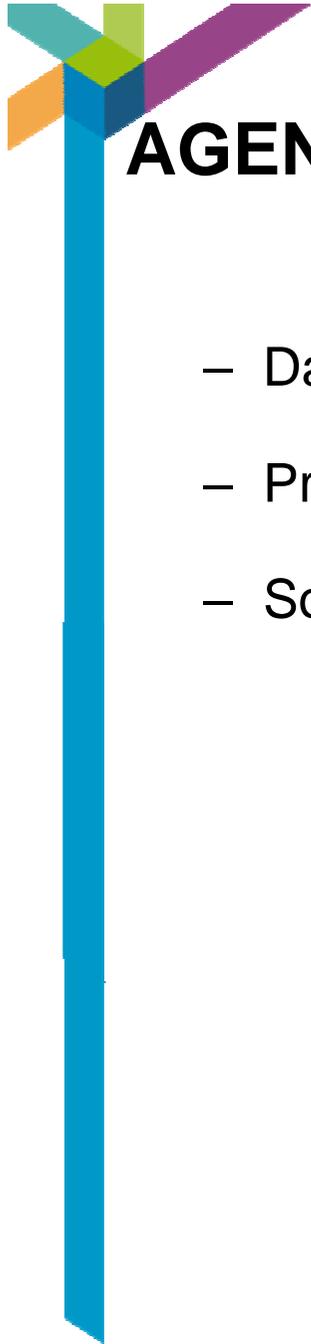




# Data Warehousing: el reto de extraer el valor de los datos

Luis Reina Juliá  
[luis\\_reina@es.ibm.com](mailto:luis_reina@es.ibm.com)  
IBM Software Group  
Arquitecto de Datos



# AGENDA

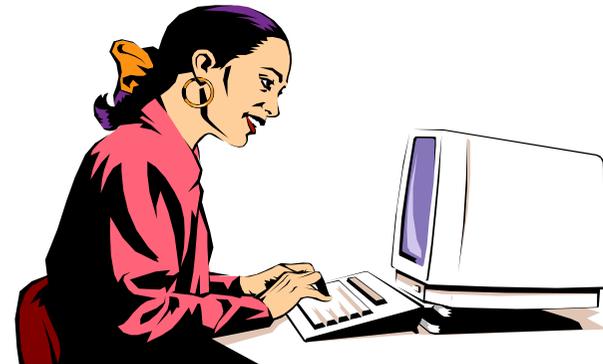
- Data Warehouses
- Problemáticas típicas de Data Warehouse
- Soluciones de IBM

# ¿PARA QUÉ SIRVE UN DATA WAREHOUSE?

El propósito del Data Warehouse: proporcionar información de la empresa útil y comprensible a los usuarios finales

- Necesidades de negocio

Necesitamos sistemas **Ayuda a la toma de decisión** que traten información histórica desde una perspectiva de negocio, no técnica y que analicen la información desde diferentes perspectivas.



- Los Sistemas **Transaccionales** no son adecuados para esto
  - Los sistemas transaccionales están orientados a la transacción, no a la consulta.
  - Las consultas tardan mucho tiempo en ejecutarse, los usuarios pueden dañar el rendimiento de los sistemas transaccionales.
  - Dependemos de perfiles técnicos.



## PROBLEMAS TÍPICOS Y RETOS DE LOS WAREHOUSE ACTUALES

### Problemas Típicos

- Rendimiento Consultas.
- Rendimiento de las Cargas (batch).
- Periodicidad de Carga del Warehouse.
- Antigüedad del dato.
- Alto Coste.
- Escalabilidad.
- Inconsistencia del Dato.
- Baja satisfacción del usuario final.
- Falta de Información.
- Información no Fiabile.
- Falta de Funcionalidad.
- .....

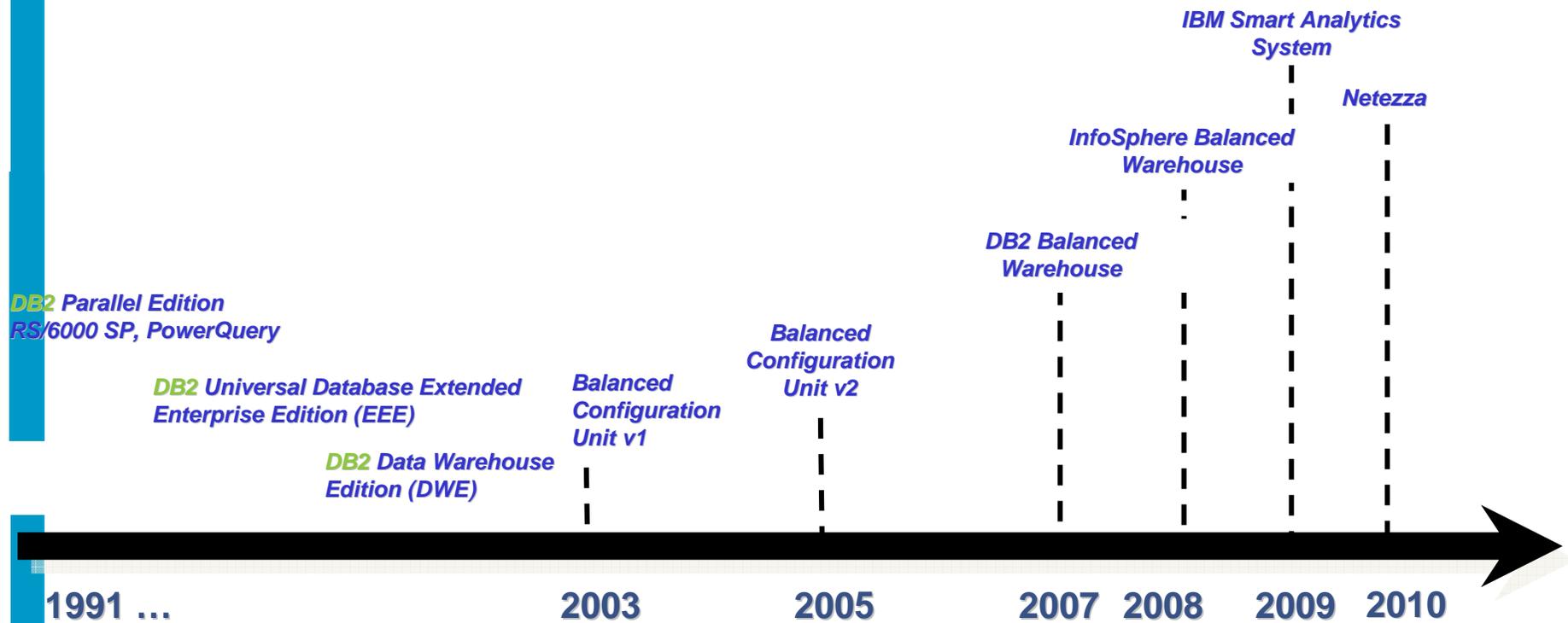
**83%**

De los CIOs citó a "*Business Intelligence and Analytics*" como parte de su plan para mejorar la competitividad.

**35%**

De los clientes estudiarán, dentro de los próximos 3 años, reemplazar su data warehouse actual con una solución de data warehouse pre-integrado, sólo el 14% lo tienen hoy.

# IBM MÁS DE 20 AÑOS DE AVANCES EN DATA WAREHOUSE



# SOLUCIONES DE IBM PARA DATA WAREHOUSE

## IBM Netezza



El verdadero appliance

## IBM Smart Analytics System



Sistema Integrato Flexible

## IBM InfoSphere Warehouse



Solución personalizada

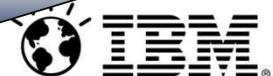
Aceleradores de Warehouse

Portafolio Information Management  
(Information Server, MDM, Streams, etc)

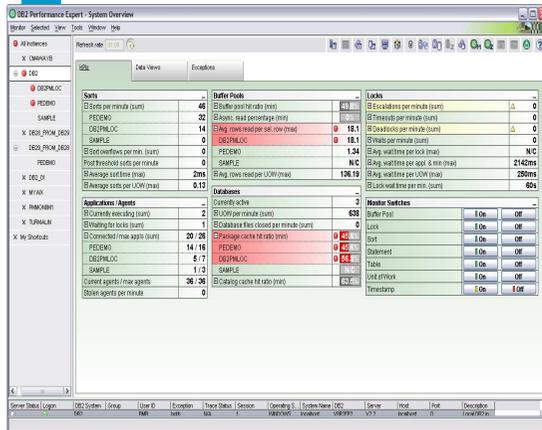
**Simplicidad**

*La combinación óptima entre simplicidad y flexibilidad*

**Flexibilidad**



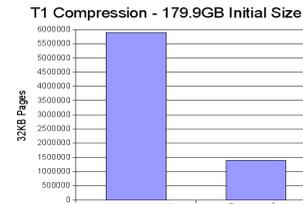
# COMPONENTES DE INFOSPHERE WAREHOUSE



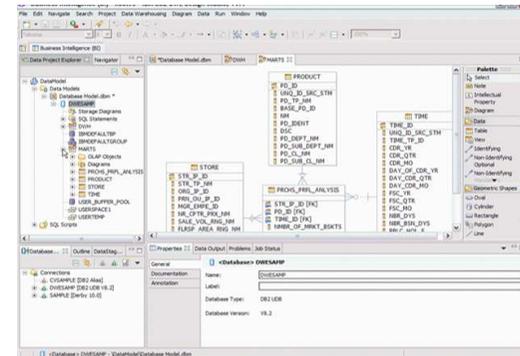
Monitorización y Alertas



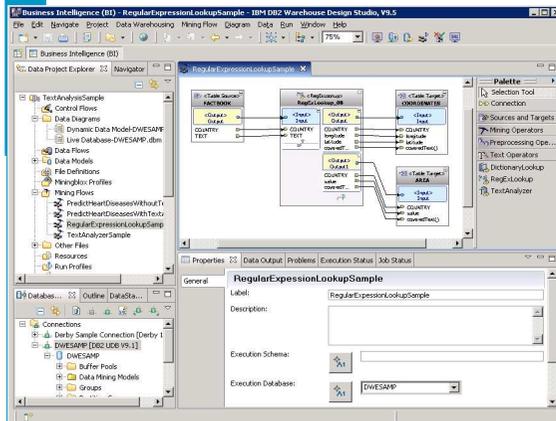
Work Load Manager



Compresión de Datos



Design Studio



SQL Warehouse (ELT)

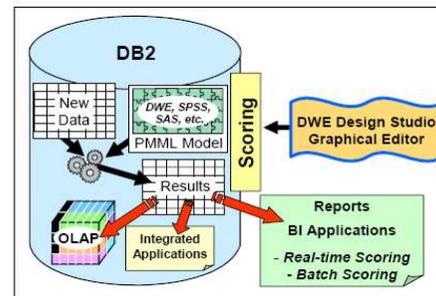


Figure 7-3 DWE Scoring architecture

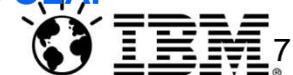
Minería de Datos



Offi<sup>d</sup> Party  
Universal Access



InfoSphere Warehouse  
Servidor de Cubos OLAP



# IBM EL INVENTOR DE LAS BASES DE DATOS



## Bases de Datos Jerárquicas

- Modelo Jerárquico.
- Proyecto Lunar Apolo .
- Base de Datos IMS



## Bases de Datos Relacionales

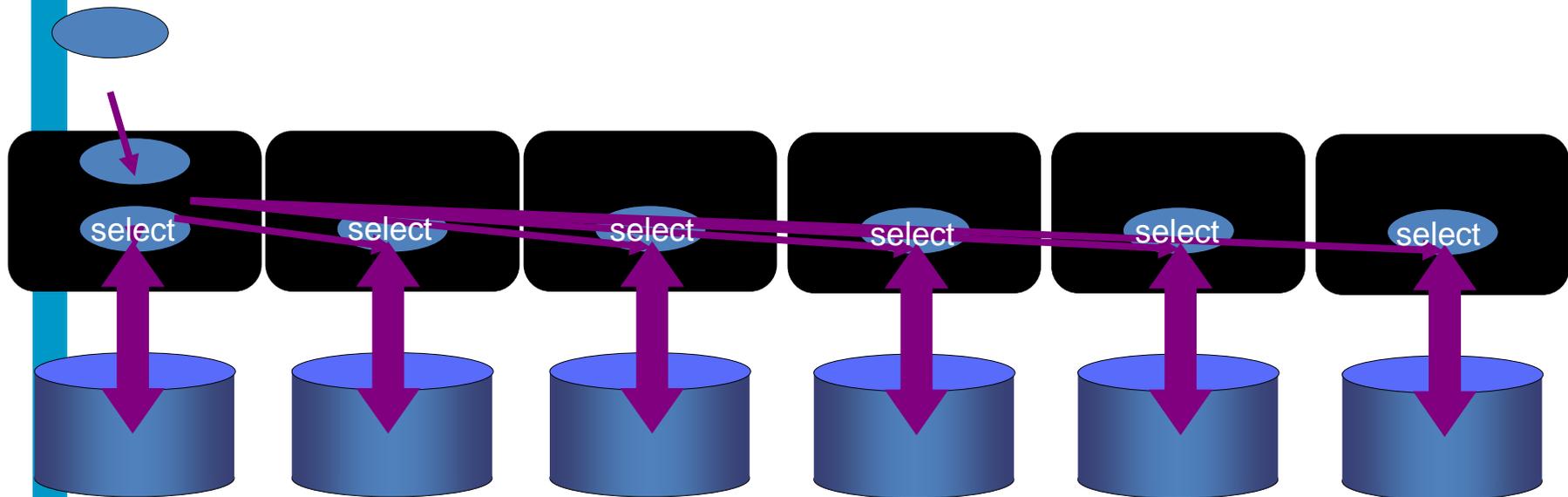
- Modelo Relacional
- F. Codd: 1970
- DB2, Informix, Sybase, Oracle, SQL Server, etc.



## DB2 ARQUITECTURA “DIVIDE Y VENCERÁS” PERMITE ESCALABILIDAD LINEAL

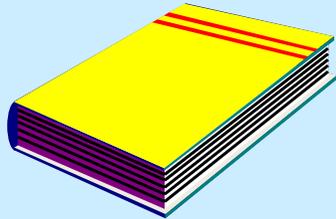
### Arquitectura «*Shared Nothing*»

- Base de datos dividida en múltiples particiones
- Cada partición puede correr en un nodo (máquina) distinta.
- Cada servidor de partición tiene recursos dedicados.
- Existe un Procesamiento Paralelo en todas las particiones.
- **Las aplicaciones y los usuarios ven una imagen única.**



# MÉTODO “TRADICIONAL” DE ACCESO A BASES DE DATOS

Necesito todos los nombres y direcciones de las personas que viven en el barrio de Chamartín



1,000 páginas  
de listín telefónico

X



1 persona puede mirar  
1 pág por minuto

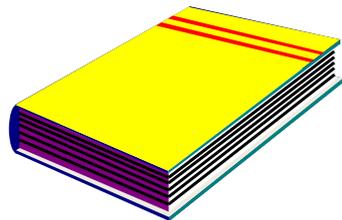
=



1,000 minutos  
En completar la tarea

# CÓMO LO HACE DB2

Procesamiento Paralelo de Bases de Datos



1,000 páginas de Listín Telefónico

Kjf	Kkj
Kjflk....1234 kdsjk....1234	Kkj
Kjflk....1234 kdsjk....1234	234
Kjdd....1234 kdsjk....1234	234
Kjflk....1234 kvvkj....1234	234
Kjflk....1234 kvvkj....1234	234
Kjflk....1234 kvvkj....1234	234

Kjf	Kkj
Kjflk....1234 kdsjk....1234	Kkj
Kjflk....1234 kdsjk....1234	234
Kjdd....1234 kdsjk....1234	234
Kjflk....1234 kvvkj....1234	234
Kjflk....1234 kvvkj....1234	234
Kjflk....1234 kvvkj....1234	234

Kjf	Kkj
Kjflk....1234 kdsjk....1234	Kkj
Kjflk....1234 kdsjk....1234	234
Kjdd....1234 kdsjk....1234	234
Kjflk....1234 kvvkj....1234	234
Kjflk....1234 kvvkj....1234	234
Kjflk....1234 kvvkj....1234	234

100 Grupo de 10 páginas

Necesito todos los nombres y direcciones de las personas que viven en el barrio de Chamartín



100 Personas mirando 1 pág por minuto

¿Cuánto tiempo tardo?



# CLIENTE DE IBM INFOSPHERE WAREHOUSE

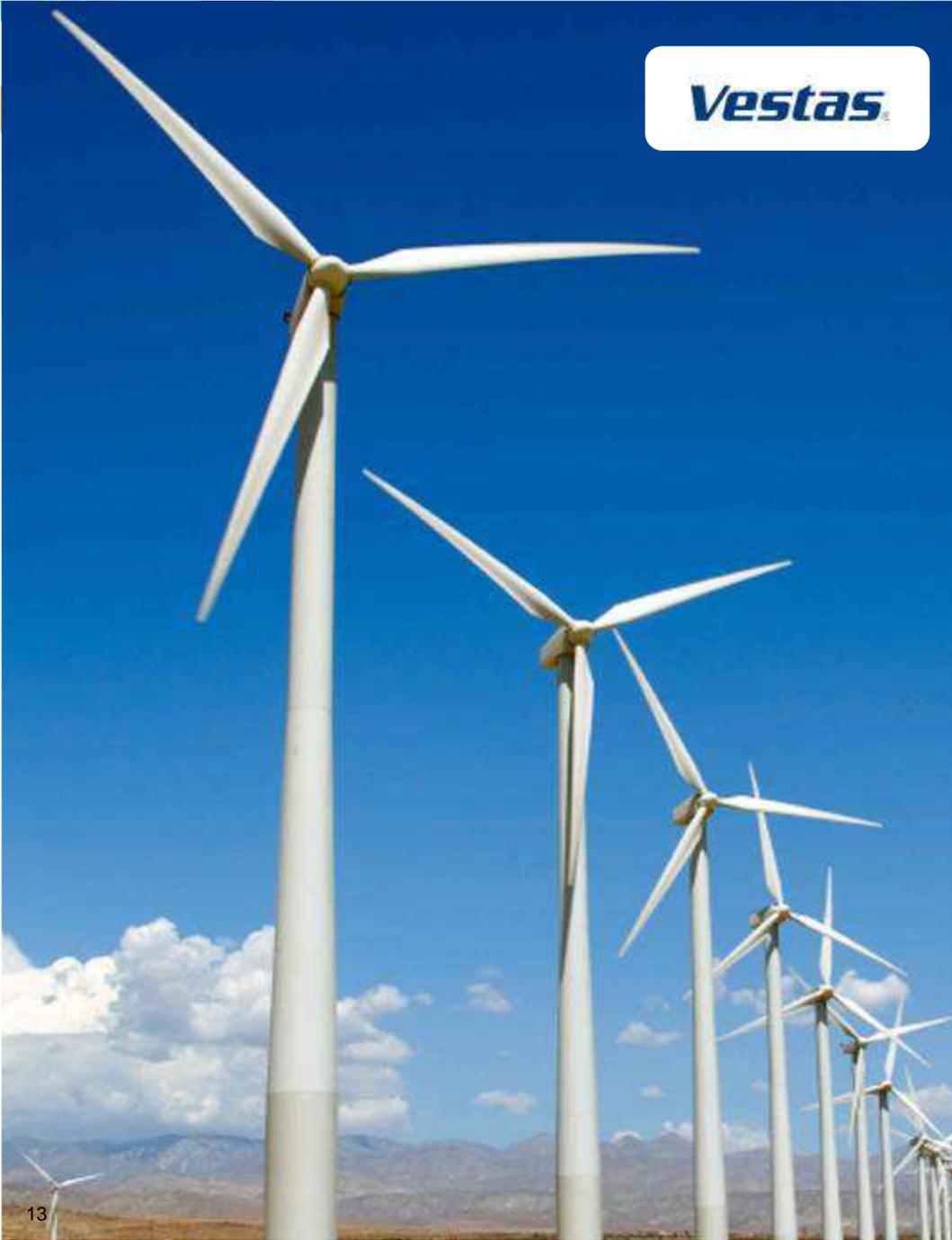
## Entorno/Problemática

- Cliente de la Industria Financiera que opera en +30 países con sede en Nueva York.
- Soporte de grandes volúmenes de datos (>200TB).
- Trabajar con clusters e Linux con múltiples servidores pero teniendo una visión única de los datos.
- Reducción de Costes.

## Solución/Beneficios

- InfoSphere Warehouse software 9.7 corriendo sobre servidores intel/Red Hat.
- Compresión de datos + Índices supuso una reducción de tres veces el volumen de datos existente.
- Visión única de las aplicaciones que atacan a una única base de datos.
- Escalabilidad cercana a la linealidad añadiendo nuevos servidores.
- Ahorro de un 40% sobre la solución existente de la competencia.





**Vestas**

## Vestas optimiza sus inversiones de capital en base a **2.5 Petabytes** de información

Capacidades utilizadas:

***InfoSphere BigInsights***

***InfoSphere Warehouse***

- Modelización del tiempo atmosférico para optimizar el emplazamiento de las turbinas, maximizando la generación de energía y su duración.
- Reducción del tiempo necesario para identificar el emplazamiento de turbinas de semanas a horas.
- Incorporación de 2.5 PB de flujos de información estructurada y semi-estructurada.
- Expectativa de crecimiento del volumen de datos hasta 6 PB.

# SOLUCIONES DE IBM PARA DATA WAREHOUSE

**IBM  
Netezza**



**El verdadero appliance**

**IBM  
Smart Analytics System**



**Sistema Integrato Flexible**

**IBM  
InfoSphere Warehouse**



**Solución personalizada**

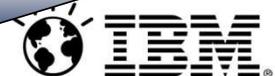
**Aceleradores de Warehouse**

**Portafolio Information Management**  
*(Information Server, MDM, Streams, etc)*

**Simplicidad**

*La combinación óptima entre simplicidad y flexibilidad*

**Flexibilidad**





# IBM Smart Analytics System

Todo lo que necesita para su Analítica de Negocio no sólo un *appliance* de data warehouse...

Incluye:

1. **Software Analítico**
2. **Potente Data Warehouse**
3. **Hardware y Servicios**



*Transformando la información en comprensión del negocio*

**3x más Rápido**

**50% menos espacio**

**Record Mundial en Rendimiento**



# IBM SMART ANALYTICS SYSTEM: SISTEMA ENSAMBLADO VERSUS UN SISTEMA POR PIEZAS



**Vs.**



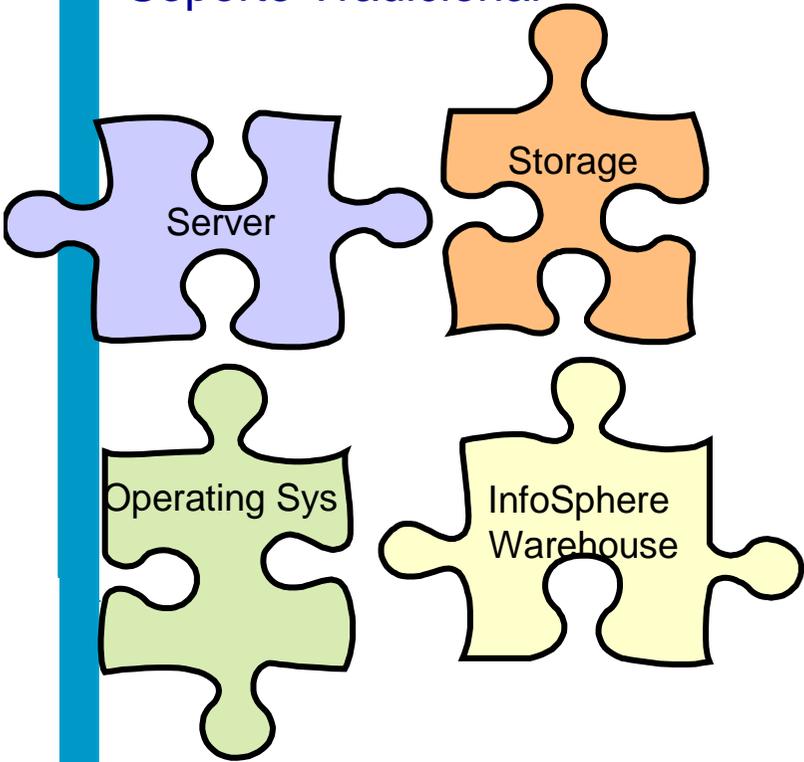
- ✓ **Menor Coste**
- ✓ **Ahorro de Tiempo**
- ✓ **Menor Riesgo**



# IBM SMART ANALYTICS SYSTEM:

## SOPORTE INTEGRADO DE LA SOLUCIÓN

Soporte Tradicional



A cada parte se le da soporte por separado

VS

IBM Smart Analytics



Soporte Integrado





# Ciente de IBM Smart Analytics

## Entorno/Problemática

- Cliente de la Industria de *Retail* español.
- Necesidad de alto rendimiento de generación informes en periodos determinados.
- Escalabilidad alta.
- Varios procesos *batch* diarios (necesidad de concurrencia batch/informes)

## Solución/Beneficios

- IBM Smart Analytic 5600
- Solución Modular: 16 módulos de datos (>200 cores, >2TB memoria)
- Barrido de datos de 44 GB/seg
- Visión única de la base de datos (tareas administrativas).
- Software de Gestión de las cargas de trabajo concurrentes (work load manager) permite concurrencia batch/informes.
- Despliegue rápido de nuevos proyectos.



# SOLUCIONES DE IBM PARA DATA WAREHOUSE

## IBM Netezza



El verdadero appliance

## IBM Smart Analytics System



Sistema Integrado Flexible

## IBM InfoSphere Warehouse



Solución personalizada

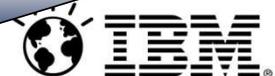
**Aceleradores de Warehouse**

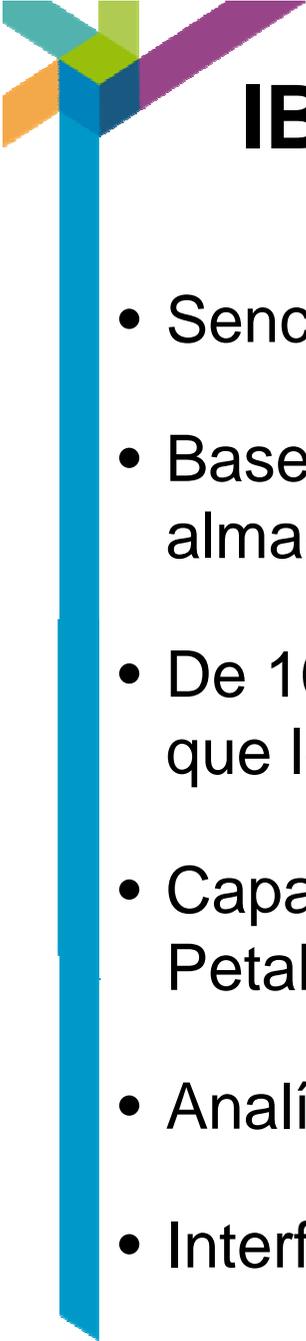
**Portafolio Information Management**  
(Information Server, MDM, Streams, etc)

**Simplicidad**

*La combinación óptima entre simplicidad y flexibilidad*

**Flexibilidad**





# IBM NETEZZA

- Sencillez de un *Appliance*.
- Base de datos, servidor y almacenamiento integrados.
- De 10 a 100 veces más rápido que los sistemas tradicionales.
- Capacidad de escalar a PetaBytes.
- Analítica avanzada.
- Interfaz estándar.





## Barnes & Noble ayuda a los proveedores al seguimiento de las ventas e inventario en tiempo real

Capacidades utilizadas:

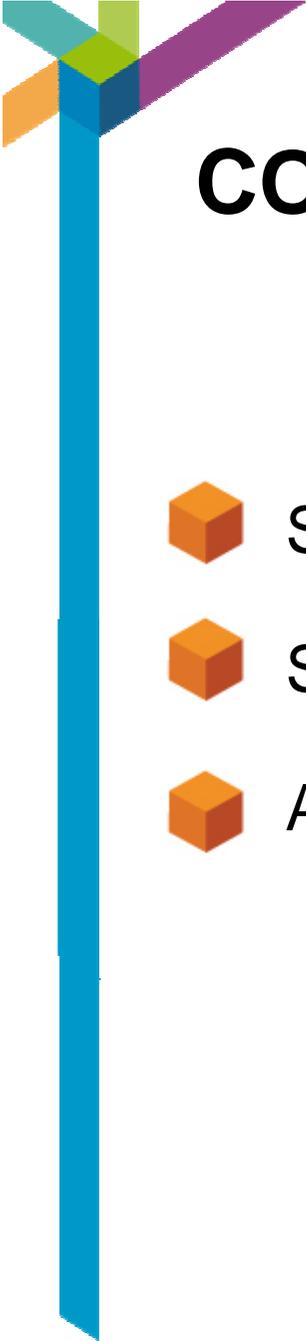
***Netezza Data Warehouse Appliance***

### Problemática

- Las editoriales absorben las pérdidas de las devoluciones cuando imprimen demasiados libros y se quedan sin stock cuando imprimen demasiado pocos.
- Para proporcionar a las editoriales la capacidad de obtener información en tiempo real sobre las ventas y establecer las tendencias de inventario en el tiempo.

### Beneficios

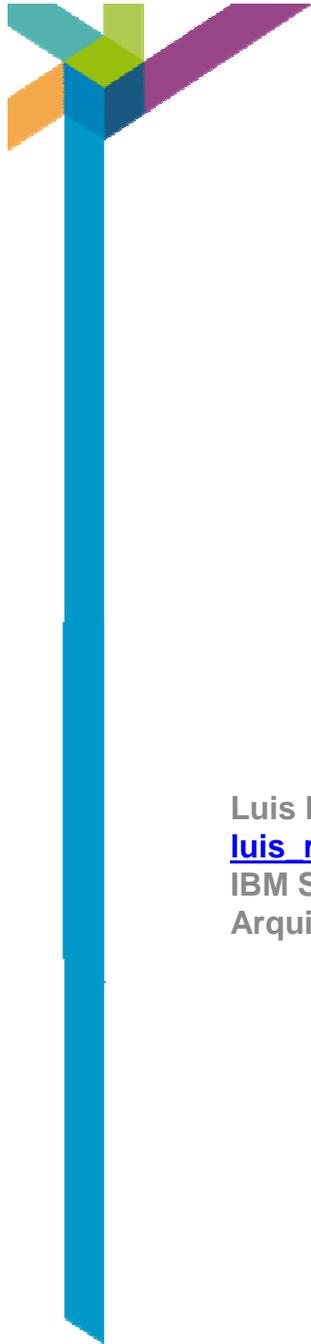
- Disminución en el tiempo de ejecución de las consultas de semanas a segundos y una reducción del 80% del tiempo comparado con su anterior sistema.
- Reducción de los niveles de inventario y los costes asociados.



# CONCLUSIONES

*“One size does not fit all”*

-  Solución a medida: **Infosphere Warehouse**
-  Solución Optimizada: **IBM Smart Analytics Software**
-  Appliance: **Netezza**



# GRACIAS

Luis Reina Juliá  
[luis\\_reina@es.ibm.com](mailto:luis_reina@es.ibm.com)  
IBM Software Group  
Arquitecto de Datos



[@ibmanalytics\\_es](https://twitter.com/ibmanalytics_es)

Big data, integración y gobierno: [www.ibm.com/software/es/info/rte/bdig/](http://www.ibm.com/software/es/info/rte/bdig/)

