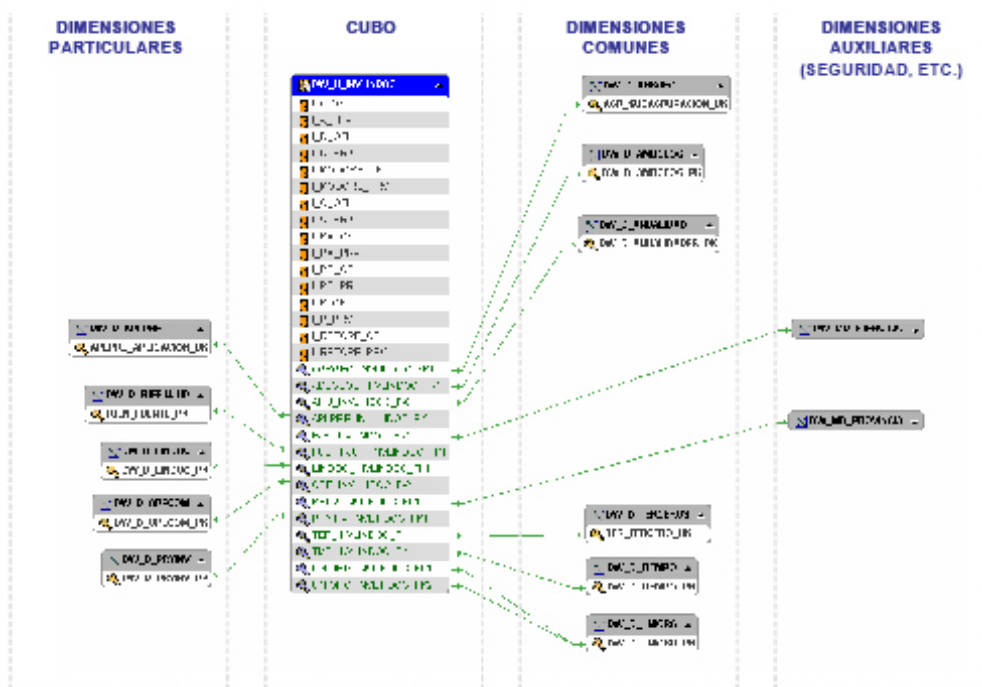


Figura 2: Ejemplo de Estrella (en Oracle Warehouse Builder).

## Metadatos

Además de la información descrita anteriormente, el almacén de datos cuenta con un repositorio de Metadatos, que permiten al usuario y al personal técnico, conocer aspectos importantísimos como son:

- Las fuentes de datos que intervienen en la información.
- Descripción del proceso de transformación/integración de los datos.
- El momento en el que se produce un cambio en las reglas de transformación/integración de los datos (producido habitualmente por un cambio en las reglas de negocio de los datos en su origen, o por la inclusión de una nuevo origen de datos).
- La estructura y contenido de las distintas áreas de negocio, para facilitar la creación de informes y consultas.
- Una descripción básica de los procedimientos de consultas e informes existentes, para fomentar la reutilización y uso compartido de estos procedimientos.

### 3. El problema de la integración de los datos

“Toda realidad puede tener distintas versiones, tantas como observadores pudieran existir”. La integración de datos es un proceso complejo que pretende obtener una única ‘realidad’ (datos fusionados) a partir de todas las ‘versiones’ disponibles (versiones de cada origen de datos).

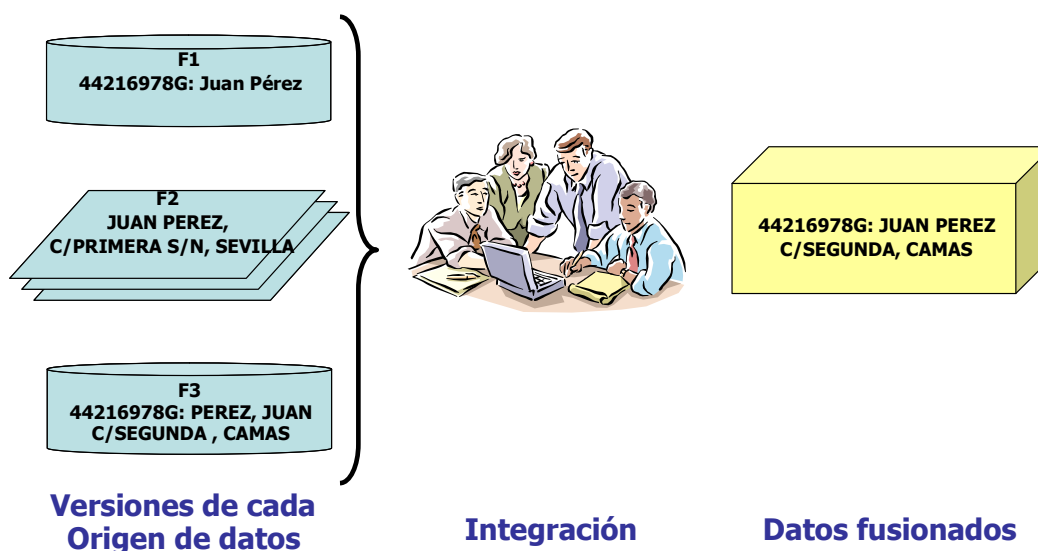


Figura 3: Ejemplo de integración de datos.

Si nos trasladamos al entorno de los sistemas de información de cualquier organización, podríamos decir que “Todo objeto puede tener distintas versiones, tantas como orígenes de datos pudieran existir”. El almacén de datos está diseñado pensando que cualquier objeto puede verse afectado por la necesidad de integración. Un objeto puede tener un único origen, pero con el paso del tiempo pueden aparecer nuevos orígenes de datos que contengan este mismo objeto, y esto obligará a implementar un proceso de integración.

---

La integración de los datos tiene que resolver problemas como los siguientes:

- Identificación de los datos: relacionar cada objeto de los distintos orígenes de datos con los objetos de otros orígenes, y con los objetos ya existentes en el almacén.
- Fusión de datos: obtener una única versión para cada objeto. Esto implica también completar las carencias de un origen de datos con los datos de otro origen.
- Resolver ambigüedades: decidir qué información es la correcta en caso de ambigüedad entre varias fuentes.
- Resolver incoherencias y errores: detectar datos que no cumplen las reglas de negocio impuestas, sustituirlos por datos correctos y/o registrar las correspondientes incidencias.

Para realizar la integración de datos de forma automática, se establecen un conjunto de reglas de integración/transформación de datos que son implementadas en la herramienta Oracle Warehouse Builder, y que deben ser conocidas, cuestionadas y revisadas en cualquier momento por los usuarios, a través de los procedimientos de consulta de Metadatos.

## 4. La construcción del almacén de datos

Para el desarrollo del almacén de datos de la Consejería de Obras Públicas y Transportes de la Junta de Andalucía se plantearon los siguientes retos tecnológicos:

- Minimizar el coste de desarrollo, y sobre todo el de mantenimiento.
- Flexibilidad ante cambios.
  - Rapidez en la implementación de cambios.
  - Fiabilidad: evitar errores colaterales producidos por los cambios.
- Alta escalabilidad, para albergar nuevos datos, nuevos orígenes de datos, y nuevas funciones de análisis, con el paso del tiempo.
- Alto rendimiento.

Se calcula que el 80 % del trabajo de desarrollo y explotación de un almacén de datos está directamente relacionado con la creación y posterior mantenimiento de los procedimientos de Extracción, Transformación y Carga. Por esta razón, es importantísimo:

- Usar una herramienta CASE especializada, que facilite esta labor.
- Contar con la participación activa de los usuarios, que deben decidir acerca de la información de interés en cada origen de datos, así como las reglas de transformación/integración de los datos.
- Establecer una metodología que permita sistematizar en lo posible la construcción de procedimientos ETL.

El éxito del almacén de datos ha sido posible gracias a que se ha tenido muy en cuenta estos tres aspectos. En especial el último punto.

## Los procedimientos ETL

Antes de que la información de los distintos orígenes de datos se carguen en el almacén de datos, tienen que pasar por varias fases conocidas por las siglas ETL: Extracción, Transformación y Carga (Load).

Al comienzo del proyecto se diseñó un Modelo para la construcción de procedimientos ETL, que se detalla a continuación:

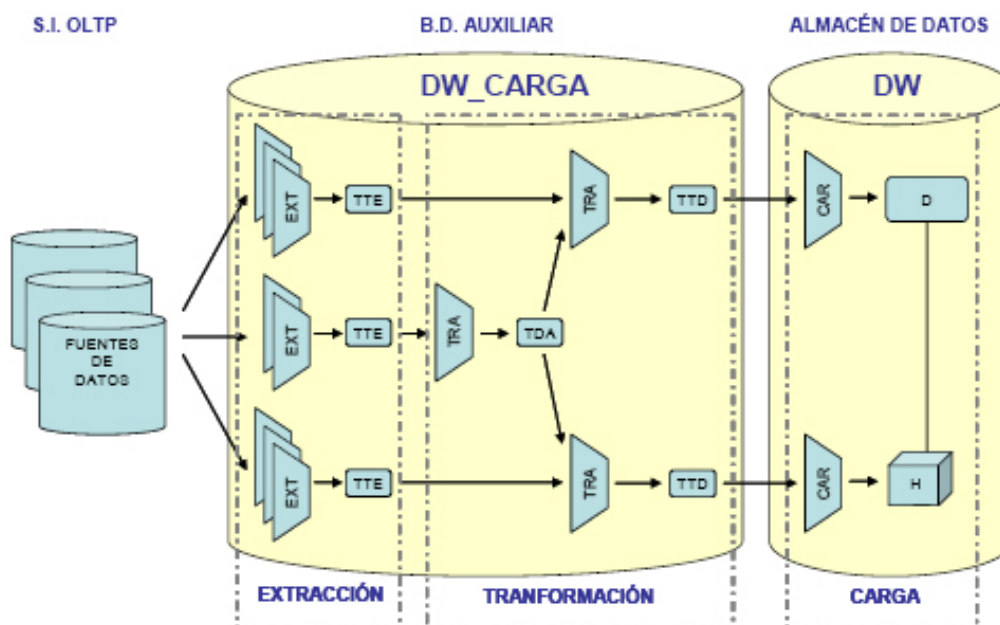


Figura 4: Estructura general del Modelo ETL

En la figura 5 se concretan los objetivos y características de cada fase y sus elementos.

## Herramientas

Para la construcción y mantenimiento del Almacén de Datos se ha elegido Oracle 9i como Gestor de Base de Datos, y la herramienta Oracle WareHouse Builder para el desarrollo y mantenimiento de los distintos elementos del almacén de datos (dimensiones, cubos, tablas, índices, etc.), y los procedimientos ETL .

Para la explotación y análisis de la información se ha usado Oracle Discoverer Desktop 10g para algunos usuarios avanzados, y Oracle9iAS Discoverer 9.0.4, para el resto de usuarios.

En la elección de estas herramientas de desarrollo y explotación se ha tenido en cuenta la voluntad homogeneizadora de la Junta de Andalucía en el uso de estándares para el almacenamiento e intercambio de información, así como la compatibilidad con los sistemas de información y los entornos de producción ya existentes.



	Objetivos	Características
EXTRACCIÓN	Obtener la situación actual de los datos en las distintas fuentes, para cada entidad del Almacén (Dimensión y/o Cubo).	La frecuencia depende de cada fuente. Generalmente DIARIA La Extracción se orienta a cada fuente de datos.
	Detectar qué datos han cambiado desde la última extracción.	Al menos un Procedimiento de Extracción (EXT) por cada fuente y entidad.
	Aislar a las siguientes fases de la complejidad de los modelos de datos de las distintas fuentes, proporcionando un modelo de datos homogéneo de cada entidad del Almacén	Es la parte más dinámica del almacén, pues tiene que ir cambiando según los cambios que se produzcan en cada fuente de datos. El resultado de la extracción se almacena en una Tabla Temporal e Extracción (TTE) La TTE permite auditar los procedimientos de transformación, y proporciona un potente instrumento de Metadatos.
TRANSFORMACIÓN	Identificar las correspondencias entre los objetos de cada fuente de datos, y con los objetos existentes en la dimensión o cubo.	La frecuencia es independiente de la extracción. Generalmente DIARIA. Es el trabajo más complicado y delicado. Suele requerir de datos de control, tablas de correspondencia de códigos, etc..
	Fusionar los datos que aportan las distintas fuentes para cada objeto de entidad en una única versión .	Se apoya en reglas de transformación consensuadas con los responsables de cada fuente y la Dirección. Es menos dinámica que la Extracción, pero está sujeta a cambios, fundamentalmente en las reglas de transformación/integración. Por esta razón es importante el mantenimiento y uso de los Metadatos.
	Normalizar los valores de cada atributo, y corregir deficiencias de datos (valores nulos, valores erróneos, etc.).	El resultado de la Transformación se almacena en una Tabla Temporal de Datos (TTD) o en una Tabla de Datos Auxiliares (TDA).
	Calcular una clave única para cada objeto de entidad que servirá de identificador en la dimensión y/o cubo	La TTD almacena los datos que van a ser cargados en la dimensión y/o cubo. La estructura es casi idéntica a la de éstos. Una vez cargados, la TTD es truncada/borrada.
	Calcular y almacenar la correspondencia entre los datos en el origen y los datos en el destino (dimensiones y cubos).	La TDA permite almacenar una versión permanente de datos ya integrados. Se usa para encapsular la lógica de extracción y transformación de algunos subconjuntos de datos que son usados en varios procedimientos de extracción y/o transformación más complejos
	En los cubos calcular el identificador de los objetos de cada una de las dimensiones relacionadas con el cubo.	
CARGA	Actualizar los datos disponibles para el usuario final en las dimensiones y/o cubos, con los datos existentes en las Tablas Temporales de Datos (TTD).	La frecuencia es independiente de las fases anteriores. Generalmente DIARIA. Es conveniente que sólo se actualicen los datos que han sufrido algún cambio. Esto mejora el rendimiento y evita problemas de ajustes y administración de la BD. En los cubos, se almacenan los incrementos/decrementos respecto a los valores anteriores. No se almacenan valores absolutos. En las dimensiones, se pueden hacer actualizaciones (UPDATE), e inserciones (INSERT). En los cubos, sólo se hacen inserciones (INSERT).

Figura 5: Descripción del Modelo ETL.

---

## Características del almacén de datos

El Modelo ETL descrito anteriormente, unido a las posibilidades y potencialidades de la herramienta elegida (Oracle Warehouse Builder) ha proporcionado al Almacén de Datos de la COPT las siguientes características:

### Construcción Sistematizada:

Los Procedimientos y Tablas tienen una estructura similar para cualquier entidad (Dimensión y/o Cubo): Cortar + Pegar + Modificar:

Se reducen tiempos de Desarrollo.

### Flexibilidad:

Incorporar nuevas fuentes de datos sólo requiere crear una nueva extracción: Las extracciones existentes no se ven afectadas.

Modificar las reglas de negocio, se realiza sólo en un sitio, de forma rápida y fácil de validar.

Se reducen tiempos de Mantenimiento

### Escalabilidad:

El almacén crece de forma ordenada. Los únicos límites son de espacio físico.

El tiempo de actualización no representa un límite, ya que cada fase puede realizarse de forma independiente y en diferido.

### Fiabilidad:

Se evitan errores colaterales con los nuevos cambios: Los posibles errores se detectan en las primeras fases, evitando que el almacén reciba datos erróneos.

### Alto rendimiento:

El Almacén de Datos (DW), es independiente a la B.D. Auxiliar: esto permite ajustes específicos para ambas bases de datos.

## 5. La explotación y el análisis de la información

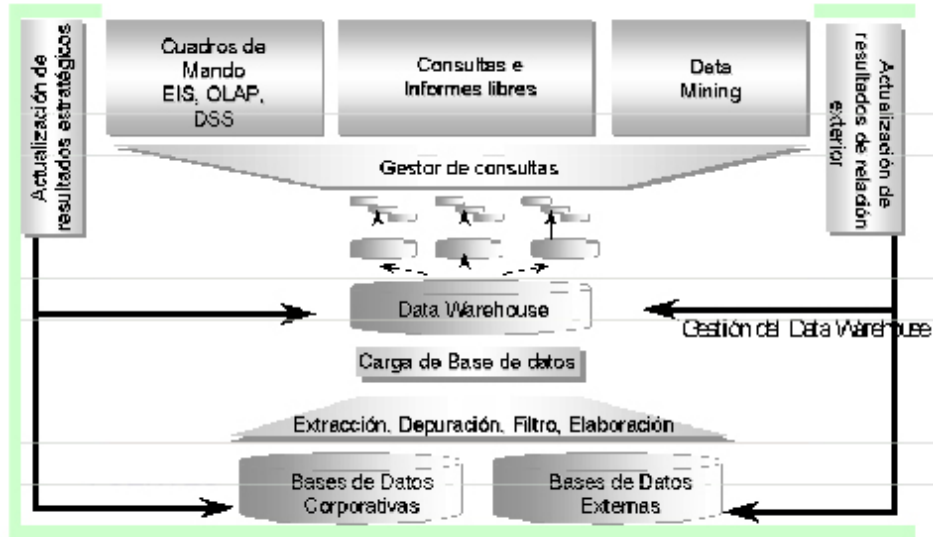


Figura 6: Contexto de un almacén de datos y flujos de información.

Al margen de la importancia de la integración de datos, el principal valor del almacén de datos radica en la capacidad de dar respuesta a un amplio abanico de funciones/servicios para la explotación y análisis de la información:

- Informes y gráficos con herramientas OLAP y de Reporting.
- Cuadros de Mando y portales Web.
- Divulgación de información a ciudadanos, empresas y/o otras administraciones, a través de los servicios de administración electrónica diseñados en cada caso.
- Proporcionar datos a cualquier S.I. para mejorar las posibilidades de sus procesos de gestión.
- Minería de Datos, para obtener nueva información oculta y/o mejorar la calidad de los datos.
- etc.

Actualmente la función de consulta más habitual es la ejecución y/o diseño de informes en la herramienta discoverer como el de la figura 7.

EJECUCIÓN PRESUPUESTARIA ACTUAL (SERVICIOS CENTRALES)											
EJECUCIÓN PRESUPUESTARIA ACTUAL (Previa) CONSEJERÍA DE OBRAS PÚBLICAS Y TRANSPORTES											
POR: CD. F. FINANCIERA.											
FECHA: 27/02/06											
Elementos de Página: Ejercicio: 2006 Capítulos: 6 y 7 Agrupación: 0											
		Crédito Definitivo	Saldo Libre	Autorizaciones	% A	Saldo Autorizaciones	Compromisos	% D	Saldo Compromisos	Obligaciones	% O
CD	F. FINANCIERA										
	CONSEJERÍA DE OBRAS PÚBLICAS Y TRANSPORTES		-298.552,00	298.552,00	0,00	,00	298.552,00	0,00	298.552,00	,00	0,00
	01: RECURSOS TRIBUTARIOS Y PROPIOS		-298.552,00	298.552,00	0,00	,00	298.552,00	0,00	298.552,00	,00	0,00
	D.G. ARQUITECTURA Y VIVIENDA	304.455.025,00	209.770.571,47	94.684.453,53	31,10	9.837.148,93	84.847.304,60	27,87	83.063.965,50	1.783.319,10	0,59
	01: RECURSOS TRIBUTARIOS Y PROPIOS	257.581.661,00	168.539.023,45	89.042.637,55	34,57	9.187.481,67	79.855.155,88	31,00	78.152.221,65	1.702.934,23	0,66
	17: F.E.D.E.R.	4.654.064,00	1.390.397,88	3.263.666,12	70,13	649.667,26	2.613.998,86	56,17	2.574.776,49	39.222,37	0,84
	18: TRANSFERENCIAS FINALISTAS	42.219.300,00	39.841.150,14	2.378.149,86	5,63	,00	2.378.149,86	5,63	2.336.987,36	41.162,50	0,10
	D.G. CARRETERAS	558.403.726,00	210.796.686,51	347.607.039,49	62,25	23.898.075,51	323.708.963,98	57,97	316.444.687,57	7.264.276,41	1,30
	01: RECURSOS TRIBUTARIOS Y PROPIOS	427.962.859,00	168.885.863,63	259.076.995,17	60,54	19.692.172,72	239.384.822,45	55,94	233.331.108,48	6.053.713,97	1,41
	17: F.E.D.E.R.	130.440.867,00	41.910.822,68	88.530.044,32	67,87	4.205.902,79	84.324.141,53	64,65	83.113.579,09	1.210.562,44	0,93
	D.G. TRANSPORTES	315.589.593,00	95.533.071,61	220.056.521,39	69,73	12.542.531,44	207.513.989,95	65,75	206.098.838,59	1.415.151,36	0,45
	01: RECURSOS TRIBUTARIOS Y PROPIOS	258.713.344,00	94.334.618,36	164.378.725,64	63,54	5.978.712,07	158.400.013,57	61,23	156.984.862,21	1.415.151,36	0,55
	17: F.E.D.E.R.	30.482.249,00	1.198.453,25	29.283.795,75	96,07	6.563.819,37	22.719.976,38	74,54	22.719.976,38	,00	0,00
	18: TRANSFERENCIAS FINALISTAS	26.391.000,00	0,00	26.391.000,00	100,00	,00	26.391.000,00	100,00	26.391.000,00	,00	0,00
	D.G. URBANISMO	20.625.521,00	10.030.347,66	10.597.173,34	36,99	49.500,00	10.537.673,34	36,01	10.507.192,10	30.401,16	0,11
	01: RECURSOS TRIBUTARIOS Y PROPIOS	24.759.099,00	14.271.706,53	10.487.392,47	42,36	49.500,00	10.437.692,47	42,16	10.407.411,31	30.481,16	0,12
	17: F.E.D.E.R.	2.647.622,00	2.647.622,00	,00	0,00	,00	,00	0,00	,00	,00	0,00
	18: TRANSFERENCIAS FINALISTAS	1.218.800,00	1.119.019,13	99.780,87	8,19	,00	99.780,87	8,19	99.780,87	,00	0,00
	INSTITUTO DE CARTOGRAFÍA	5.348.395,00	3.079.849,41	2.268.545,59	42,42	,00	2.268.545,59	42,42	2.180.376,74	88.168,85	1,65
	01: RECURSOS TRIBUTARIOS Y PROPIOS	5.348.395,00	3.079.849,41	2.268.545,59	42,42	,00	2.268.545,59	42,42	2.180.376,74	88.168,85	1,65
	S.G. DE O.L. Y URBANISMO	2.496.470,00	1.540.841,34	945.628,66	38,03	21.600,00	924.028,66	37,16	913.388,67	10.639,99	0,43
	01: RECURSOS TRIBUTARIOS Y PROPIOS	2.272.840,00	1.327.211,34	945.628,66	41,61	21.600,00	924.028,66	40,66	913.388,67	10.639,99	0,47
	17: F.E.D.E.R.	213.630,00	213.630,00	,00	0,00	,00	,00	0,00	,00	,00	0,00
	S.G. TÉCNICA	16.714.370,00	9.961.020,40	6.753.349,60	40,40	543.207,54	6.210.142,06	37,15	6.026.092,84	184.049,22	1,10
	01: RECURSOS TRIBUTARIOS Y PROPIOS	16.714.370,00	9.961.020,40	6.753.349,60	40,40	543.207,54	6.210.142,06	37,15	6.026.092,84	184.049,22	1,10
TOTAL COPT SIN DAS		1.231.623.100,00	548.421.836,40	683.201.263,60	55,47	46.892.063,42	636.309.200,18	51,66	625.533.114,09	10.776.086,09	0,87

Figura 7: Ejemplo de Informe de Discoverer con del Almacén de Datos.

## Desarrollo e implantación paulatina

El desarrollo e implantación del almacén de datos, y en especial los distintos servicios de explotación y análisis de datos se han ido realizando paulatinamente, siguiendo un ciclo de vida propio del prototipado:

1. Se diseña un área de negocio.
2. Se analizan los datos con informes genéricos e informes de auditoría orientados a validar los datos.
3. Se identifican errores, carencias, relaciones ocultas de los datos que pueden incorporarse a los procedimientos de extracción/transформación.
4. Se acuerdan aspectos de nomenclatura, descripciones, aspectos de presentación, criterios de agregación, etc. que se modelan en las herramientas de administración de Discoverer y/o Oracle Warehouse Builder. Además se asegura su correcta actualización en el repositorio de Metadatos.
5. Se realizan los cambios en el área de negocio, y se vuelve a reiterar en el punto 2.

## El usuario: una pieza indispensable

Sin duda alguna, es imposible acometer un proyecto de estas características sin contar con la participación de los usuarios. Esta participación debe acompañar toda la vida del proyecto, desde el análisis de los distintos orígenes de datos, hasta la confección de los informes, consultas, etc.

En la fase final de explotación y análisis de la información, los usuarios deben ser los verdaderos protagonistas, confeccionando y distribuyendo directamente los informes que consideren necesarios. Para que

esto sea posible es importante elegir una herramienta de análisis de datos fácil de usar y de aprender por los usuarios 'no informáticos'. En nuestro proyecto elegimos Discoverer como herramienta de análisis. No obstante, se está trabajando en el uso de herramientas de análisis más sofisticadas para satisfacer las demandas de análisis de algunos usuarios más expertos.

## Estadísticas del uso de los usuarios

En el proyecto ALMACÉN a principios de 2005 se diseñó un Plan de Formación destinado a conseguir que al menos un usuario de cada Centro Directivo y cada Delegación Provincial tuviese los conocimientos y habilidades necesarias para la edición de informes. Se han realizado cuatro ediciones del curso de formación a lo largo de un año. En la figura 8 puede apreciarse el número de usuarios con perfil de edición que usa el almacén de datos. De los 4 iniciales (técnicos de desarrollo y usuarios responsables del proyecto) a los 43 de finales de febrero de 2006.

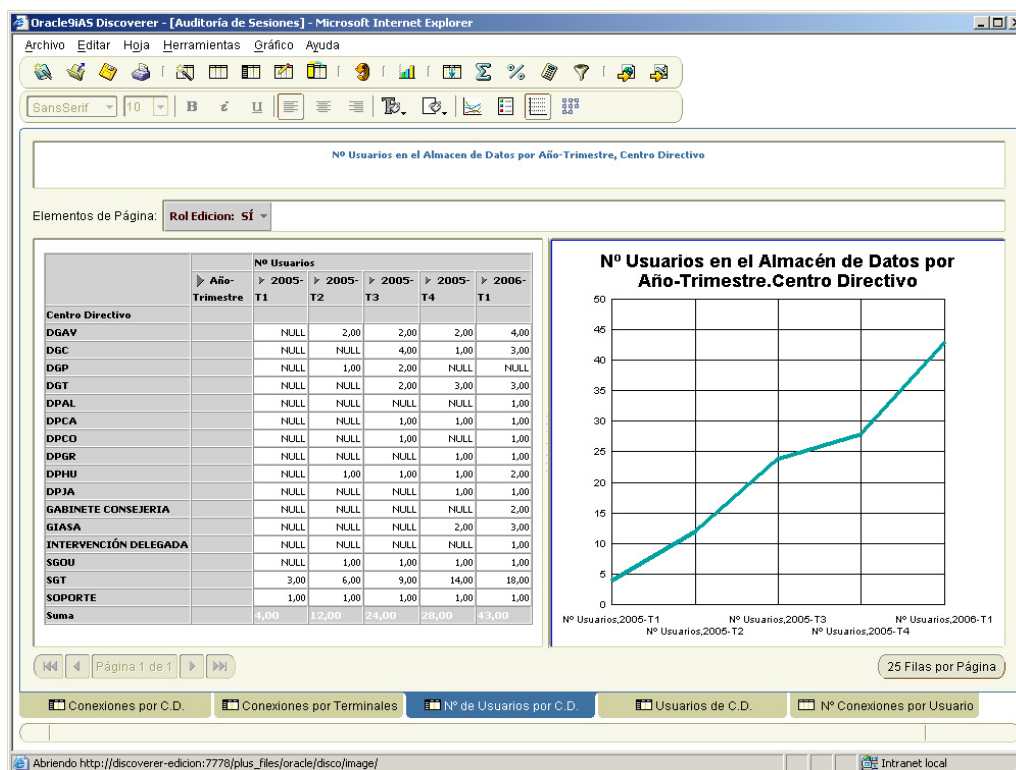


Figura 8: Informe de Auditoria sobre el uso de los usuarios de Edición.

Además del número de usuarios que usan el almacén de dato, que en la actualidad llega a 63 (teniendo en cuenta los perfiles de Consulta y Edición), se ha producido un incremento significativo en la frecuencia de uso de los informes de discoverer. En la figura 9 se analiza la evolución del número de sesiones (conexiones a la BD) realizadas en el último año.

En el transcurso de un año se ha multiplicado por 10 el número de conexiones a la base de datos. En el último trimestre (enero y febrero), la media de conexiones por día laborable es de 60 conexiones, lo que indica que la frecuencia de uso de cada usuario se acerca a 1. Es decir, de media, cada usuario del almacén de datos usa el sistema todos los días.

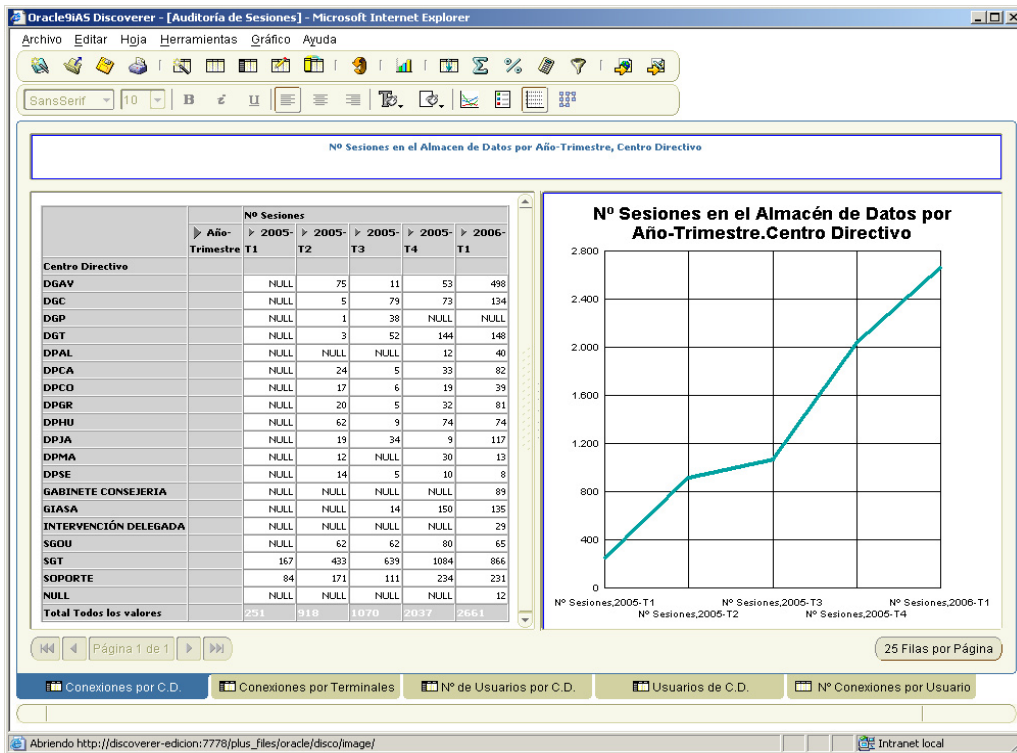


Figura 8: Informe de Auditoria sobre la frecuencia de uso de los usuarios.