

# Soluciones IBM Data Warehouse para el Siglo XXI

**INFORMATION INTEGRATION  
& GOVERNANCE FORUM 2011**  
*Delivering Trusted Information for Smarter Business Decisions*



# IBM el Inventor de las Bases de Datos



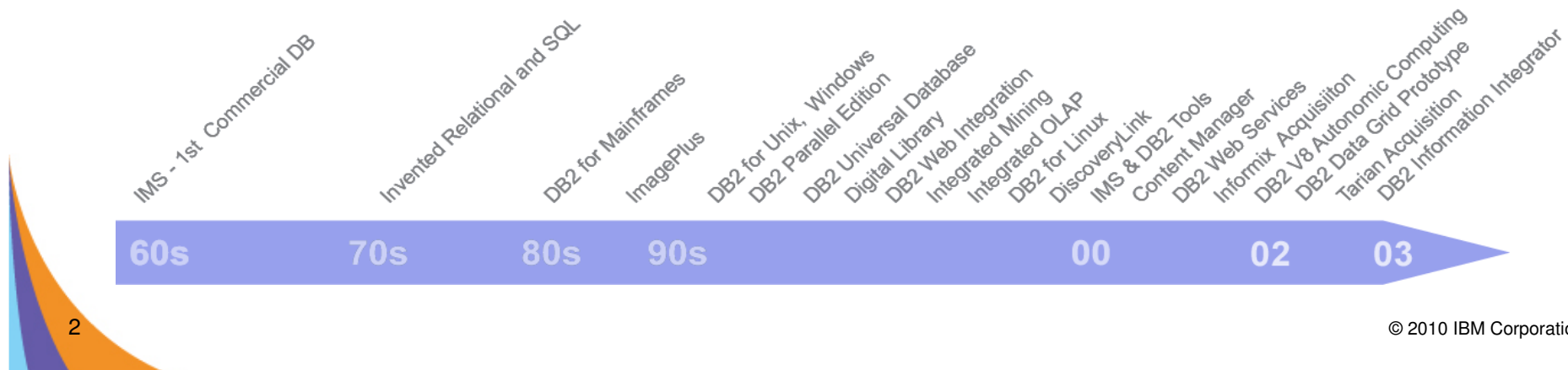
## Bases de Datos Jerárquicas

- 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.



## IBM el Líder Tecnológico

EMPRESAS | 5.896 patentes en EEUU

# IBM consiguió en 2010 más patentes que Microsoft, HP, Oracle, EMC y Google juntas

Portaltic | ELMUNDO.es | Madrid





Actualizado martes 11/01/2011 13:12 horas



IBM alcanzó un récord de **5.896 patentes en EEUU en 2010**, siendo el décimo octavo año consecutivo que ha encabezado la lista de las empresas más activas en este terreno del mundo, según los datos de [ifi CLAIMS](#). La compañía consigue así más patentes que varias de las compañías del sector juntas.

IBM es la primera empresa que ha conseguido hasta **5.000 patentes de los EEUU en un solo año**. Esta cifra se ha ido incrementando con los

## Agenda

-  Introducción
-  Infosphere Warehouse
-  IBM Smart Analytics System
-  Netezza

## Modelo de las 3Vs

### Volumen

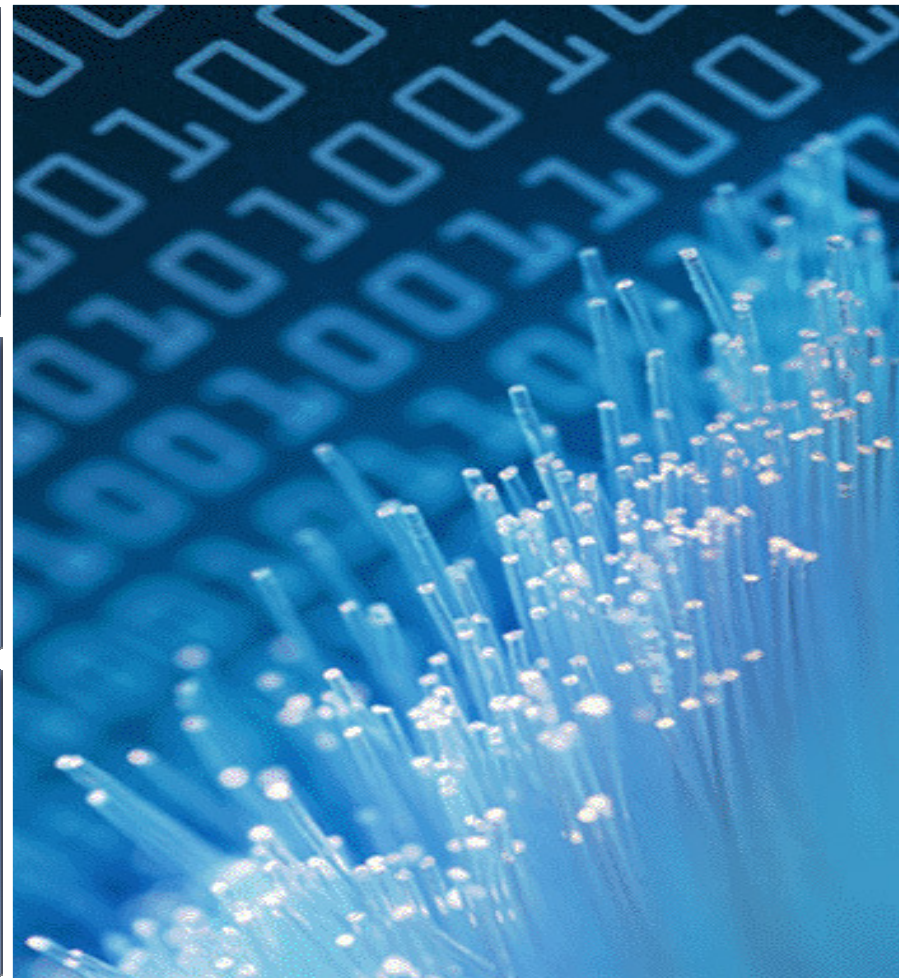
- Cada día, se genera 15 petabytes de nueva información.

### Variedad

- 80% de los nuevos datos son de tipo no estructurado, generados en su mayoría por correos electrónicos, imágenes y video.

### Velocidad

- Un compañía media de 1000 empleados gasta unos \$5,3 millones al año en encontrar información.



## ¿Cual son los retos?

### 1 de cada 3

Líderes de negocio frecuentemente toman decisiones basados en información de la que no se fían o que no tienen.

.

### 1 de cada 2

Líderes de negocio dicen que no tienen acceso a toda la información que necesitan para su trabajo.

### 83%

De los CIOs citó a “*Business intelligence and analytics*” como parte de su plan para mejorar la competitividad.

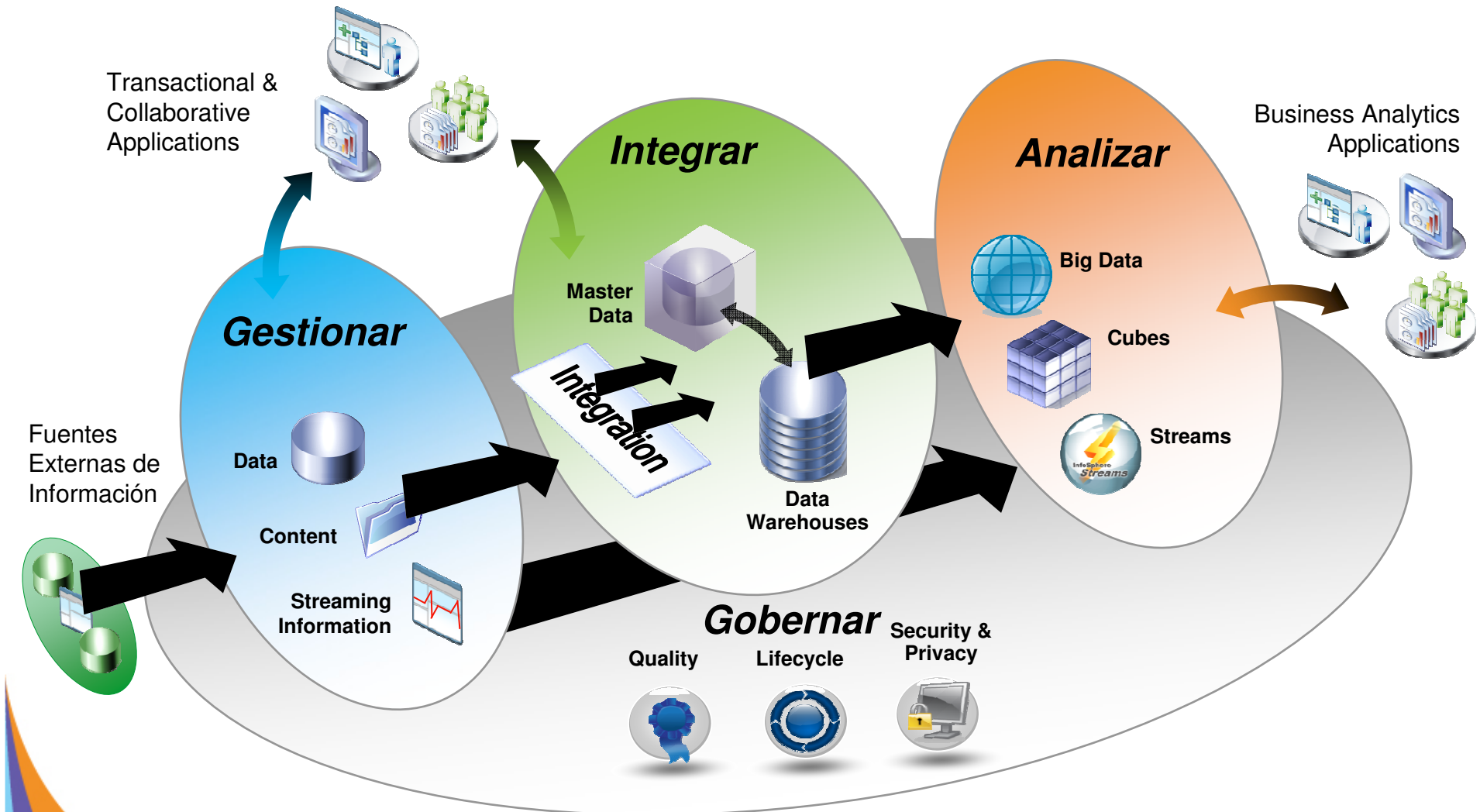
.

### 35%

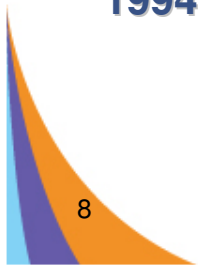
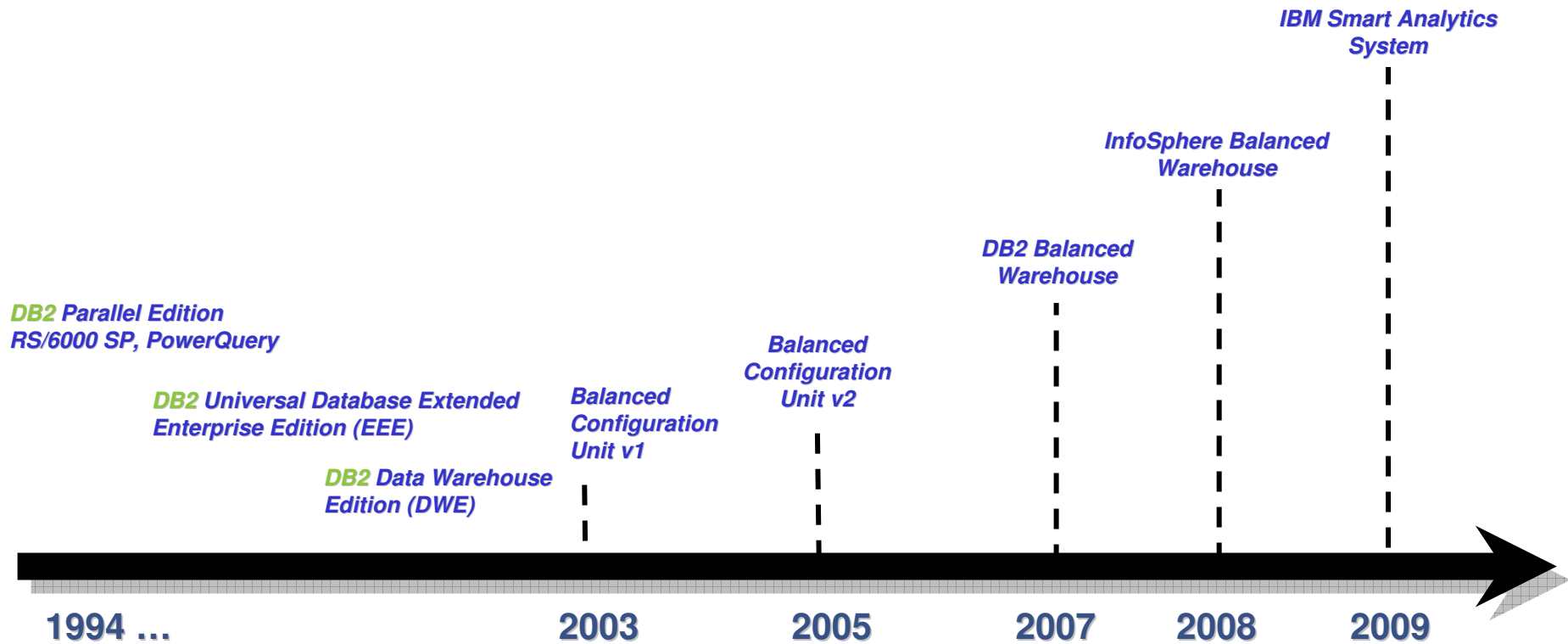
De los clientes estudiarán, dentro de los próximos 3 años, reemplazar su data warehouse actual con una solución de data warehouse pre-integrado, sólo el 14% lo tienen hoy.

Fuente (TDWI: Next Generation Data Warehouse Platforms Q4 2009)

# El Panorama de la Información



# IBM más de 15 años de avances en Data warehouse





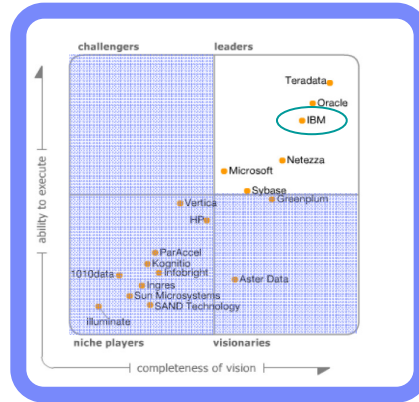
# IBM es líder reconocido en todas las areas de Gestión de la Información y Analítica de Negocio

**Business Intelligence**



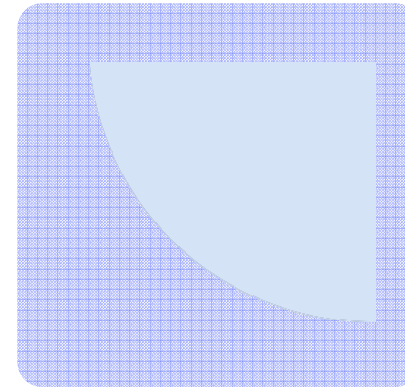
Gartner

**Data Warehouse**

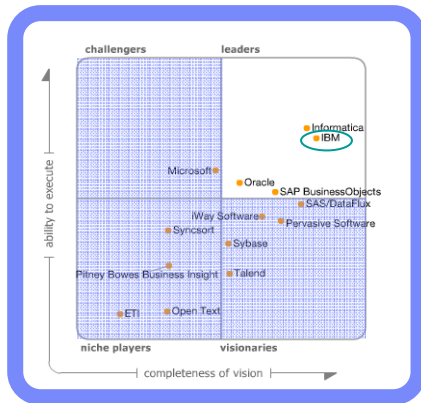


Gartner

**Data Mining**

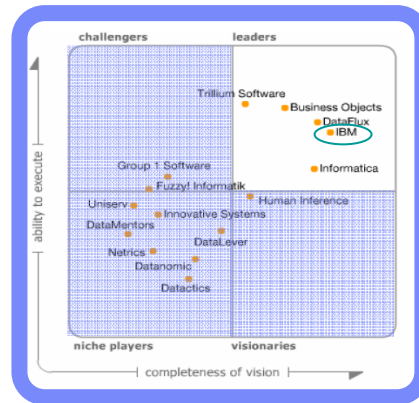


**Data Integration**



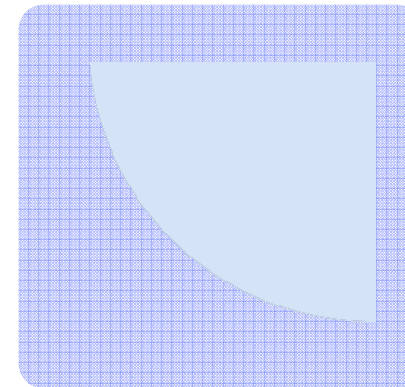
Gartner

**Data Quality**



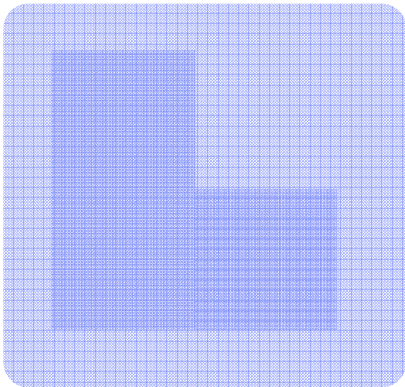
Gartner

**ETL**

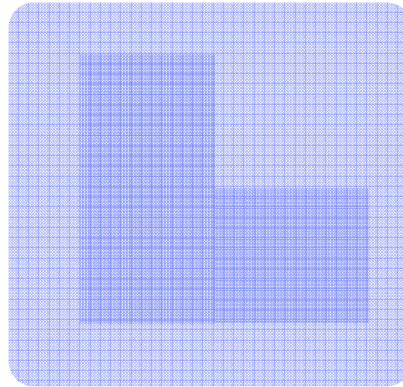


# IBM es líder reconocido en todas las areas de Gestión de la Información y Analítica de Negocio

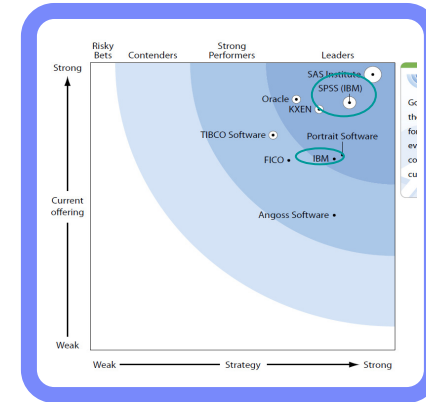
Business Intelligence



Data Warehouse

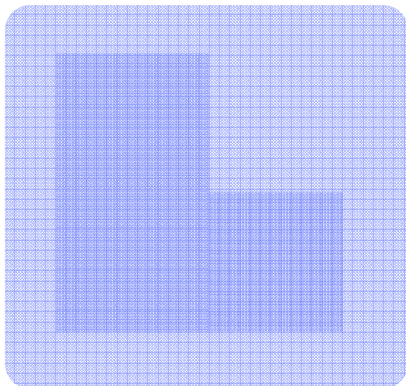


Data Mining

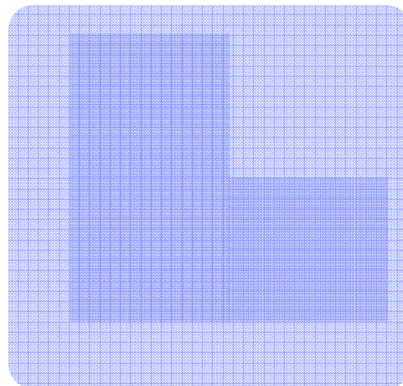


FORRESTER

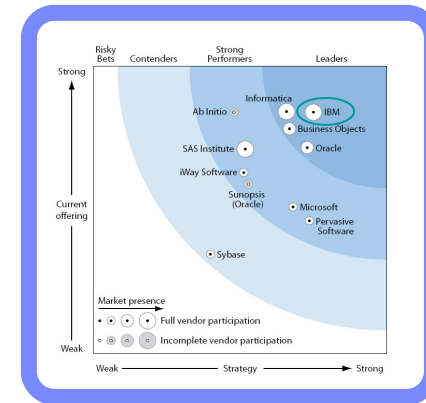
Data Integration



Data Quality



ETL



FORRESTER

# Soluciones de IBM para Data Warehouse

## IBM Netezza



True Appliance

## IBM Smart Analytics System



Flexible Integrated System

## IBM InfoSphere Warehouse



Custom Solution

Warehouse Accelerators

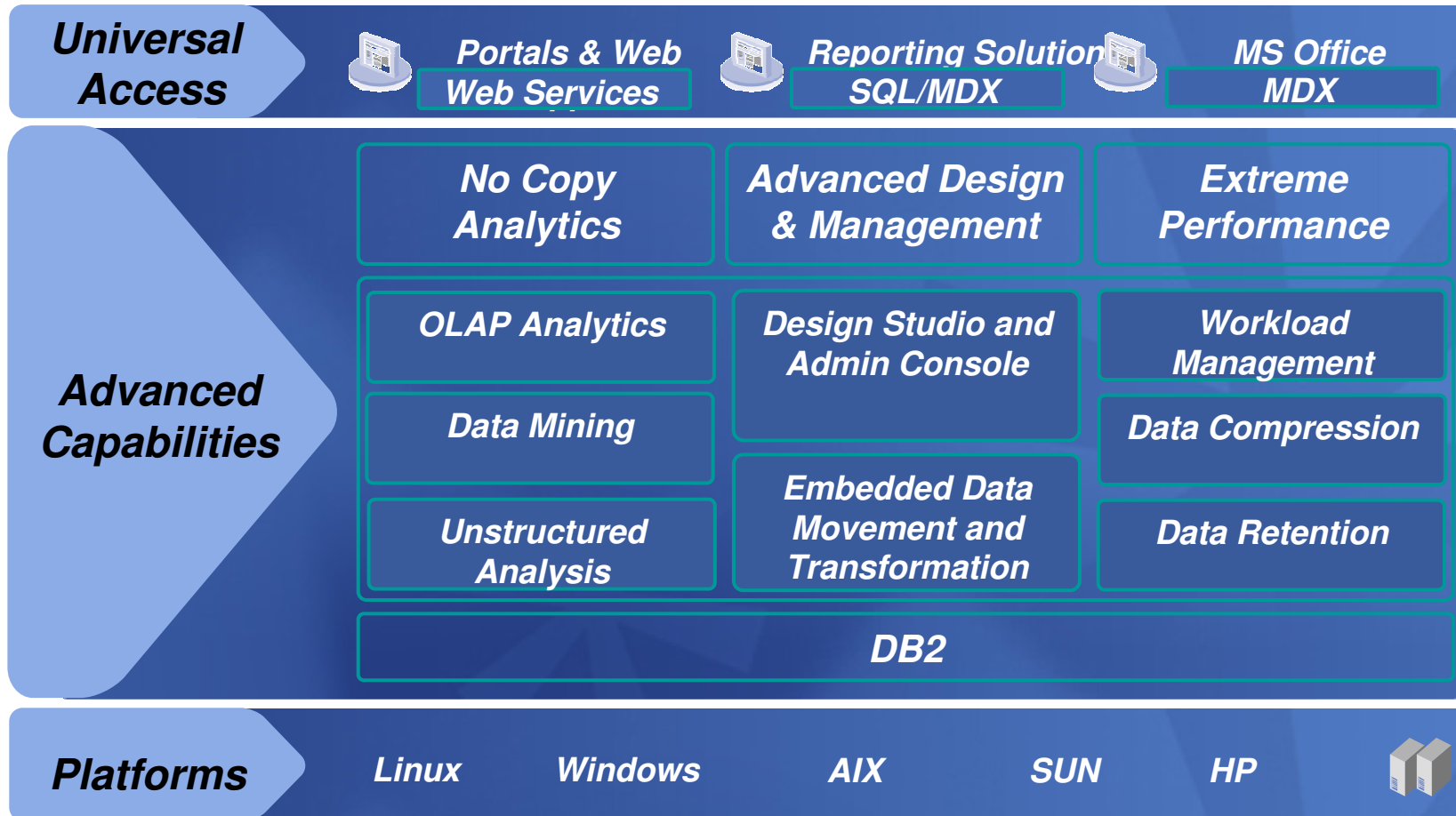
Information Management Portfolio  
*(Information Server, MDM, Streams, etc)*

Simplicity

*The right mix of simplicity and flexibility*

Flexibility

# Infosphere Warehouse: Plataforma para su Data Warehouse

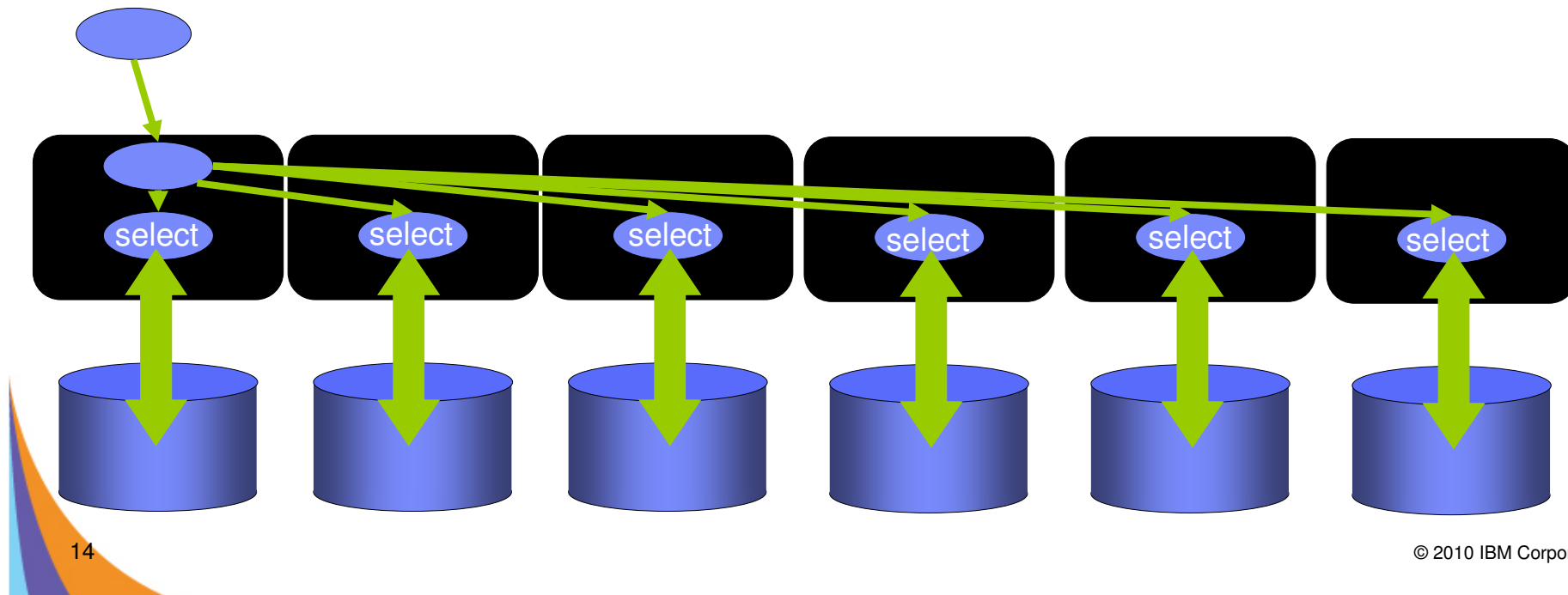


## *DB2*

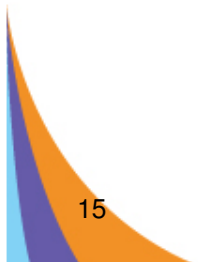
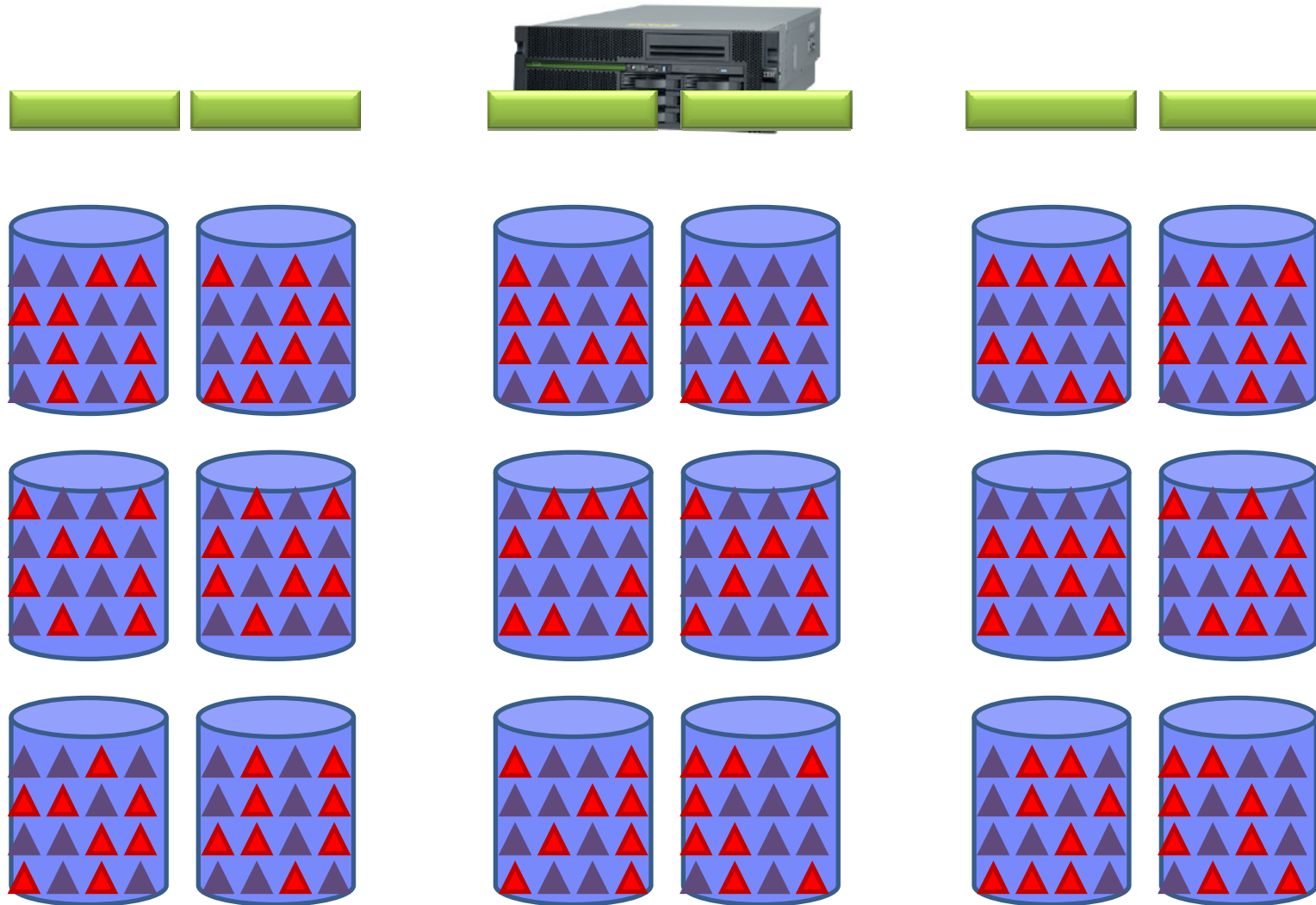
## Arquitectura “Divide y Venceras” permite escalabilidad lineal

### Arquitectura «*Shared Nothing*»

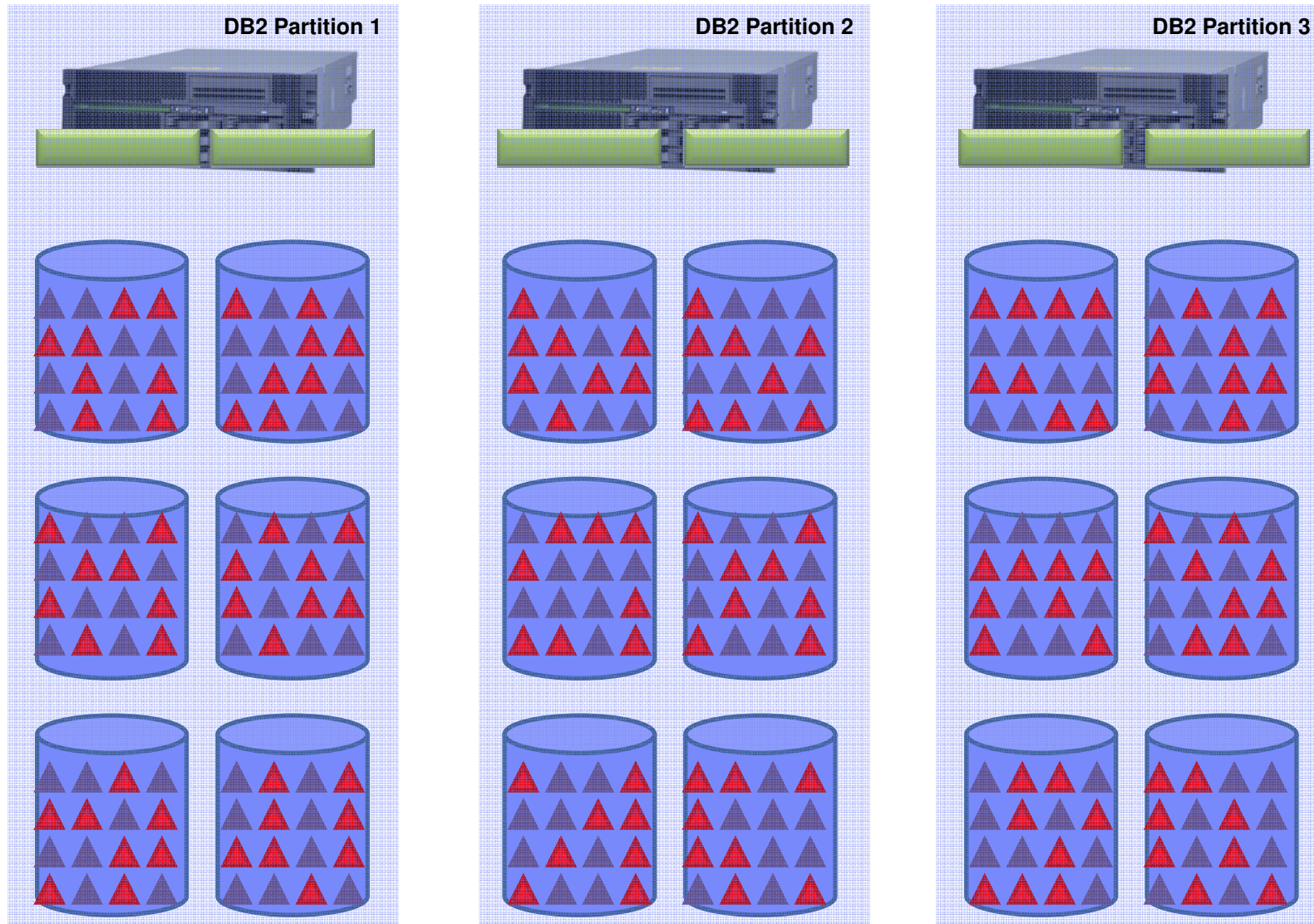
- Base de datos dividida en múltiples particiones
- Cada partición puede correr en un nodo (máquina) distinta.
- Cada servidor de partición tiene recursos dedicados.
- Existe un Procesamiento Paralelo en todas las particiones.
- **Las aplicaciones y los usuarios ven una imagen única.**



# Mejora del Rendimiento mediante la eliminacion de la E/S

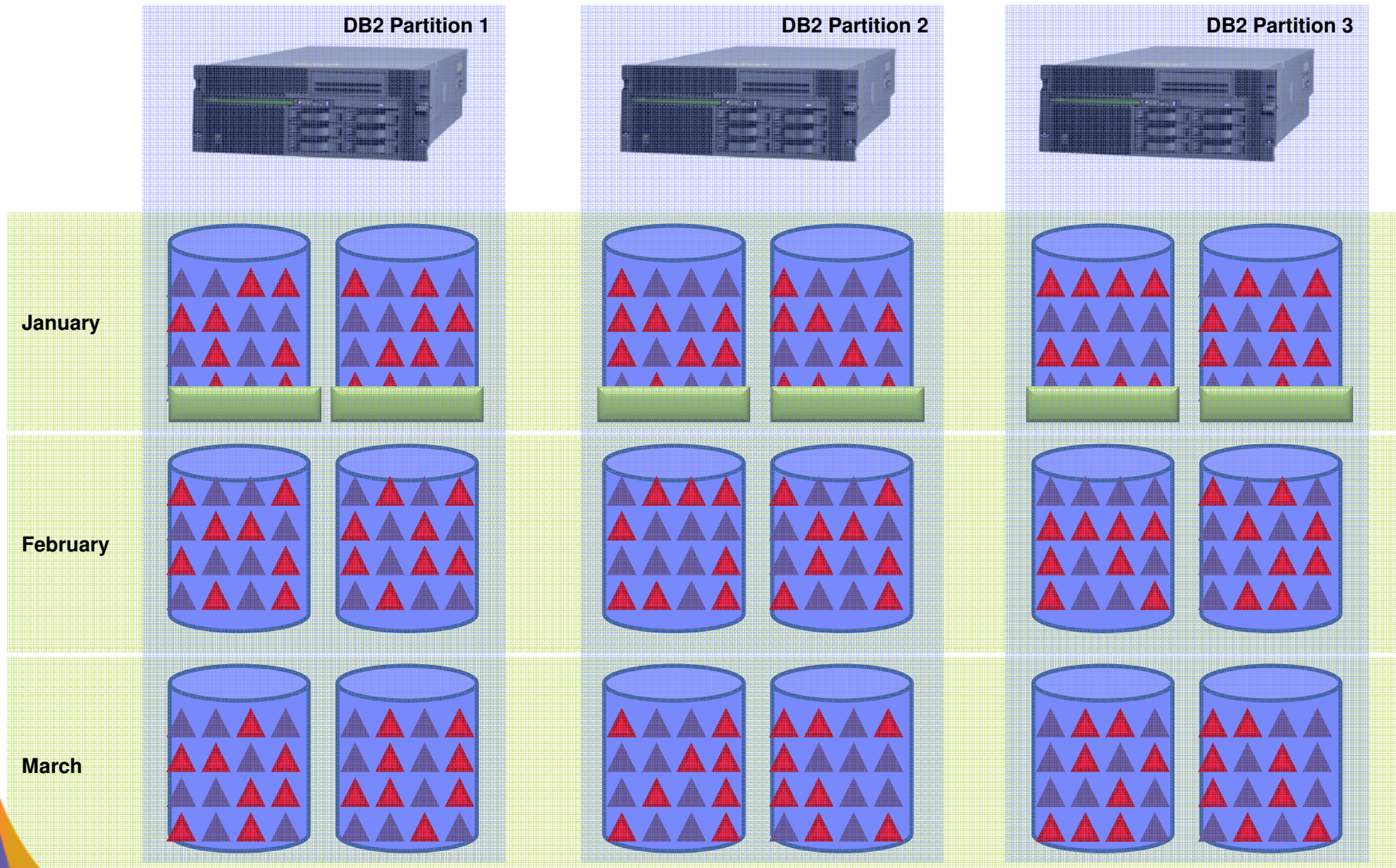


# Particionamiento de Base de Datos





# Particionamiento de Base de Datos + Particionamiento por Rango



# *Cubing Services*

## Servidor de Cubos Multidimensionales

# Cubing Services: Acelerador de consultas

*Portals, Web Applications, Dashboards, Interactive Reports, Ad Hoc Analysis, Common Desktop Tools*

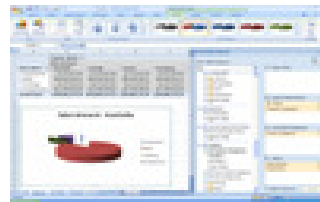
**Cognos 8 BI**



**IBM Alphablox**



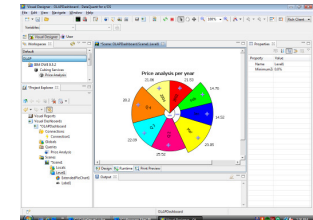
**Microsoft Excel**



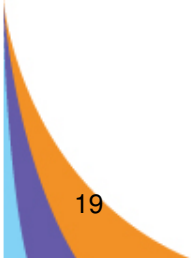
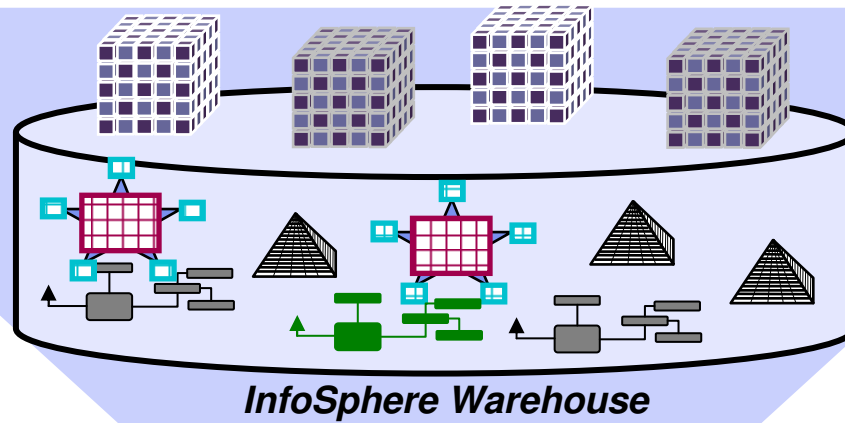
**Cubeware**



**IBM DataQuant & DB2 QMF**

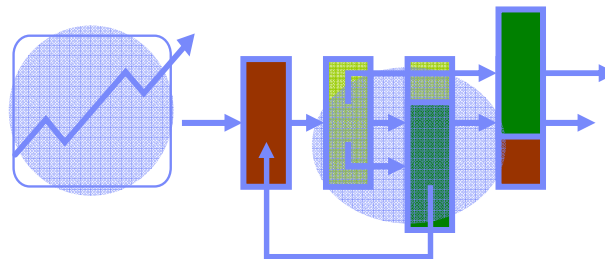


**Universal Cube Access**  
(MDX, ODBO, XMLA)

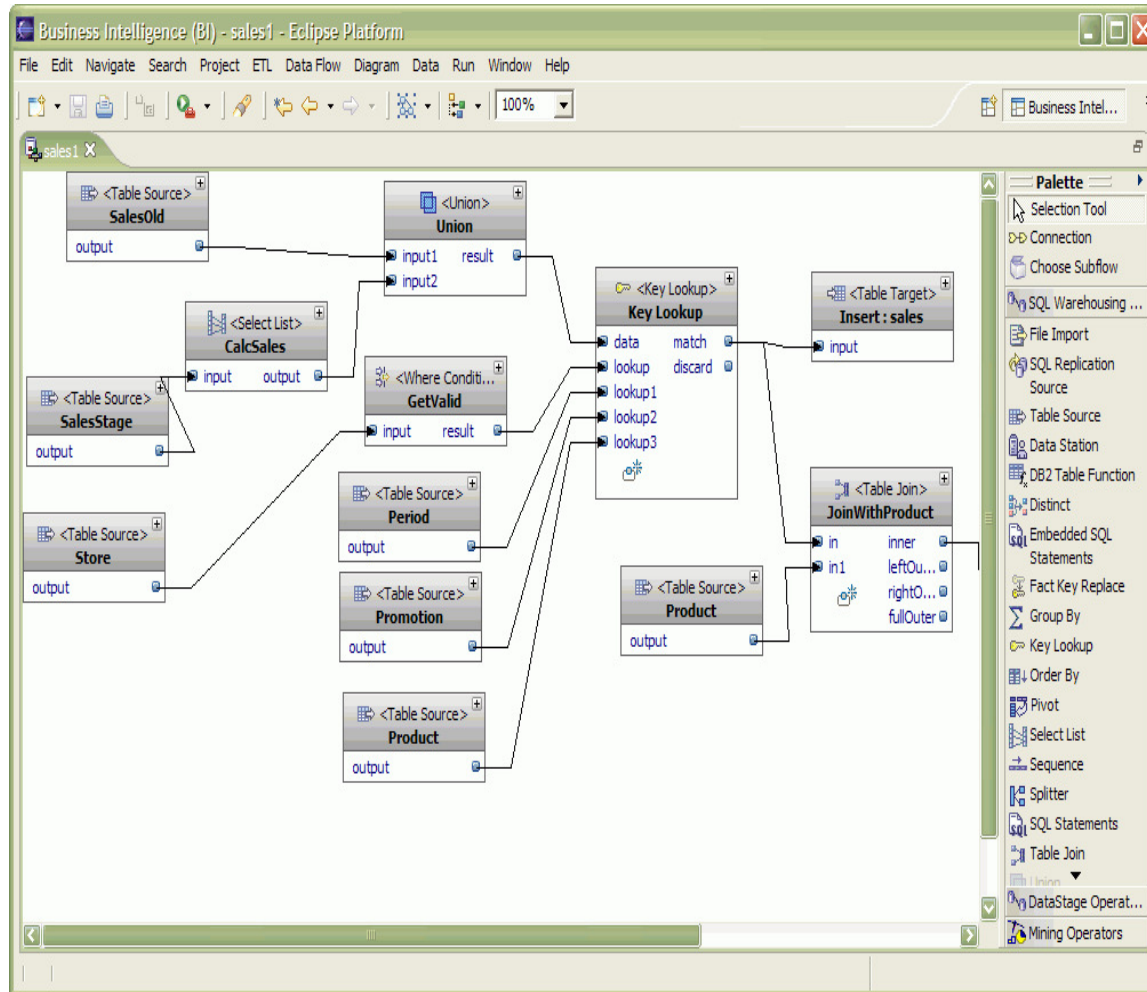


# SQL Warehouse (SQW)

## Movimiento y Transformación de datos



## SQL Warehouse: Movimiento y Transformación de Datos



- *Fácil de Usar*
- *Interfaz Gráfico*
- *Workflow y Planificación*
- *Automatiza los procesos de Text Analytics y Data Mining.*
- *Fuentes DB2 y no DB2 (e.g Oracle)*
- *Se integra con Information Server (Datastage)*
- *Gestión de versiones*
- *Monitorización de jobs*

# *Data Mining* Minería de Datos

# Visión general del data mining en Infosphere warehouse

## ■ **Modeling:** Crea los Modelos.

- Associations
- Sequences
- Clustering
- Decision tree classification
- Regression

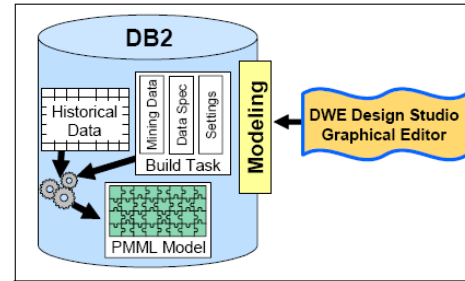


Figure 7-1 DWE Modeling architecture

## ■ **Visualization:** Visualiza los Modelos (para explorar y validar).

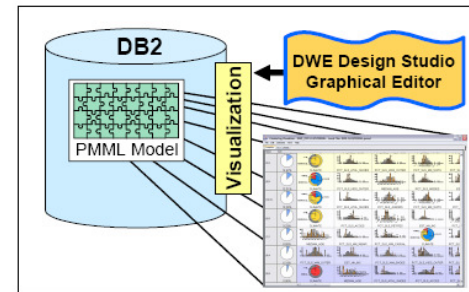


Figure 7-2 DWE visualization architecture

## ■ **Scoring:** Aplica los Modelos

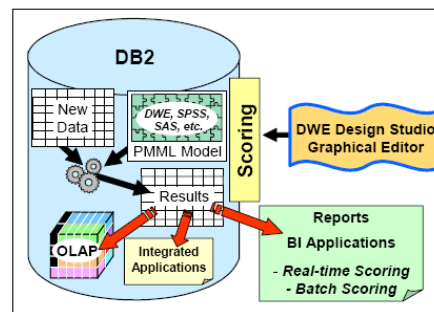


Figure 7-3 DWE Scoring architecture

# *Text Analytics*

## **“Minería” en campos de texto libre**



## Text Analytics: Análisis de Campos de Texto Libre

ID	LONGDISTANCE	INTERNATIONAL	LOCAL	DROPPED	PAYMETHOD	LOCALBILLTYPE	LONGDISTBIL...	AGE	GENDER
2211	26.23	0.0	32.45	0.0	CC	Budget	Standard	39.2	M
3224	27.3	0.0	38.43	1.0	CC	FreeLocal	Standard	29.45	M
3322	34.3	0.0	38.5	1.0	CC	FreeLocal	Standard	37.4	M
3325	31.0	0.0	38.4	0.0	CC	Budget	Standard	44.2	M
3445	17.4	12.4	42.21	0.0	CC	FreeLocal	Standard	36.22	F
3654	42.3	0.0	32.9	0.0	CC	Budget	BusinessPlan ...	42.3	F
3744	28.2	0.0	12.22	0.0	CC	Budget	Standard	64.2	M
3767	28.3	23.3	2.4	1.0	CC	Budget	BusinessPlan ...	43.2	F
4432	29.6	14.6	27.0	0.0	CC	Budget	Standard	52.3	F

Volumes of structured, well-organized demographic and transactional data



ID	NOTES
2211	She claims the bill always seems to have calls she didn't make
3224	He lost his phone and called to cancel service. Told him we would suspend until we hear back from him. He will ...
3322	Really is unhappy with website. Claims self-service stuff doesn'T work
3325	Can't believe he was stupid enough to sign a year contract. he thinks we provide the worst service he has eve...
3445	Needs help learning how to use phone. Suggested he go back to store and have the rep teach him.
3645	Wants to change rate plans to get family all on same bill. Transferred him back to cust. care
3744	Claims he never received his bill. Was really unhappy that I aksed him fro his cc. Wanted me to turn on his pho...
3767	he loves the phone when it works. Problem is that the phone hardly ever works. He thinks the battery is lame

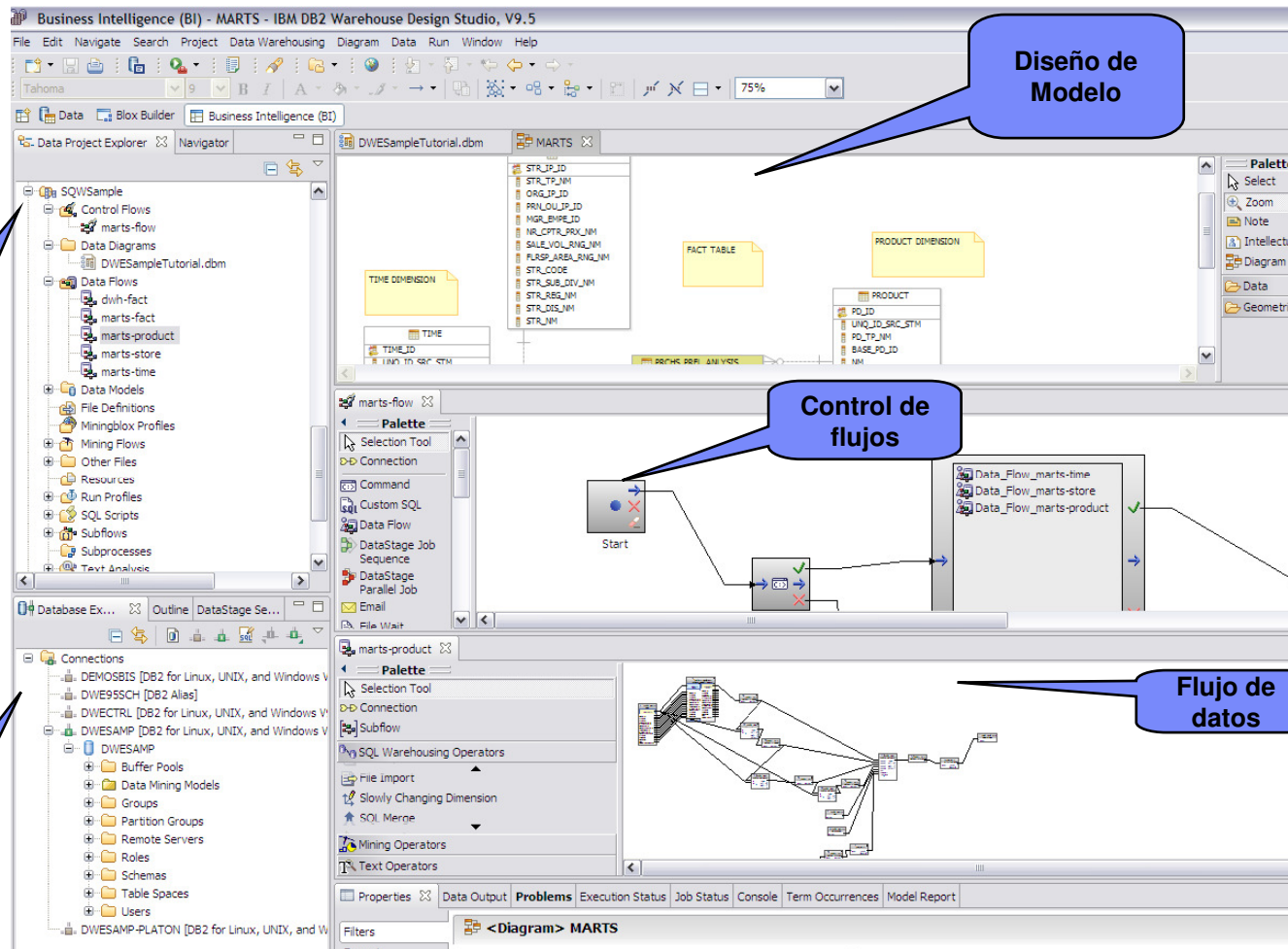
Volumes of unorganized, unstructured data from call-center notes

# *Design Studio*

(Modelado de datos Físico y  
Herramienta de Desarrollo)

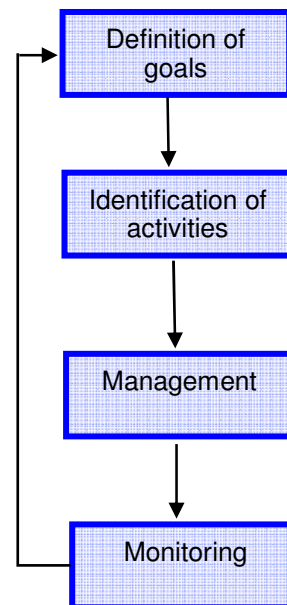
# Design Studio: Herramienta de Desarrollo

*Entorno único de desarrollo y administración, basado en eclipse*



# *Workload Manager*

(Control de las Cargas en DB2)



## Workload Manager: Control sobre las diferentes cargas de trabajo

Permite controlar sistemas complejos.

- **Introduce nuevas cargas de trabajo sin impacto sobre las ya existentes.**
- **Priorización de los procesos en función de las necesidades de negocio.**

Ayuda a conseguir los acuerdos sobre niveles de servicio

- **DB2 automáticamente modifica y ajusta los recursos y prioridades dependiendo de la carga de trabajo**
- **Permite generar controles basados en usuarios y aplicaciones**



DB2 Workload Manager

- **Forma parte del motor de DB2**

# *Performance Expert* Monitorización de DB2 y Alertas

# Performance Expert: Monitorización y Alertas

**Specify the screen refresh rate. Most KPIs are displayed as delta values.**

**Problems/warnings are highlighted in red/yellow**

**Portlets show for each KPI a database ranking**

**For multipartition instances additionally a combo box let you choose the partition or GLOBAL**

**Switch on/off monitor switches for this instance**

**Refresh rate:** 01:00

**Instances:** All Instances, CM4WAYB, **DB2**, DB2PMLOC, PEDEMO, SAMPLE, DB28\_FROM\_DB29, DB29\_FROM\_DB28, PEDEMO, DB2\_01, MYAIX, PMMON6M1, TURMALIN, My Shortcuts

**KPIs:** Data Views, Exceptions

**Sorts**

Sorts per minute (sum)		46
PEDEMO		32
DB2PMLOC		14
SAMPLE		0
Sort overflows per min. (sum)		0
Post threshold sorts per minute		0
Average sort time (max)		2ms
Average sorts per UOW (max)		0.13

**Applications / Agents**

Currently executing (sum)		2
Waiting for locks (sum)		1
Connected / max appls (sum)		20 / 26
PEDEMO		14 / 16
DB2PMLOC		5 / 7
SAMPLE		1 / 3
Current agents / max agents		36 / 36
Stolen agents per minute		0

**Buffer Pools**

Buffer pool hit ratio (min)		49.8%
Async. read percentage (min)		0%
Avg. rows read per sel. row (max)		18.1
DB2PMLOC		18.1
PEDEMO		1.34
SAMPLE		N/C
Avg. rows read per UOW (max)		136.19

**Databases**

Currently active		3
UOW per minute (sum)		638
Database files closed per minute (sum)		0
Package cache hit ratio (min)		45.6%
PEDEMO		45.6%
DB2PMLOC		56.8%
SAMPLE		N/C
Catalog cache hit ratio (min)		63.5%

**Locks**

Escalations per minute (sum)		1
Timeouts per minute (sum)		0
Deadlocks per minute (sum)		1
Waits per minute (sum)		0
Avg. wait time per lock (max)		N/C
Avg. wait time per appl. & min (max)		2142ms
Avg. wait time per UOW (max)		250ms
Lock wait time per min. (sum)		60s

**Monitor Switches**

Buffer Pool	<input checked="" type="checkbox"/> On	<input type="checkbox"/> Off
Lock	<input checked="" type="checkbox"/> On	<input type="checkbox"/> Off
Sort	<input checked="" type="checkbox"/> On	<input type="checkbox"/> Off
Statement	<input checked="" type="checkbox"/> On	<input type="checkbox"/> Off
Table	<input checked="" type="checkbox"/> On	<input type="checkbox"/> Off
Unit of Work	<input checked="" type="checkbox"/> On	<input type="checkbox"/> Off
Timestamp	<input checked="" type="checkbox"/> On	<input type="checkbox"/> Off

**Server Status:** Logon, DB2 System, Group, User ID, Exception, Trace Status, Session, Operating S..., System Name, DB2, Server, Host, Port, Description

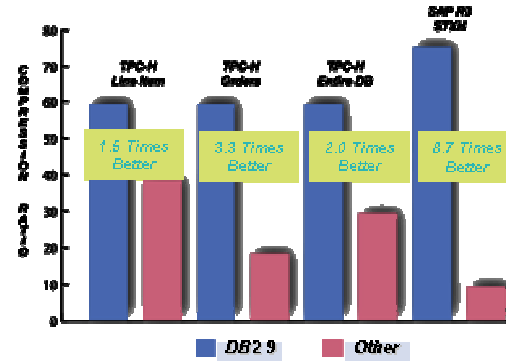
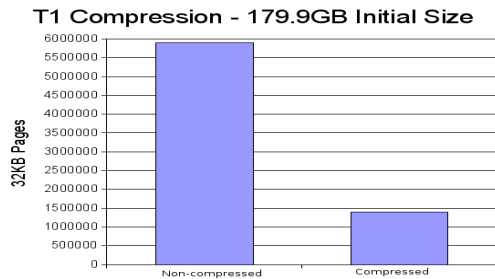
# *Storage Optimization*

## Compresión de datos en DB2

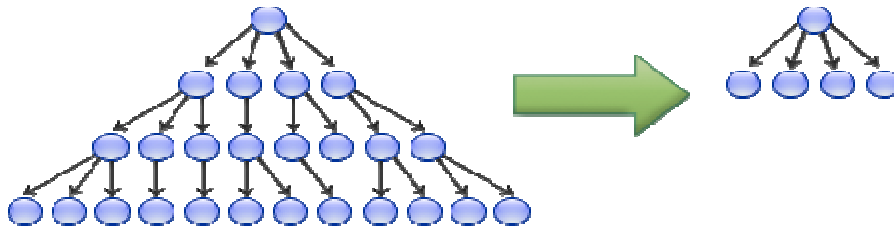


# Compresión de Datos

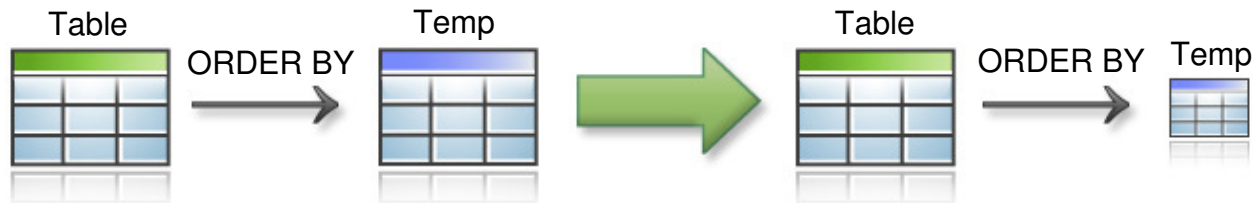
– Compresión de Datos



– Compresión de Índices



– Compresión de Temporales



# Soluciones de IBM para Data Warehouse

**IBM  
Netezza**



**True Appliance**

**IBM  
Smart Analytics System**



**Flexible Integrated System**

**IBM  
InfoSphere Warehouse**



**Custom Solution**

**Warehouse Accelerators**

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

**Simplicity**

*The right mix of simplicity and flexibility*

**Flexibility**

## IBM Smart Analytics System

*Lo que la organizaciones inteligentes quieren y necesitan para tener resultados más rápidos*

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

Incluye:

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



*Transformando la  
información en  
compresión del  
negocio*

***3x más Rápido***

***50% menos espacio***

***Record Mundial en  
Rendimiento***

# IBM Smart Analytics System: Sistema montado versus un Sistema por Piezas



**Vs.**



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



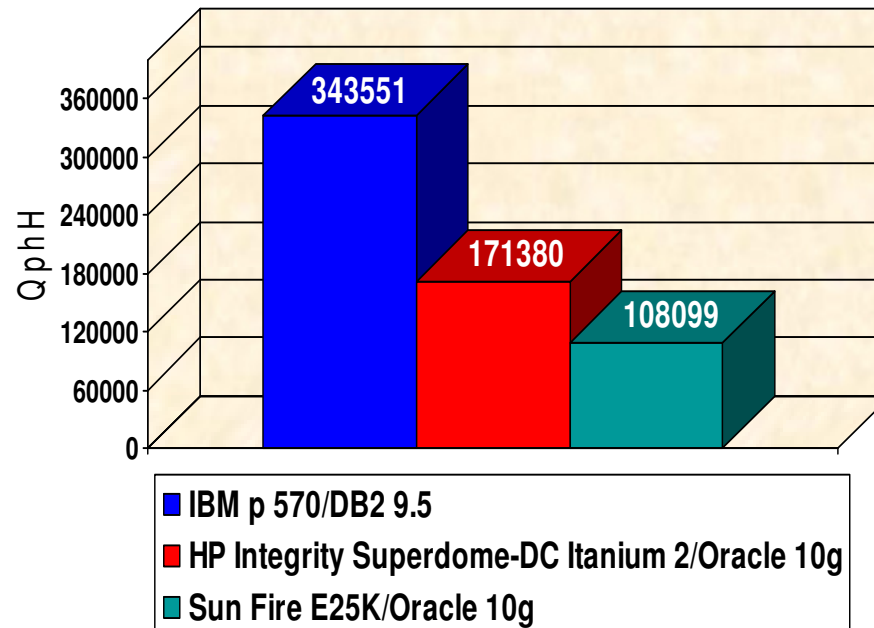
# IBM Smart Analytics System: Solución de Alto Rendimiento

**TPC** Transaction Processing  
Performance Council

<http://www.tpc.org>

*The TPC defines transaction processing and database benchmarks and delivers trusted results to the industry.*

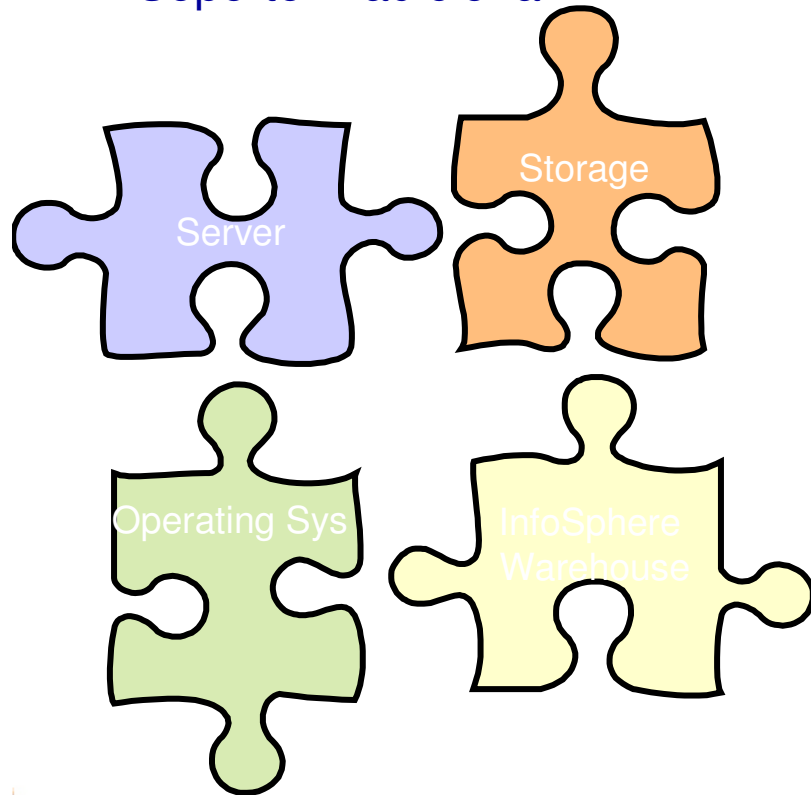
## World-record 10TB TPC-H



<http://www.ibm.com/software/data/db2/performance.html>

# 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

# IBM Smart Analytics System: Sistema Modular y Escalable

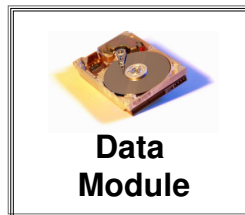


- ✓ Se construye mediante módulos que permiten un crecimiento flexible y controlado
- ✓ Cada Modulo consta de Procesamiento + Almacenamiento + Memoria.



**Modulos de Usuario**

1 o más



**Data Module**

1 o más



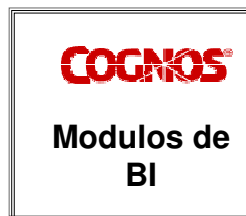
**Modulos de failover**

1 o más



**Servidor de Gestión**

1



**Modulos de BI**

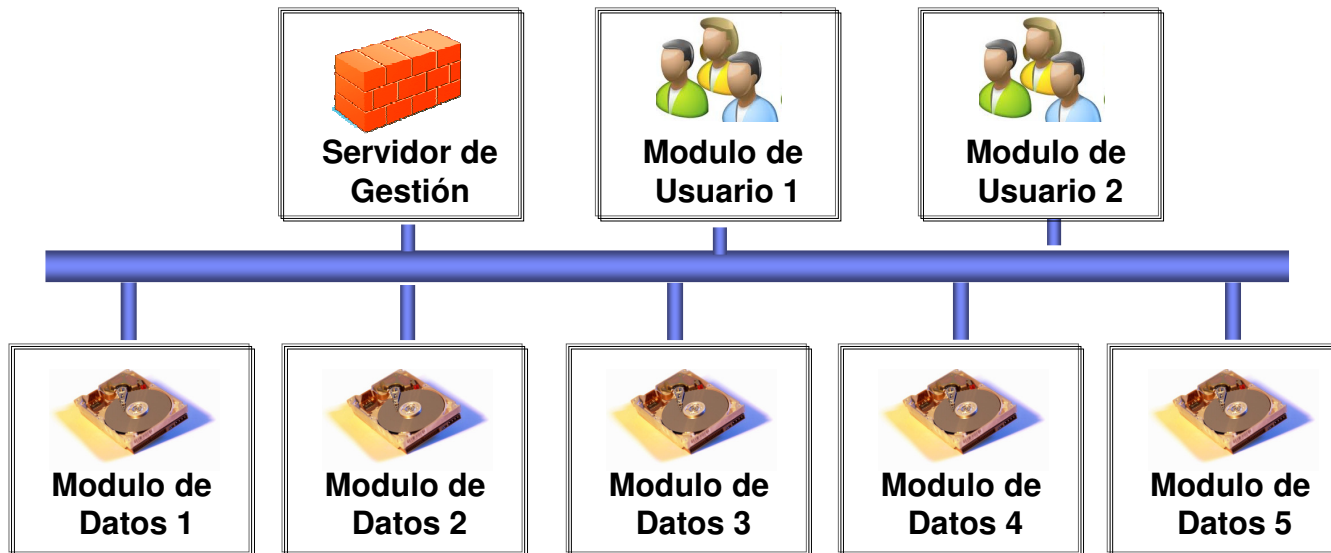
(opcional)



**Modulos de Aplicaciones**

(opcional)

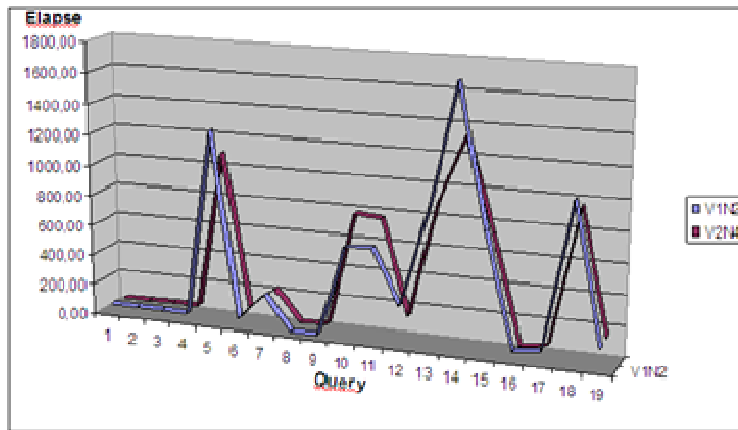
# IBM Smart Analytics System: Sistema Flexible con Crecimiento Controlado



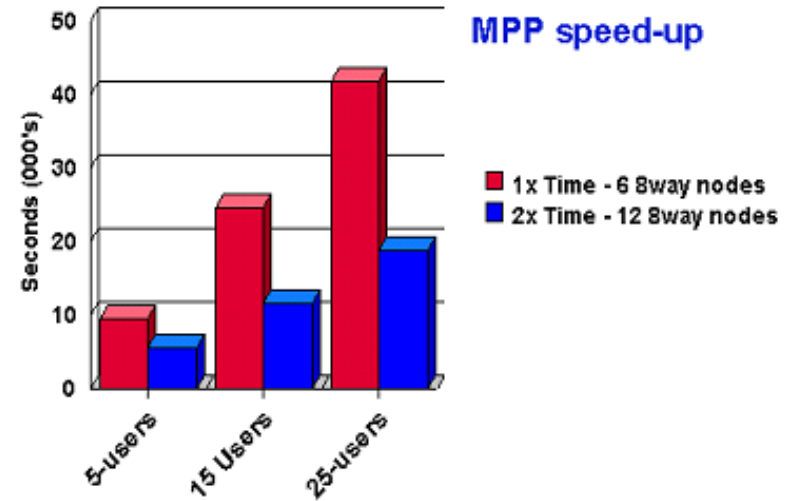


# IBM Smart Analytics System: Escalabilidad Lineal

Industry Application	Scaling Description	Scale Result
Insurance	8 to 16 partitions, 4 workloads	95% to 109%
Financial	32 to 40 partitions, 1 workload	107%
Distribution	16 to 24 partitions, 3 workloads	85% to 112%
Telco	48 to 96 partitions, 5 workloads	105% to 122%



2TB x 2NODES  
4TB x 4NODES



MPP speed-up

1x Time - 6 8way nodes  
2x Time - 12 8way nodes

# IBM Smart Analytics System: Administración Sencilla

## Monitorización

- Status Global
- Status de Componentes

## Mantenimiento

- Completo (firmware, software, SO)
- Crecimiento

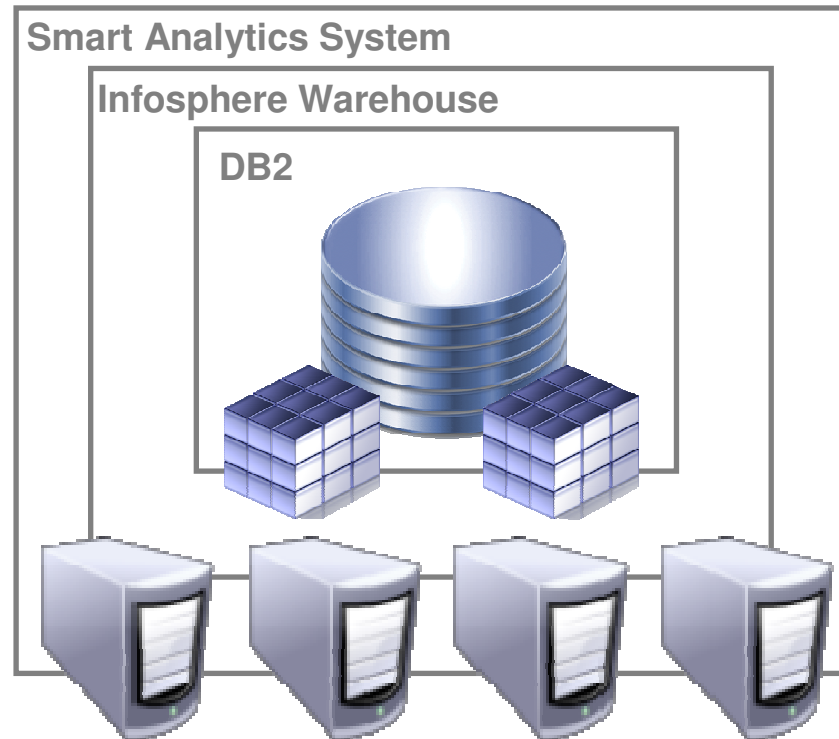
## Consola Web/ Vía Comandos

The screenshot displays the 'Data Studio Web Console' interface. On the left, there is a navigation menu with options: Stop Appliance, Update Appliance, Network Config, User Management, Maintenance, Infrastructure, Cognos BI, and Balanced Warehouse. The main area is titled 'Overall Appliance Health' and shows 'Flex Data Visualization Platform' for appliance 'server1.dublinlab.ibm.com'. Below this, 'Detailed Solution Views' are shown for 'Storage Manager' and 'Component XYZ2', each with CPU, MEM, and DISK usage bars and 'Expand View'/'Stop' buttons. A terminal window is overlaid on the console, displaying a list of files and their sizes, including 'S14.inq.ppt', 'S14.inq I.0.ppt', 'S14.inq Charts.ppt', 'Balanced Warehouse Overview for B...', 'Sandy\_Warehouse\_Charts.ppt', 'S14.exe', 'warehouse@CPU1\_1.ppt', 'ManagementInterfaces.ppt', 'people.txt', 'Pinned\_Delivery.ppt', 'warehouse@CPU1.ppt', 'S-Services-Based Management Node.ppt', 'McKessonManagementInterfaces.ppt', and 'Management Interfaces Architecture Overp...'. The terminal also shows directory paths like 'C:\Program Files\IBM\...' and 'C:\Program Files\IBM\...'. The terminal window title is 'C:\Program Files\IBM\...'. The terminal output shows a list of files and their sizes, including 'S14.inq.ppt', 'S14.inq I.0.ppt', 'S14.inq Charts.ppt', 'Balanced Warehouse Overview for B...', 'Sandy\_Warehouse\_Charts.ppt', 'S14.exe', 'warehouse@CPU1\_1.ppt', 'ManagementInterfaces.ppt', 'people.txt', 'Pinned\_Delivery.ppt', 'warehouse@CPU1.ppt', 'S-Services-Based Management Node.ppt', 'McKessonManagementInterfaces.ppt', and 'Management Interfaces Architecture Overp...'. The terminal also shows directory paths like 'C:\Program Files\IBM\...' and 'C:\Program Files\IBM\...'. The terminal window title is 'C:\Program Files\IBM\...'.

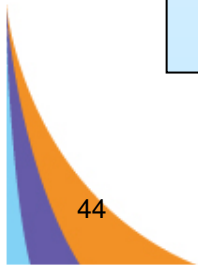
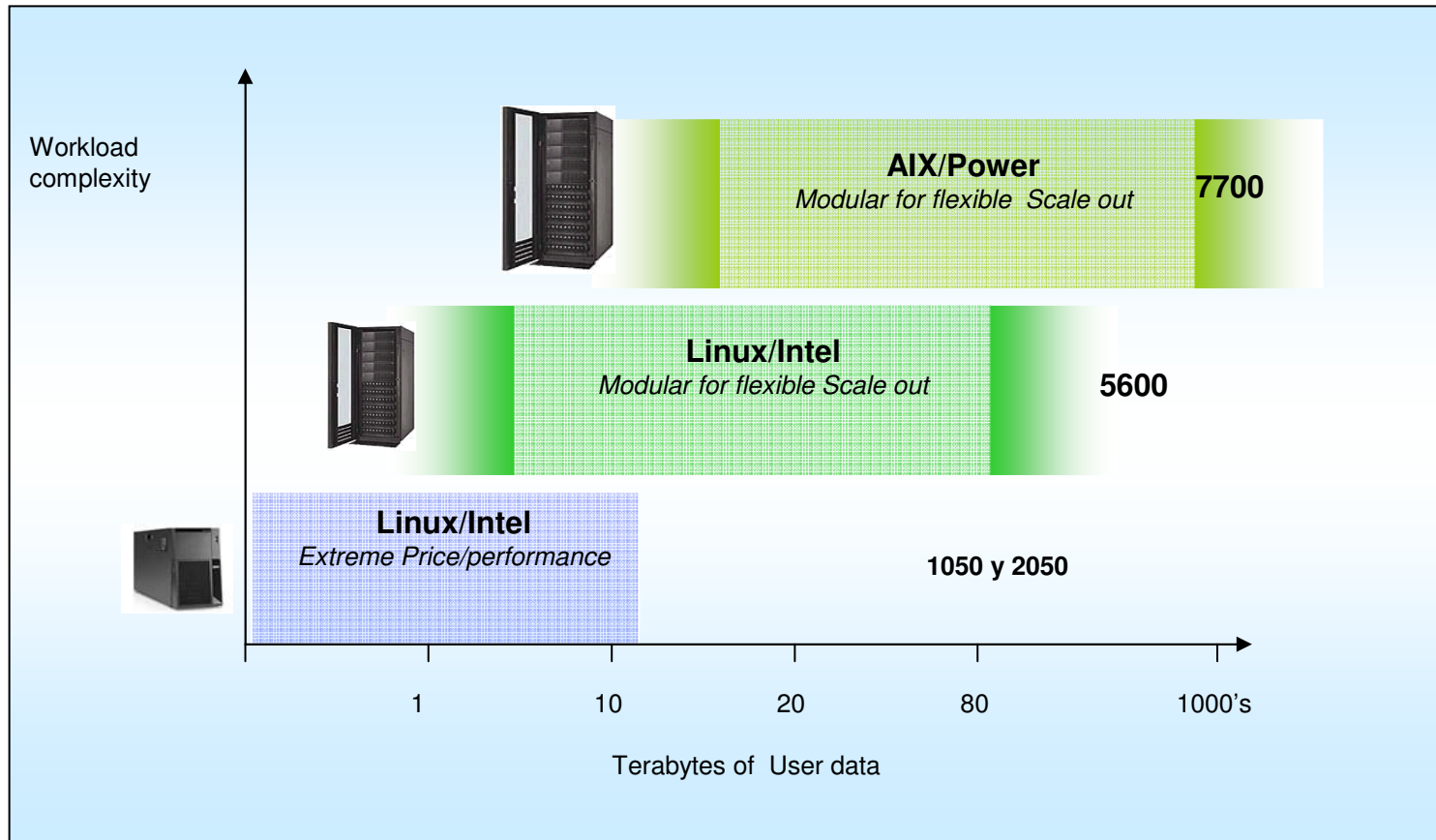
## Gestión

- Configuración
- Control

# IBM Smart Analytics System



# IBM Smart Analytics System: Distintos Modelos



# IBM Smart Analytics System Sistemas Compactos

## IBM Smart Analytics Systems 1050 y 2050

### Hardware/OS

- SUSE 11 or Windows Server 2008
- IBM System x3500 M3 hardware for System 1050
- IBM System x3850 X5 hardware for System 2050
- IBM System Storage DS3500



### Software

- InfoSphere Warehouse Departmental o Departmental Base Edition
- Cognos 8 BI Reporting and Query (opcional)
- Compacto, de bajo coste, un único sistema que proporciona un perfecto balanceo entre potencia y simplicidad.

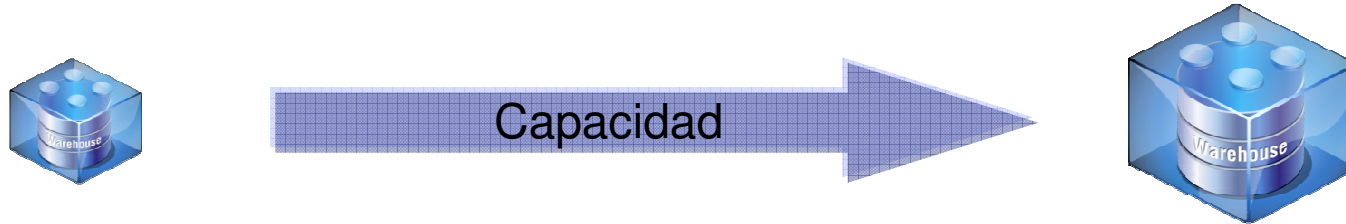
### Seis Formatos

- *1050 Small, 1050 Medium, 1050 Large*
- *2050 Small, 2050 Medium, 2050 Large*

1050 2050



# IBM Smart Analytics Systems 1050 y 2050: Capacidades



Medidas de Tamaños	1050			2050		
	S	M	L	S	M	L
User space (TB)	0.33	1.65	3.3	3.3	6.6	13.2
Total space (TB)	0.6	3	6	6	12	24
<b>Minimum # Users</b>	-	-	-			
Departmental	30	30	30	N/A	N/A	N/A
Departmental Base	20	20	20			
<b># BI Users</b>	>= 0	>= 0	>= 0	>= 0	>= 0	>= 0

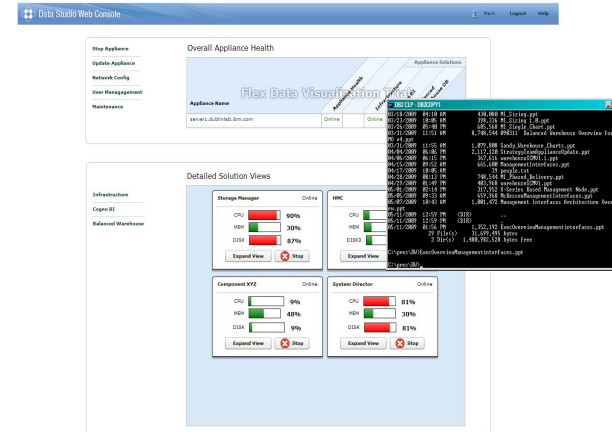
# IBM Smart Analytics System Sistemas Modulares



# IBM Smart Analytics System: Linux

## Software

- Smart Analytics System 5600 Control Console
  - Allows for coordinated updates and maintenance of software, operating system, firmware and drivers
- InfoSphere Warehouse Enterprise 9.7.2
  - Number of performance, administration, and usability enhancements
- Cognos 8.4.1 (optional)
- Linux SLES 10 SP3



## Hardware

- System x3650 M3 server
  - Energy-efficient design supporting more cores, memory & data capacity
  - Scalable 2U package that's easy to service & manage
- DS3500 Storage
  - Provides affordable performance
  - flexibility without sacrificing scalability, efficiencies or ease of use



5600

## IBM Smart Analytics System: AIX

### Características

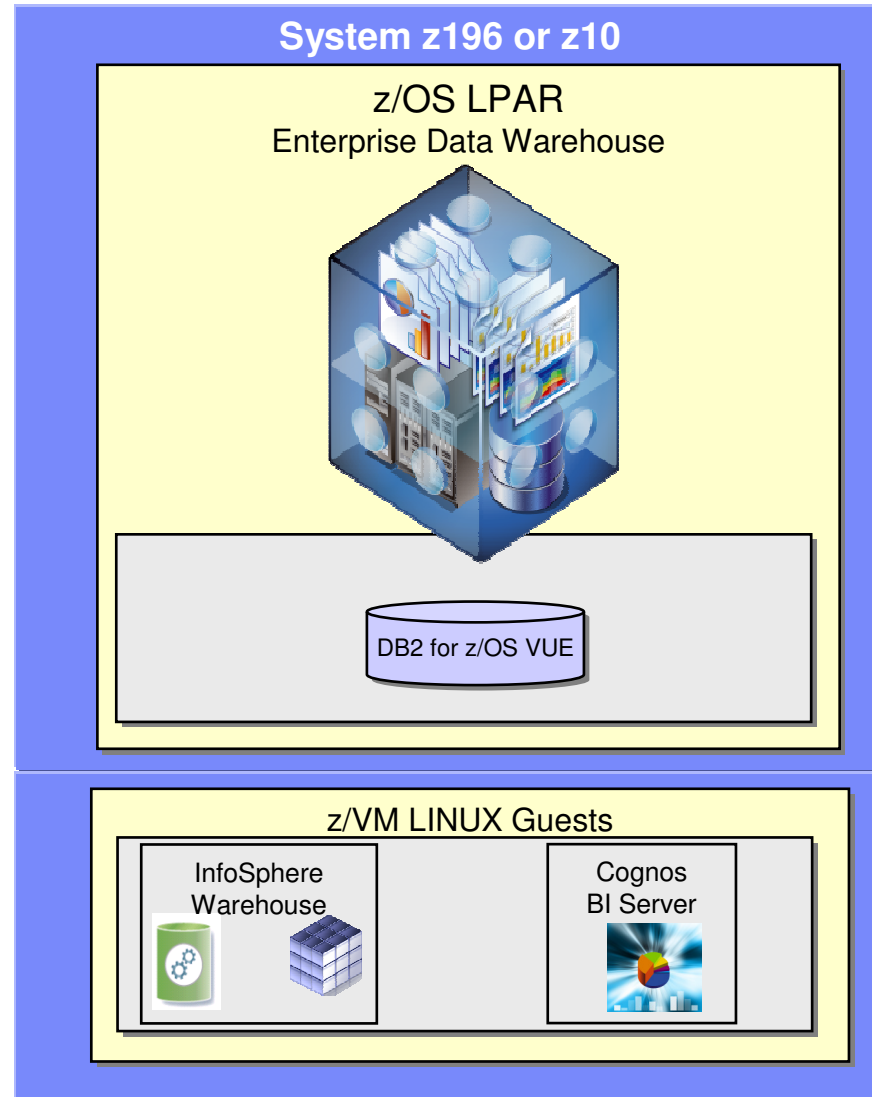
- InfoSphere Warehouse Enterprise Edition 9.7.2
- Cognos 8.4.1
- AIX
- IBM POWER 740
- IBM System Storage DS3500
- Default SSD using Blue Darter with optional further SSD expansion
- Smart Analytics System 7700 Control Console



# 7700

# IBM Smart Analytics System Solución para Mainframe Modelo 9600

# IBM Smart Analytics System para *Mainframe*





Modelos de Industria



Warehouse Packs



## Modelos de Industria

- Enfoque y Metodología probada.
- Incluyen el conocimiento de la industria.
- Más de 500 clientes.

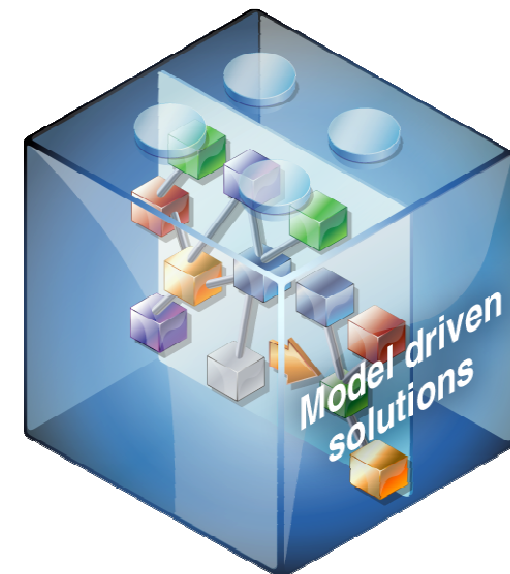


## Valor para el Cliente

- Reducen el análisis y el diseño de requerimientos en más de un 40%.
- Acelera la obtención de la aprobación por los interesados (*stakeholders*) en más de un 50%.
- Cubren en media el 85% de las necesidades de datos de un cliente.
- Reduce el tiempo comparado a un proyecto de *data warehouse* a medida.

## InfoSphere Warehouse Packs

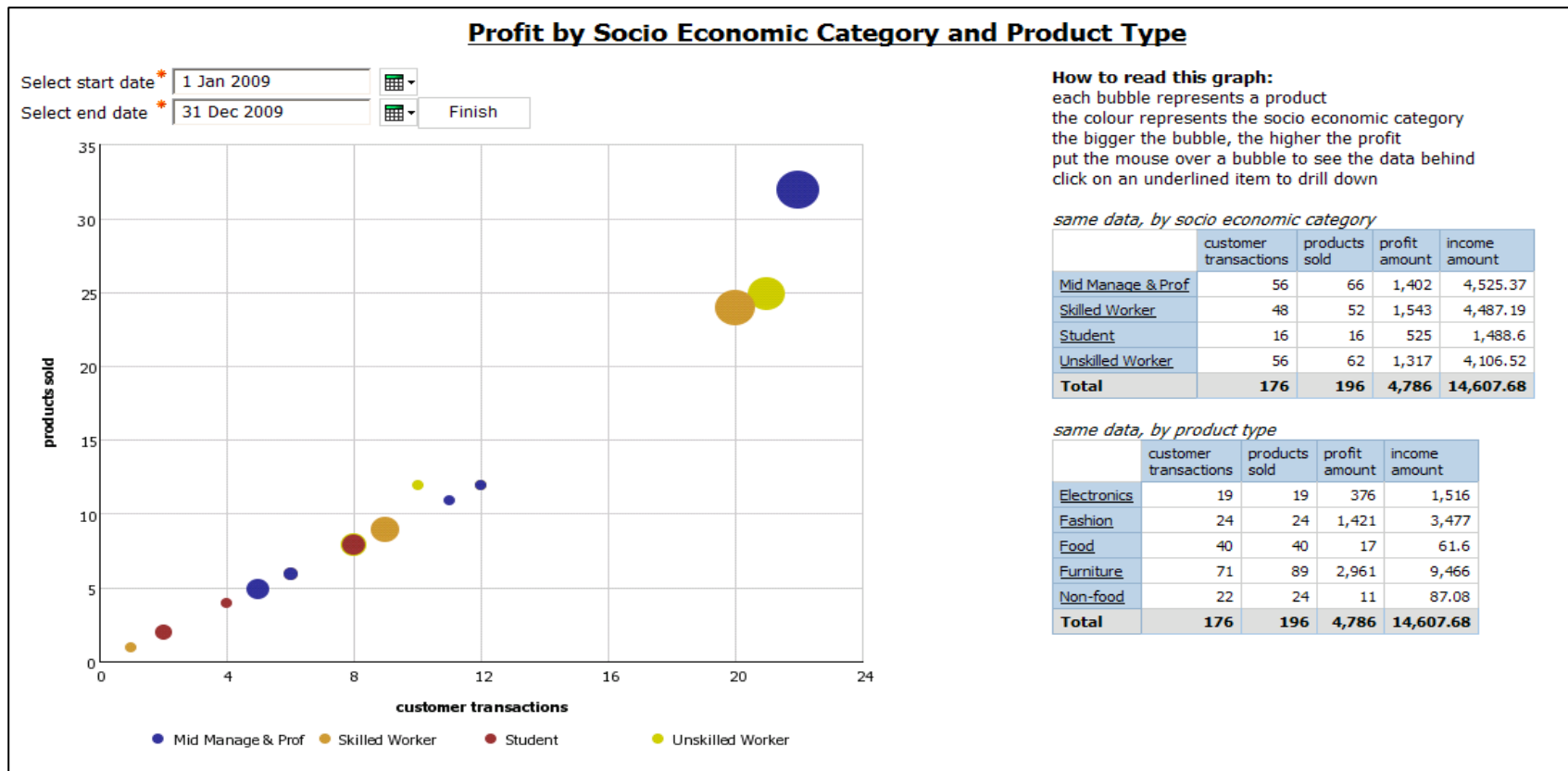
- Un conjunto de estructuras de datos e informes en las áreas de:
  - *Customer Insight*
  - *Market and Campaign Insight*
  - *Supply Chain Insight*
- Permiten el desarrollo rápido de Informes de BI desde una estructura compartida y extensible de *data warehouse*.
- Se pueden aplicar a cualquier solución de InfoSphere Warehouse o IBM Smart Analytics System.



# Customer Insight Pack

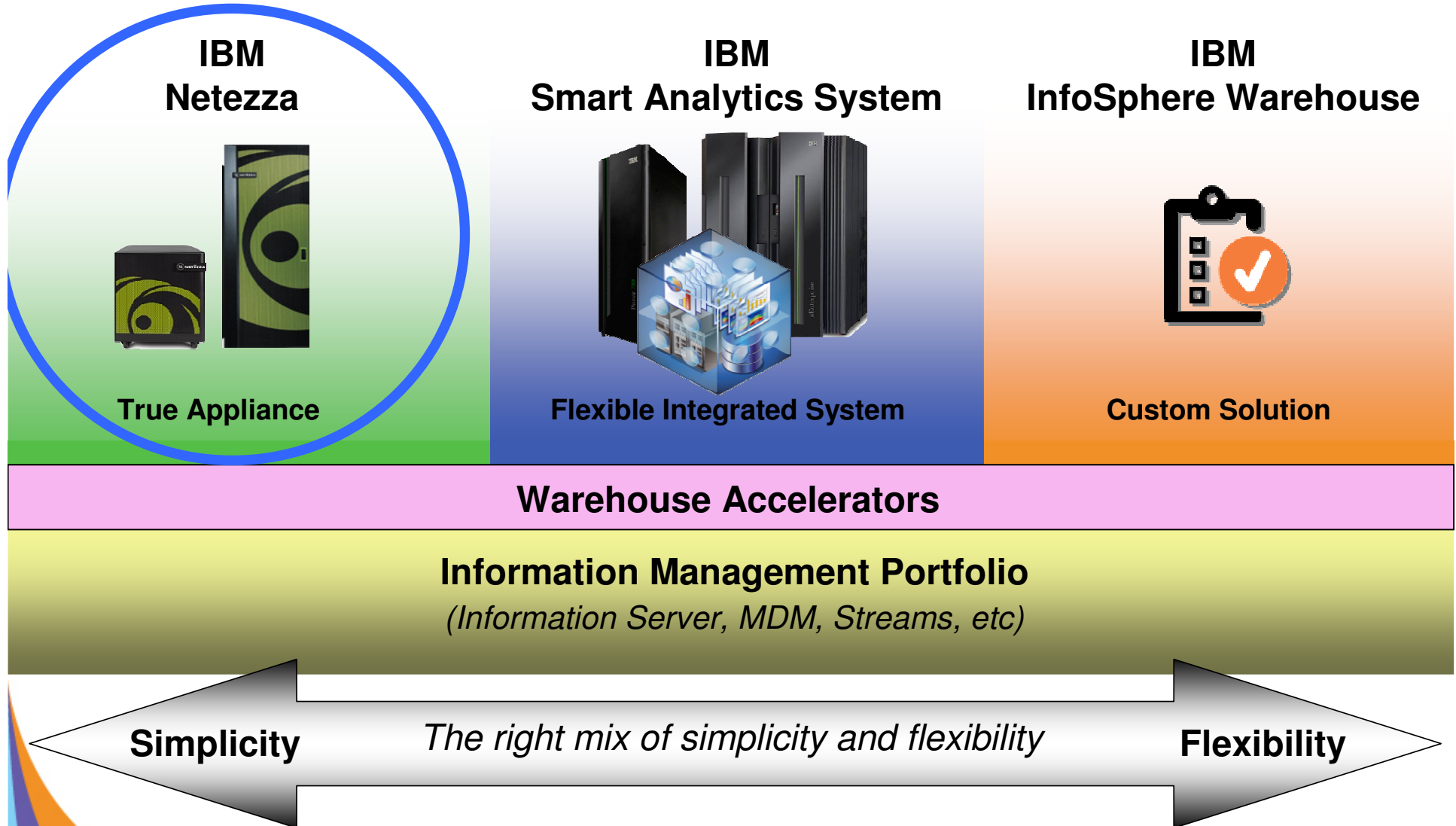
## Ejemplo de Informe: Beneficio por Categoría Socioeconómica.

Para **Customer Insight** se proporcionan más de 20 informes predefinidos con Cognos Reports.





# Soluciones de IBM para Data Warehouse



# IBM Netezza

## Who is Netezza

Leading provider of data warehouse & analytics appliances

Leader in the Gartner Magic Quadrant (2009)



## Global Offices

North America / EMEA / Japan / APAC

## Global Industries Served

- Retail
- Telecommunications
- Digital Media
- Health & Life Sciences
- Financial Services
- Government
- Energy

## What we do

- Performance leading appliances
- Proven enterprise-class solution with tens of PB of customer data.
- Rapid time to value
- Data warehousing & analytics

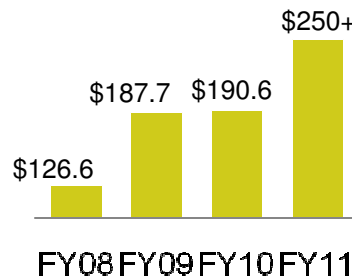


**Employees**  
500+

**Customers**  
375+

## Revenues

\* \$ Millions



\* Fiscal Year Ending Jan 31

## Acquired by IBM

Netezza is part of IBM Information Management

Backed by IBM's world-class resources

## Philosophy

Easy to do business with



## 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.



## Simplicidad



# Delivers out-of-the-box performance with **no tuning**

*“Allowing the business users access to the Netezza box was what sold it.”*

Steve Taff,  
Executive Dir. of IT Services



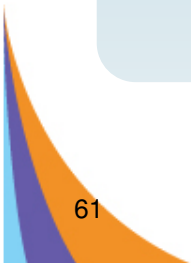
# Integración Perfecta en el Ecosistema de Business Intelligence

Interoperabilidad Certificada con los proveedores líderes

Data Integration

Business Intelligence and Analytics

Advanced Analytics and Data Mining



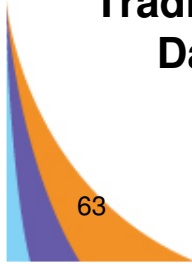
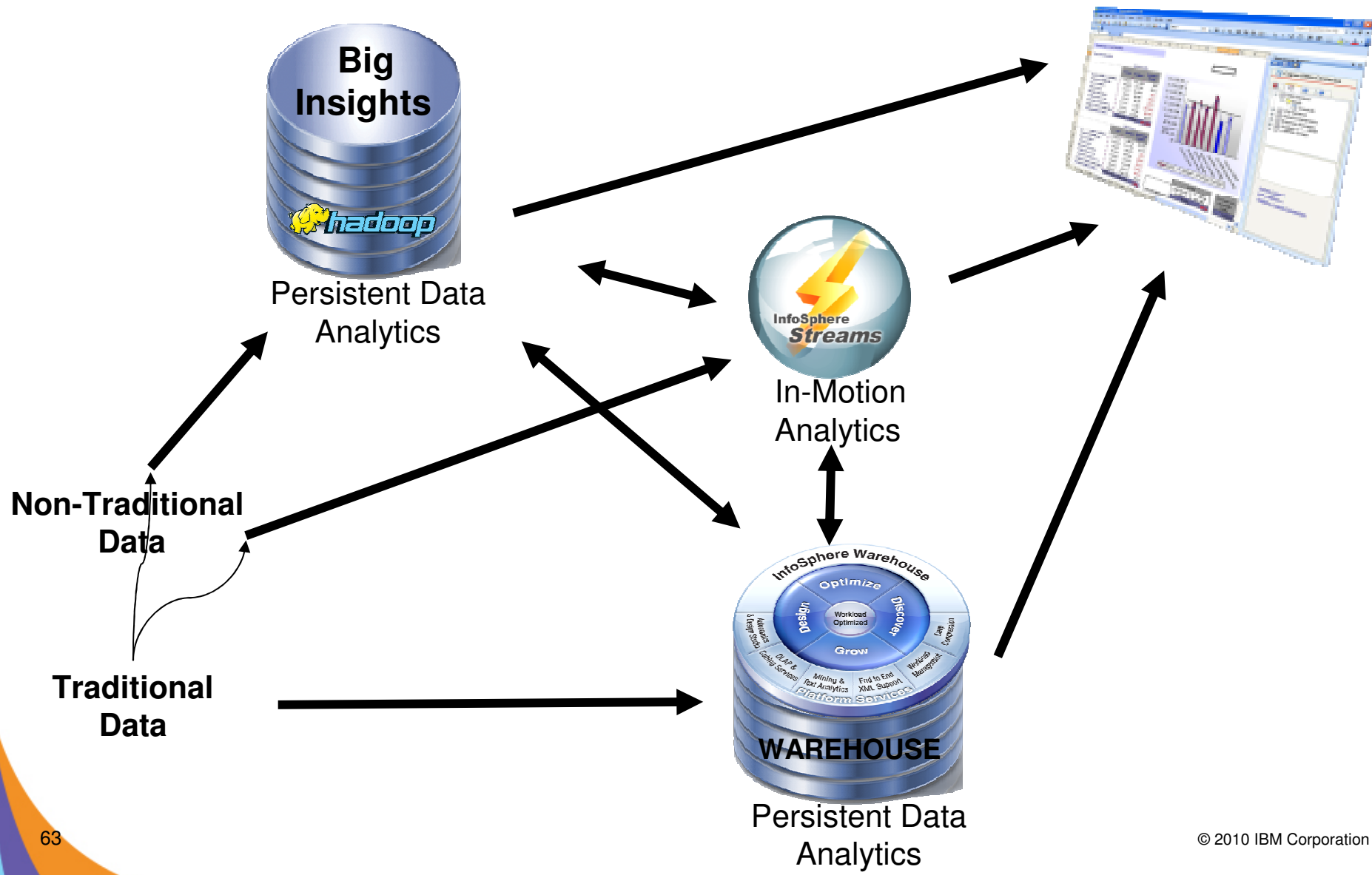


Infosphere BigInsights  
(hadoop)



Infosphere Streams

# Infosphere Big Insights e Infosphere Streams



## Conclusiones

*“One size does not fit all”*

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



*Agradecido*  
*Agradecido*  
**Muchas Gracias**  
*Agradecido*  
*Agradecido*



# INFORMATION INTEGRATION & GOVERNANCE FORUM 2011

*Delivering Trusted Information for Smarter Business Decisions*

