



IBM Software Group

# Data Warehousing on System z

## Integrating Business Information and Warehousing



**DB2** Information Management Software

List  
speakers

© 2005 IBM Corporation

# Agenda

- **Why Data Warehousing on System z**
- **System z support for Business Intelligence**

# Why Data Warehousing on DB2 and System z? Why Now?

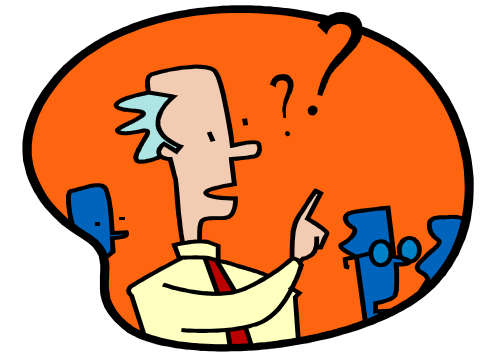


## System Z Customer Requirements

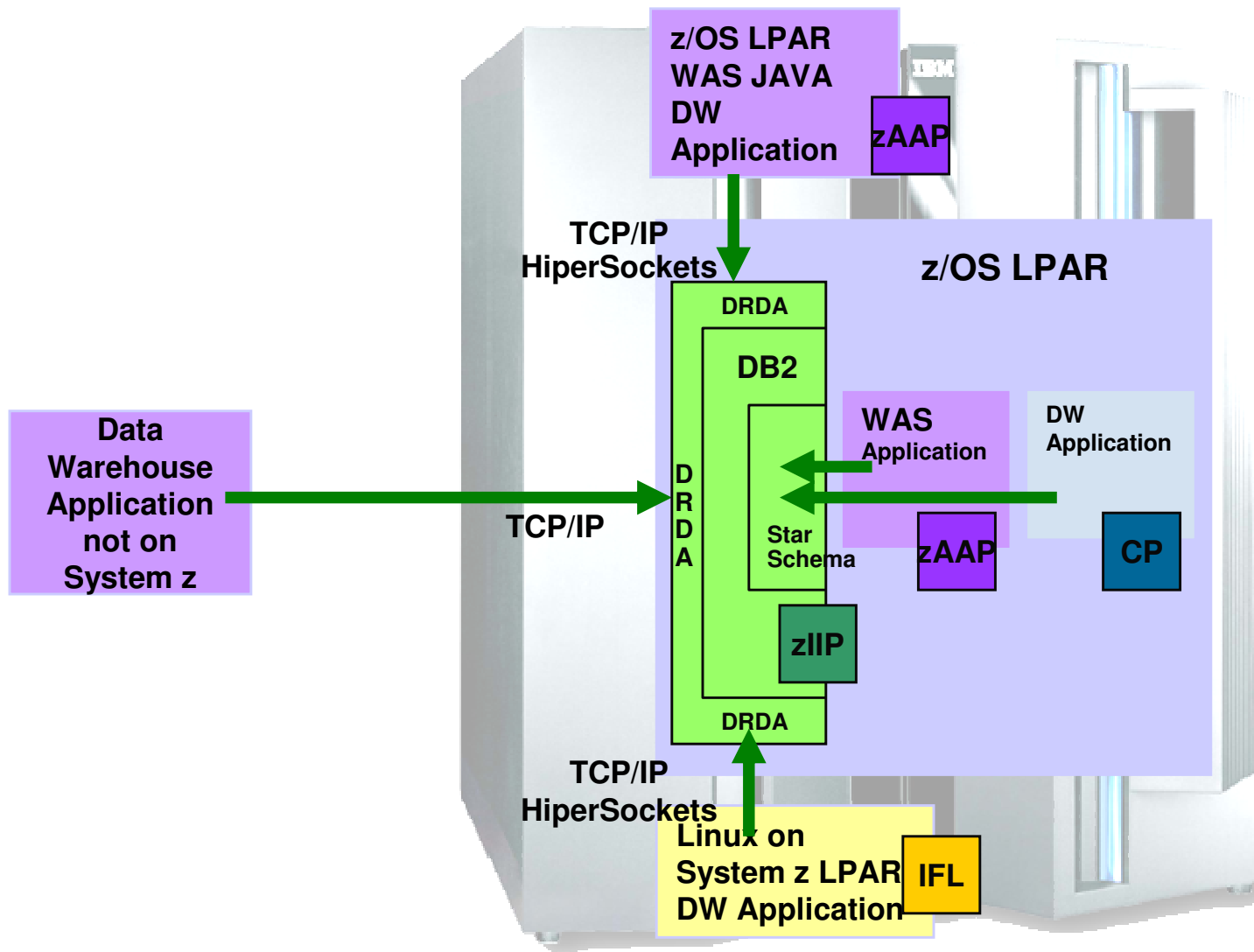
- **Increasing demands for sophisticated analysis with real time operational data — *BI is becoming mission critical***
- **High requirement for system, platform and data security**
- **Requirements for compliance across end-to-end data integration and analytic components**
- **Desire to optimize & leverage existing z infrastructure and skills**
- **Existing solution requires complex and costly data movement**

# When is System z the Preferred Platform?

- **Need true real-time Operational Data Store (ODS)**
  - Operational data is on System z
  - Data must virtually be in sync with the operational data
  - Availability, Security and Resiliency needs are high
  - Meet auditable data warehouse requirements
  
- **Specialty engine allows for IT optimization**
  - zIIPs, zAAPs, IFLs.
  
- **Want to leverage & optimize existing z skills and investment**
  - Can balance workload based on business policies/needs (WLM), separate OLTP and DW workloads
  
- **Consolidation of distributed marts or DW to an existing System z Data Serving platform**



## Specialty engines work together



The IBM System z9 specialty engines can run independently or complement each other

(shown at left, instances where specialty engines can be employed)

# System Z Green Credentials

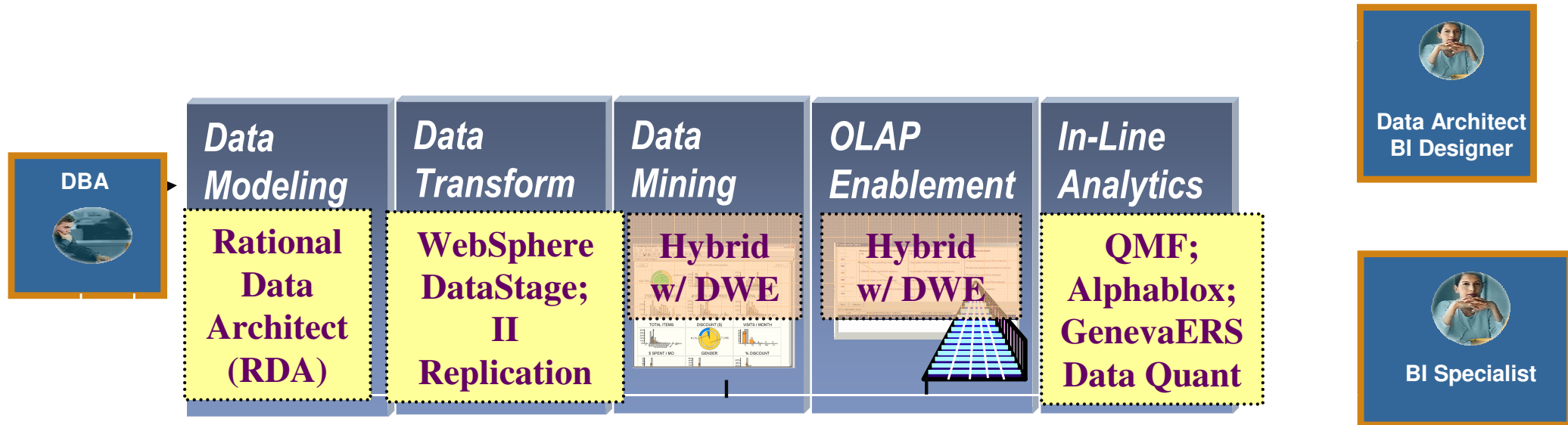
- **Environmental Issues are becoming a concern**
  - **Power**
    - Proliferation of servers increase demand on the external infrastructure
  - **Cooling**
    - Heat generated by increase number of servers has to be considered
  - **Space**
    - Will the expanding number of servers fit into the available space

# System z support for Business Intelligence Requirements





# Where IBM has invested for DW on z (so far!)



## System Admin Tools for

- Security and Regulatory Compliance
- Performance
- Application Management

# IBM DATA QUANT FOR Z/OS

- Adds compelling new Warehouse/Business Intelligence component to WH on z
- Visual Dashboards, Enhanced Graphical Reporting, Security and Personalization, SOA Layer, Enhanced Analytics
- Offers a “thick” client with DataQuant for Workstation, or a pure HTML, browser based client with DataQuant for WebSphere

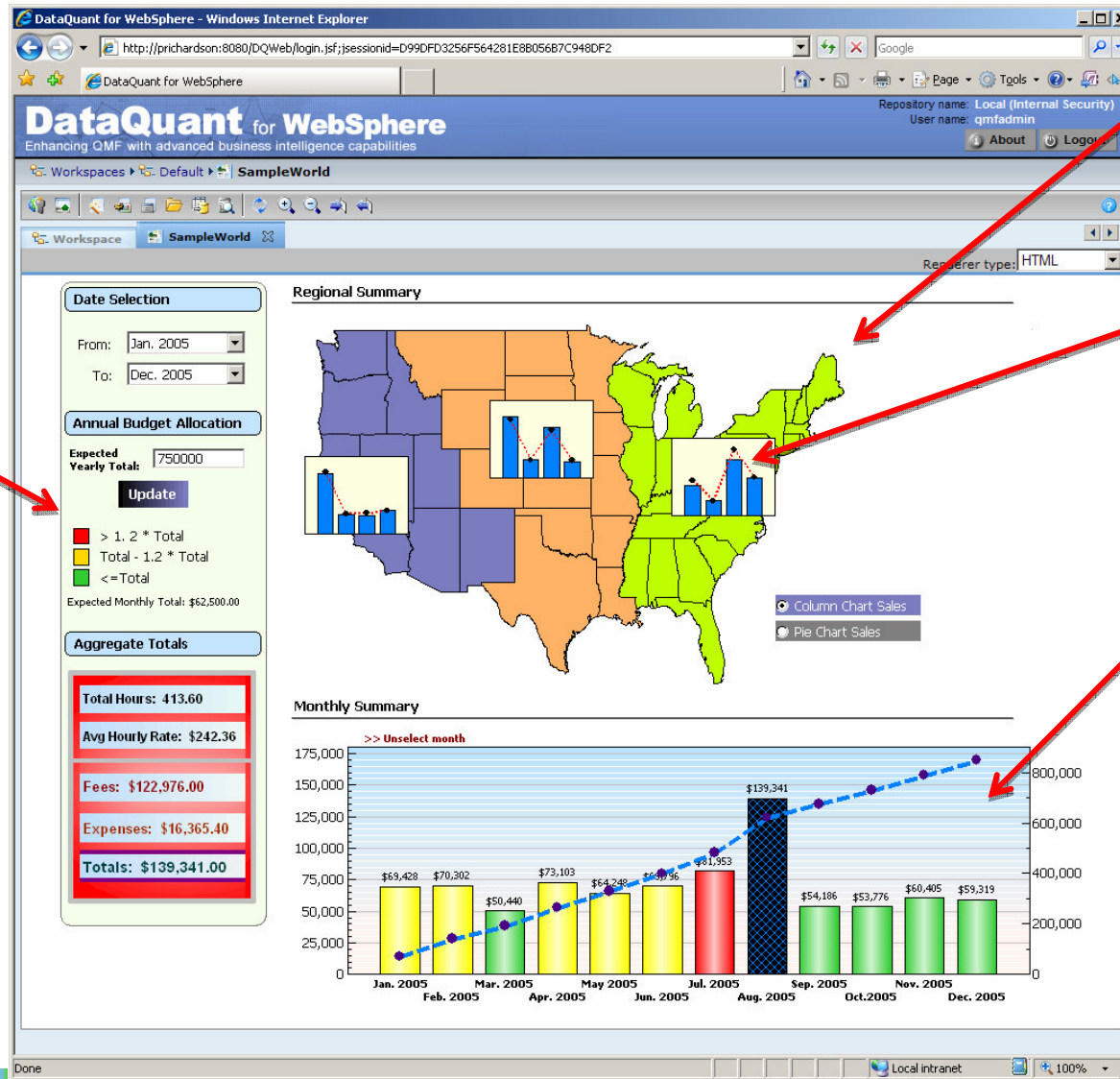


# IBM DataQuant Positioning within the BI Space

- **IBM DataQuant competes at the core BI layer and is a component of the broader IBM BI solution 'stack'**



# Visual Dashboards – BI that spans the Enterprise



Geospatial data from DB2 on Linux

Annualized sales data from Informix

Monthly transactional summaries from DB2 on z/OS

Budgeting and business rules from DB2 on Linux

# DB2 V8 enhancements benefiting DW

- **Usability, Availability & Scalability**
  - **Online Partitioning Changes, More Partitions**
  - **Schema Evolution**
- **Data Warehousing Support**
  - **Star Join Improvements**
  - **Materialized Query Tables**
- **Overall System & Query Performance**
  - **Locking Improvements**
  - **Multi-row INSERT & FETCH**
  - **Index only Access for VARCHAR**
  - **DDF Performance Improvements**

## DB2 V8: More Than 50 Features Relevant to BI

### Performance

- Data-partitioned secondary indexes (DPSI)
- Multiple DISTINCT clauses in SQL statements
- Reduced lock contention on volatile tables
- Coupling Facility lock propagation reduction
- Multi-row INSERT/FETCH
- REOPT(ONCE) to reduce host variables impact on access paths
- Index-only access for VARCHAR columns
- Backward index scan
- Faster short PREPARE
- IN access path performance
- DDF performance enhancements

### Business warehouse

- Sparse index for star join
- More tables in join
- Common table expressions
- Recursive SQL
- Materialized query tables

### Continuous availability

- Changing clustering index as online operation
- Elimination of BUILD2 phase of REORG with DPSIs
- Online schema evolution for many column types
- Volume-level, automated backup and recovery
- CI size larger than 4 KB
- More log data sets
- Conditional restart enhancements
- Support for synchronizing log point

### Architecture

- Unicode support
- Introduction of DB2 Connect
- DB2 Universal Driver for JDBC
- 64-bit virtual storage for most DB2 storage areas
- Up to 4096 partitions
- Longer table/column names
- SQL statements up to 2 MB
- ASCII precompiler

### Ease of use

- Clustering decoupled from partitioning
- New REORG option to reorganize all partitions in Reorg-pending state
- CREATE INDEX invalidates statements from dynamic statement cache
- Indexes created as deferred are ignored by DB2 optimizer
- LOB ROWID transparency
- Collecting distribution statistics on arbitrary sets of columns with RUNSTATS
- Fast cached SQL invalidation
- Automatic space management
- Statements IDs of cached statements as input to EXPLAIN
- Statement ID in IFCID 124
- Long-running non-committing reader alerts
- Lock escalation reporting
- Transaction-based DB2 accounting and workload management
- Stored procedures to facilitate database administration
- Network statistics with DB2 Connect
- DRDA ping
- Comments in dynamic SQL
- CTE-based optimizer hints



# What else is coming for DW on z?

## ■ SHORT TERM -

- DB2 9
  - Even more DB2 compatibility
  - Warehouse, reporting & optimizer enhancements
  - XML

## ■ MEDIUM TERM –

- Shark
  - Large table scans pushed down to Shark
  - Better price/performance
- Continue DB2 warehousing improvements
  - MQT Advisor
- Additional capabilities supporting DW on z
  - Cube Views support
  - Alphablox on Linux on z
- BCU for z/OS

## ■ LONGER TERM

- Additional capabilities supporting DW on z
  - Common administration
  - Data Mining
  - Rest of DWE capabilities



## DB2 9: Another Feature Rich Release for BI

### Performance

- New row internal structure for faster VARCHAR processing
- Fast delete of all the rows in a partition
- Numerous enhancements in 'smaller' LOB performance
- Fast LOB streaming
- Reducing log latch contention
- Deleting first n rows
- Skipping uncommitted inserted/updated qualifying rows
- Faster release of LOB locks
- Reducing data sharing overhead for global indexes
- Functional indexes

### Business warehouse

- Dynamic index ANDing
- Reduce temporary tables materialization
- Generalizing sparse index/in-memory data caching

### Continuous Availability

- Partition-by-growth as a means to remove non-partitioned tablespace size limit
- Full support for system-level backup and recover (automatic offload to tapes and individual objects recovery)
- Renaming SCHEMA and VCAT to facilitate fast database provisioning
- Rename index
- Reorganization of LOBs to reclaim space
- Online REORG enhancements
- Online REBUILD index

### Architecture/SQL

- Thin DB2 Connect Client
- FOR BIT DATA collating sequence (VARBINARY)
- Full JDBC compliance
- Enable Decimal Float data type (pre-conditioning)
- BIGINT data type
- Index compression

### Architecture/SQL (con't)

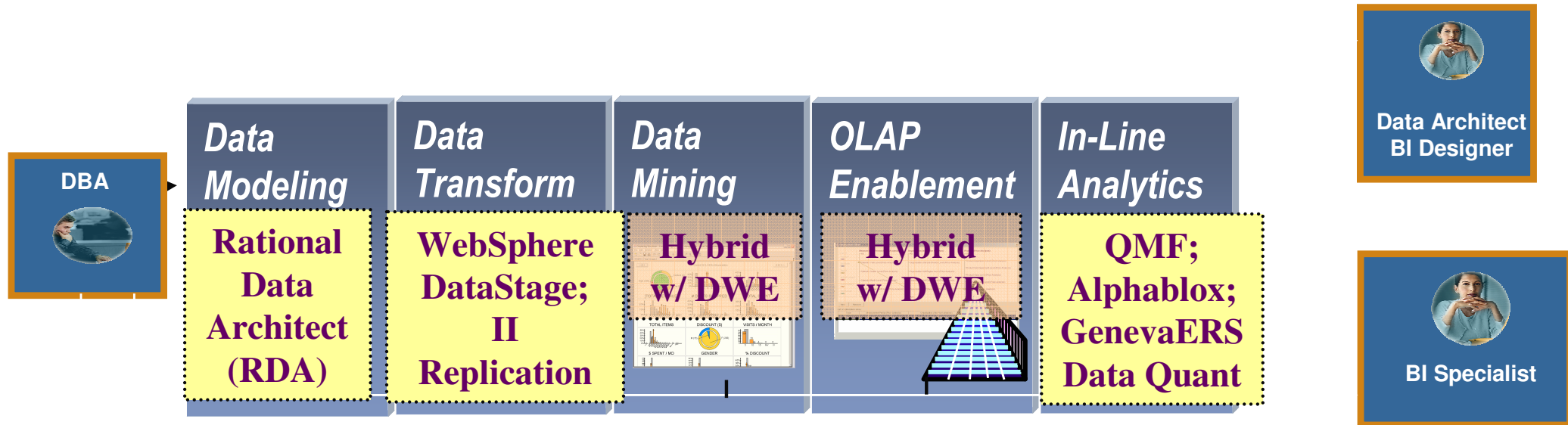
- Provide more VS relief for thread related storage (partially)
- Unicode support for all CLI functions
- MERGE statement
- SET operations

### Ease of Use

- Implicit objects creation
- Enhancing real time statistics (Optimization Service Center)
- Autonomic reoptimization
- Integration of Real Time Statistics tables into the catalog
- Simulating indexes in EXPLAIN (Optimization Service Center)
- More autonomic bufferpools tuning (WLM synergy)
- RLF support for end-user correlation
- TRACE support for end-user correlation
- Enhance tracing in DB2 Connect
- Identifying unused indexes
- Enhancing IFC for IRLM diagnostics
- DSNACCOR enhancements



# Where IBM has invested for DW on z (so far!)



## System Admin Tools for

- Security and Regulatory Compliance
- Performance
- Application Management

# Alternative architectures

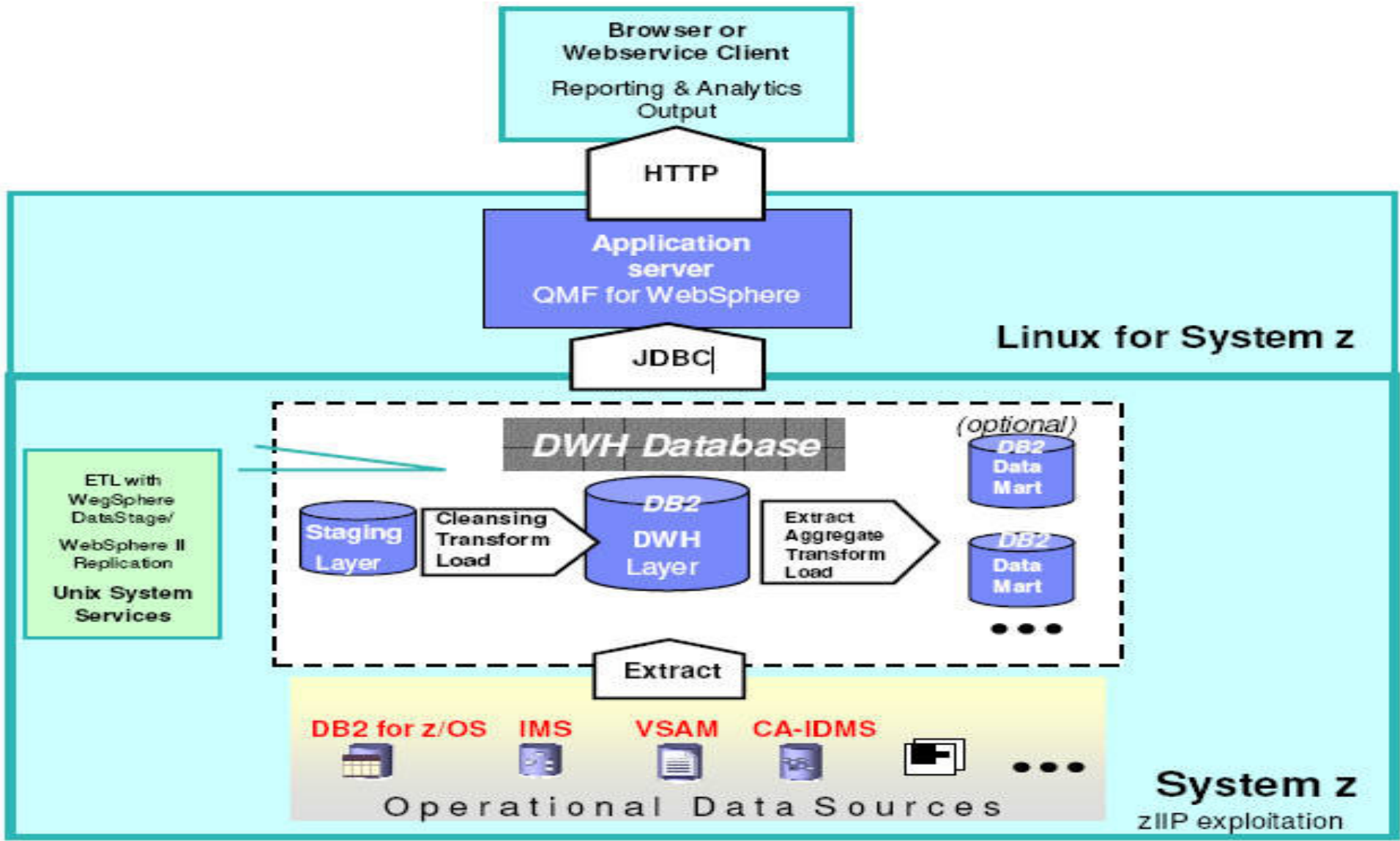
## ■ “Pure” System z BI Solution

- ODS, Data Warehouse & Datamart(s) in DB2 z/OS
- End User Tools (e.g. QMF, Business Objects, Cognos, Data Quant) access DB2 z/OS directly (fat client implementation) or via browser (web server implementation)

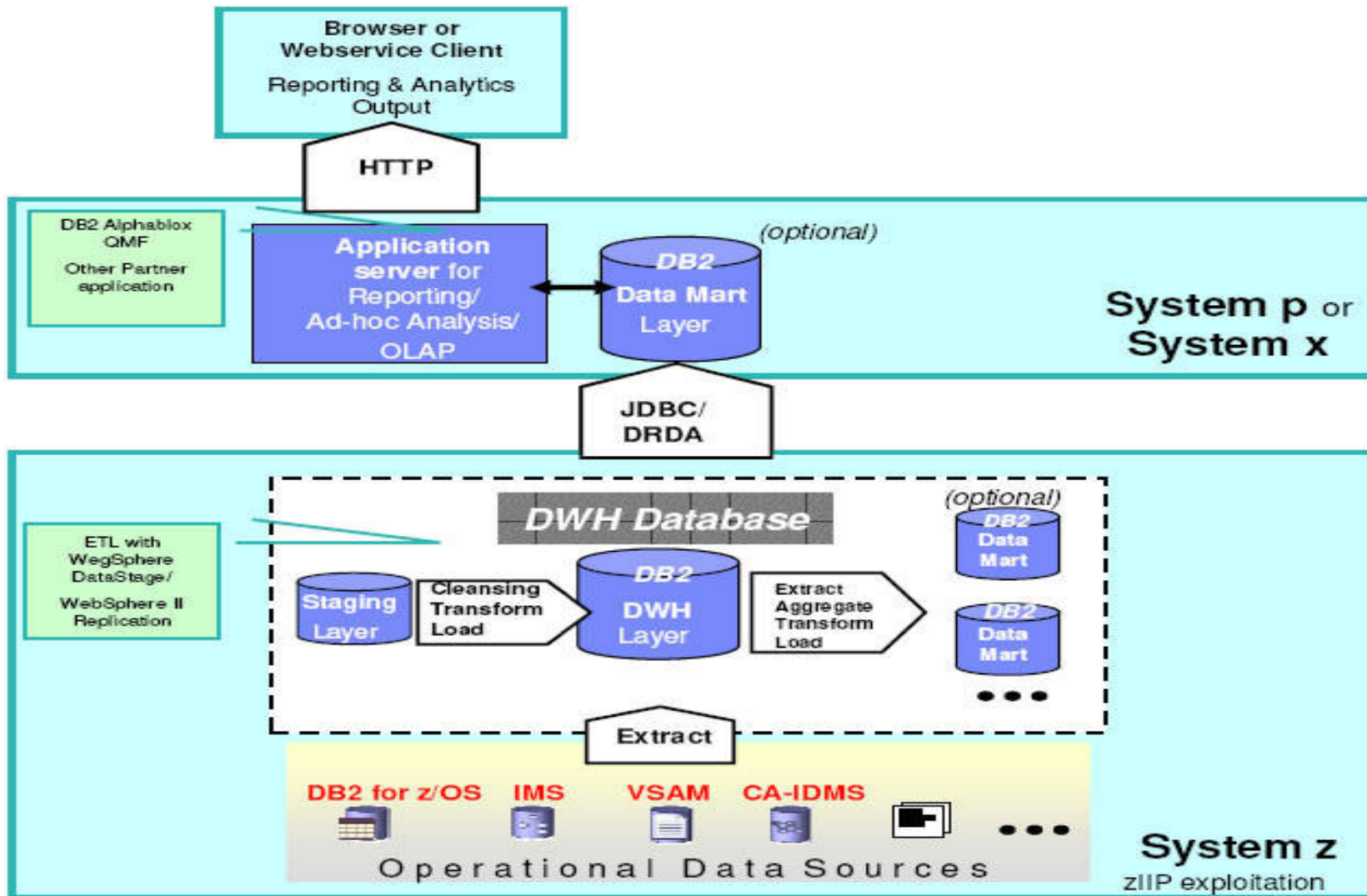
## ■ “Hybrid” BI Solution

- ODS & Data Warehouse in DB2 z/OS
- Reporting solution runs on distributed WAS, e.g. Alphablox, QMF, Cognos ReportNet, Business Objects Server
- Relational, Multidimensional (OLAP) and Statistical Datamarts on pSeries and/or xSeries supporting End User Tools, e.g. Hyperion Essbase, Cognos PowerPlay

# z/OS Warehousing architecture – Pure System z



# z/OS Warehousing architecture – Hybrid



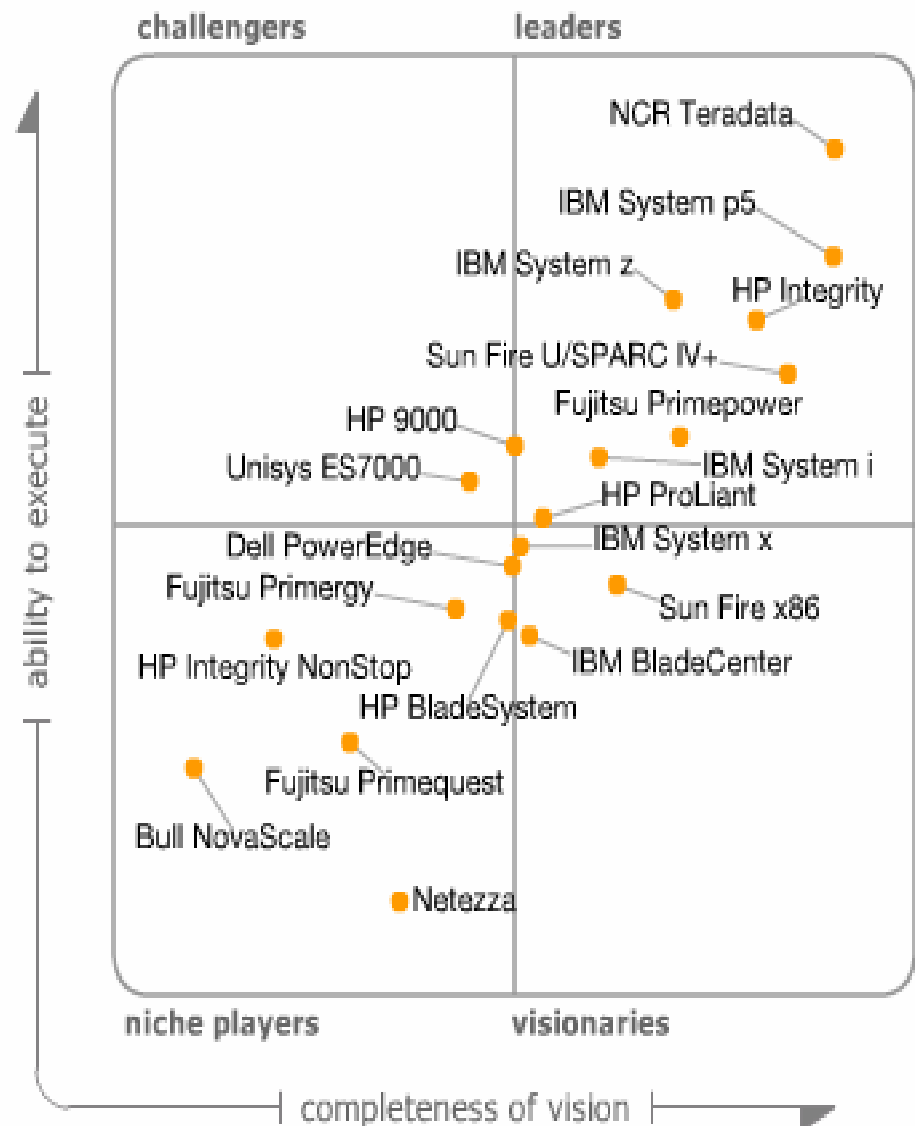
# Summary

## ■ System z

- Supports Compliance
- Delivers real time solutions
- Supports mixed workload requirements
- Leverages existing investment
- Delivers world class security
- Unparalleled reliability, availability and security
- Addresses Environmental Issues
- Provides tooling to interrogate and manage the data

# Summary

- **System z has returned to the Warehouse space and is now a leading player in the Data Warehouse and Data Integration market**



As of August 2006

Source: Gartner (August 2006)

# Back up

# What is Data Warehousing?

- **Data Warehousing** provides the underlying data storage facilities required to support any BI Solution.
- **An Operational Data Store (ODS)** is a subject oriented database organized by business area. It is up to date (vs historical) and detailed (vs summarized).
- **A Data Warehouse (DW or sometimes WH)** is a multi-subject oriented database populated from operational systems and/or ODS's. It is historical (vs point-in-time) in nature and typically contains detailed data. It is often looked upon as the single source of corporate "truth".
- **A Data Mart (DM)** is a database designed to support the analysis of a particular business subject area. Data has usually been transformed and aggregated from the source DW or operational system. Data Marts can be relational, multidimensional (OLAP) or statistical in nature.

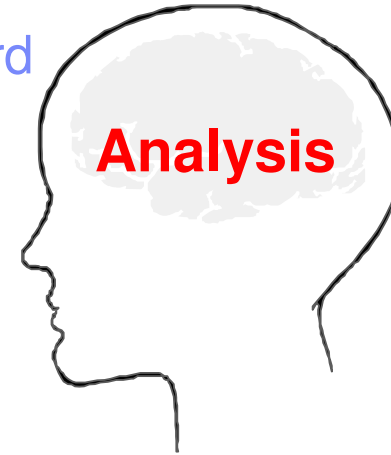


# Understanding the terminology

- **Business Intelligence (BI) and Data Warehousing (DW) are sometimes used interchangeably**
  - **Typically BI includes end user tools for query, reporting, analysis, dashboarding etc.**
  - **Both concepts depend on each other**
    - BI almost always assumes a Warehouse (WH), Operational Data Store (ODS) or Data Mart (DM) exists with timely, trusted information
    - A DW depends on end user tools that turn data into information.
- **Both terms (DW and BI) address desire for timely, accurate, available data delivered when, where and how the end users want it**

# What is Business Intelligence?

In a word



**Business Intelligence**: the process of gathering, consolidating, and analyzing data from multiple sources for strategic and tactical decision making.

- **derives new value from transactional data**
- **supports strategic planning, monitoring, and efficiency**
- **delivers *knowledge* of the customer, suppliers, and channels**
- **unifies the enterprise with actionable information for operational Business Intelligence**
- **Top quality BI relies on a *secure, high performing, warehouse oriented* infrastructure to deliver Information on Demand—based on *open standards***

## Specific components of DW/BI

- **Data Transformation (ETL)- Extract/Transform/Load** removing data from its source, changing the it into a usable format and putting it into a target WH, DM or application
- **Data Modeling** —mapping data into a relationship structure to make it more usable by an application or by analytical processing
- **Data Mining** —identifying non obvious patterns and relationships in data
- **OLAP Enablement** —Online Analytic Processing, as opposed to OLTP (Transactional); looking at vast quantities of data in aggregate form across multiple business dimensions (product by region by channel by date)
- **Query & Reporting** —accessing data and presenting it in an understandable format
- **Analytics**—manipulation of data to make or validate a business decision; **ad hoc analysis** typically takes the results of one query to form the basis of the next query—on the spot
- **System Management Tools** to keep the IT environment sound and productive

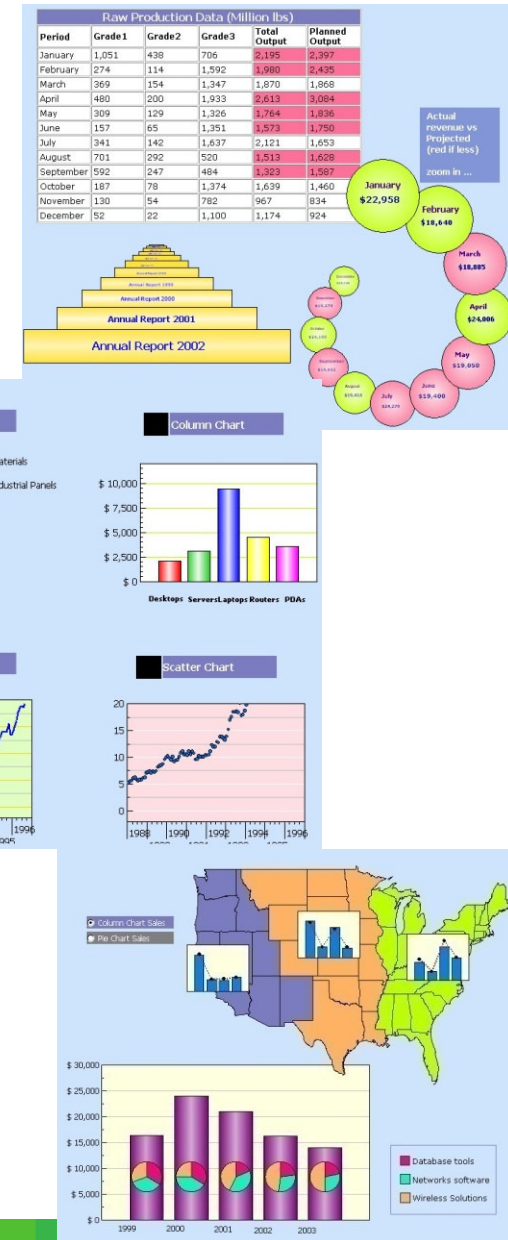
# Visual Dashboards (cont.)

- **Supports over 20 'stock' visual layouts**

- Line and bar charts
- Geographic maps
- Pie charts
- Horizon charts
- Hierarchical charts
- Tabular layouts

- **Flexible rendering**

- Motifs can be nested indefinitely (e.g. pie charts upon bars within maps etc.)
- Customized graphical 'templates' – users can build their own chart types



# IBM Tools and Technologies to support for Compliance

- **Superb z9 Cryptography**
  - **Hardware assisted**
  - **End-to-end**
- **IBM tools to assist in Compliance**
  - **DB2 Audit Management Expert**
    - ✓ Centralized easy-to-use support for your auditing needs Enables auditors to collect, view, analyze and report on data and save it into an audit repository
    - ✓ Provides accurate record of who did what, when and where
  - **DB2 Test Data Base Generator**
    - ✓ Use transformation capabilities to protect sensitive production data for use in test environments
    - ✓ Take one more area off the auditor's checklist
  - **DB2 Data Archive Expert**
    - ✓ Large amounts of data need to be kept to comply with retention requirements
    - ✓ Regulations require a centralized approach to archiving as opposed to application by application
    - ✓ Operational benefits to archiving inactive production data

# What DW Capabilities Exist Today For System z

- **DB2 V8**
  - Functional and performance enhancements
  - Easier online reporting and data management capabilities
  - DB2 family compatibility – including MQTs
- **Integration**
  - Ascential Data Quality, Data Stage, WebSphere II Replication & Classic Federation
- **Analyze/Report**
  - Data Quant, Alphablox, QMF, Brio, Hyperian, Business Objects, SAS, IBI
- **Security and Regulatory Compliance**
  - DB2 Data Archive Expert, DB2 Test Database Generator, DB2 Audit Manager Expert, IBM Encryption for DB2 and IMS Databases
- **Performance**
  - IBM Tivoli Omegamon XE for DB2 Performance Expert on z/OS, DB2 High Performance Unload (HPU)
- **Application Management**
  - DB2 Table Editor, DB2 Web Query Tool, DB2 Automation Tool, DB2 Query Monitor

# DW Issues and how System z can solve them

- **Continuous (near-real-time) data loading**
  - **Parallel Sysplex makes it easy to have DW and OLTP system on the same hardware**
  - **Workload Manager to prioritise resource for ETL jobs**
    - Prioritise Reporting suites
    - Prioritise Ad hoc Users
    - Ensures OLTP has higher priority
- **An increasing level of analytics and BI – Oriented functionality in OLTP**
  - **Summary and analytical output is handled by parallel access to OLTP and DW**
  - **MQTs within the ODS deliver data warehouse capabilities with pre-processed OLTP data**
  - **DB2 ensures that materialized data remains actual**

# DW Compliance

## ■ System z Supports Compliance

- **Single version of the truth**
  - Data confined to one location regardless of where the application is running
- **Single process to manage compliance**
- **Fewer people to manage compliance process**
  - DB2 Audit Management Expert
- **Single place for auditors to look**
- **Industry specific data models assist DW compliance**