



*The future runs on System z*

# **New Opportunities for Enterprise Data Warehousing on System z**



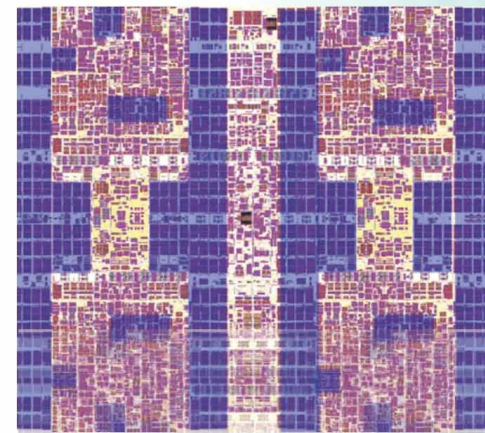
***Beth Hamel***

***Product Manager for Data Warehouse Solutions on System z***

# Agenda

- **Why Data Warehouse and Business Intelligence on System z?**
  - Industry Trends – Changing Business Requirements
  - System z Features
- **An Introduction to Data Warehousing on System z**
- **Business Intelligence and Data Analytics on System z**
- **Best Practices and User Experiences**

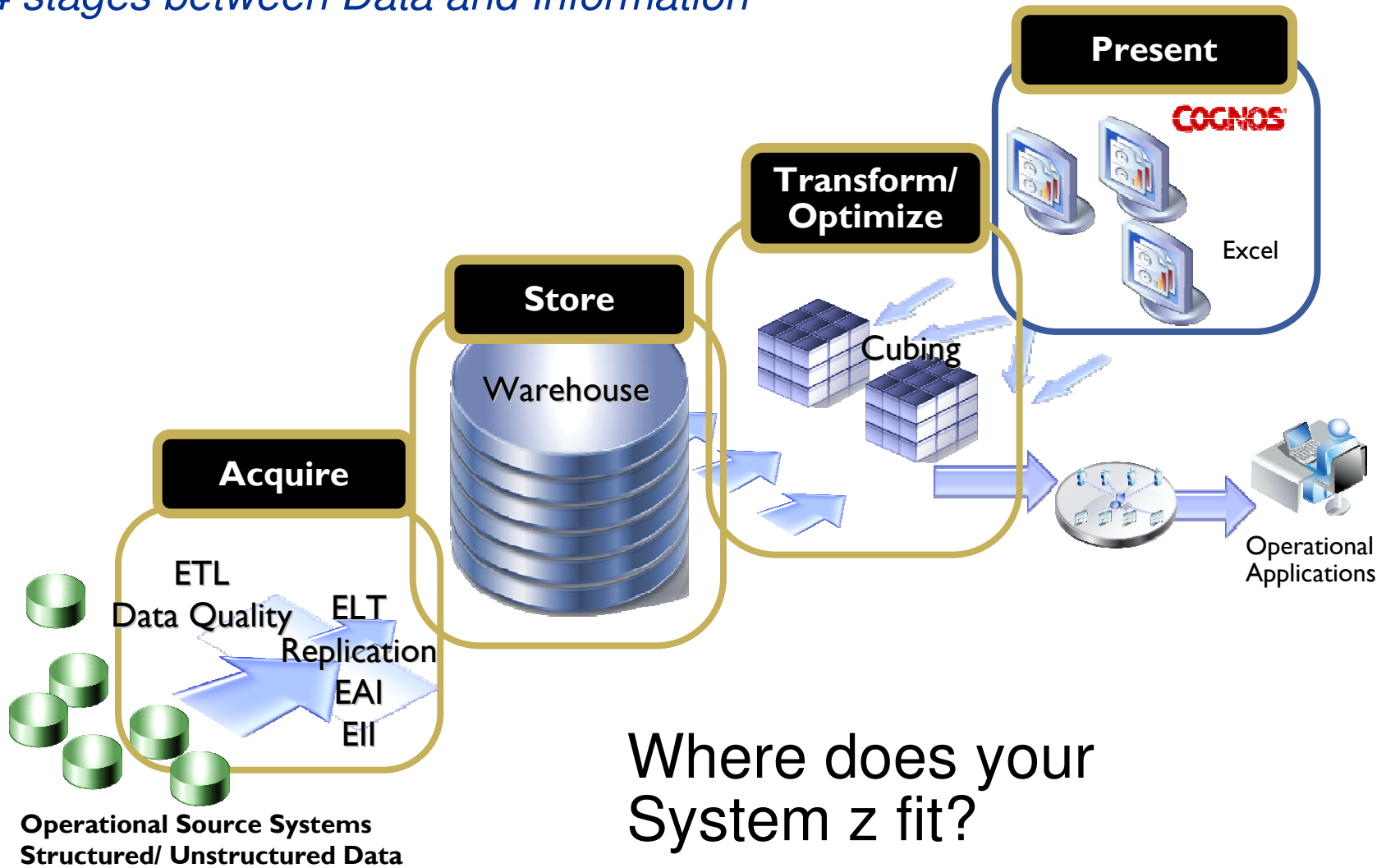
# *Why System z for Data Warehouse and BI?*



**Enterprise Quad Core  
z10 processor chip**

# Data Warehouse/BI Basics

- 4 stages between Data and Information



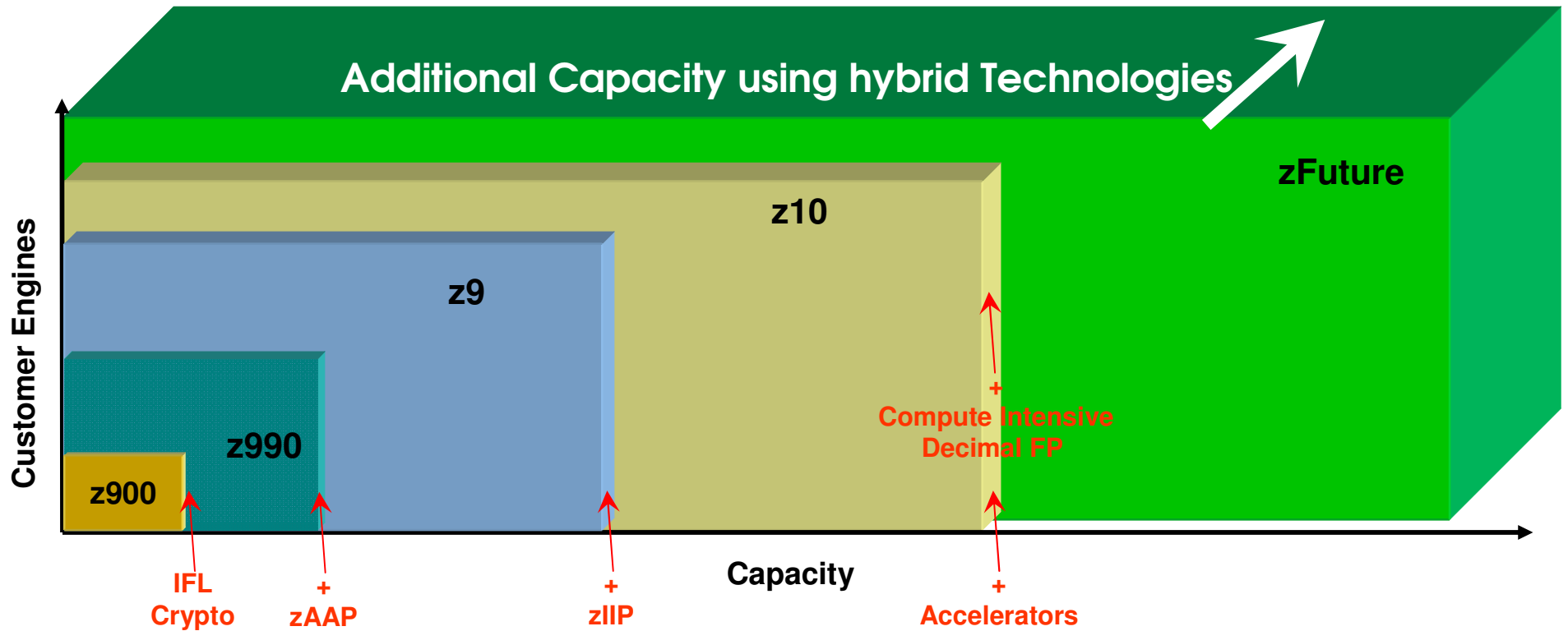
Where does your System z fit?

# Industry Trends that Favor the System z Platform

- **Cost efficiencies of a consolidated, virtual environment**
  - Reduced: labor costs, energy consumption, data center footprint
  - Improved: ability to manage operations, upgrades, performance
  
- **Mixed workload environments and operational business intelligence**
  - When warehouse and analytics become more operational, System z platform qualities of service become more critical:
    - Availability, Reliability, Security
  
- **Compliance pressures**
  - Centralized data can dramatically improve ease of data governance by reducing the number of servers and copies of data to be managed
  - Single version of the truth

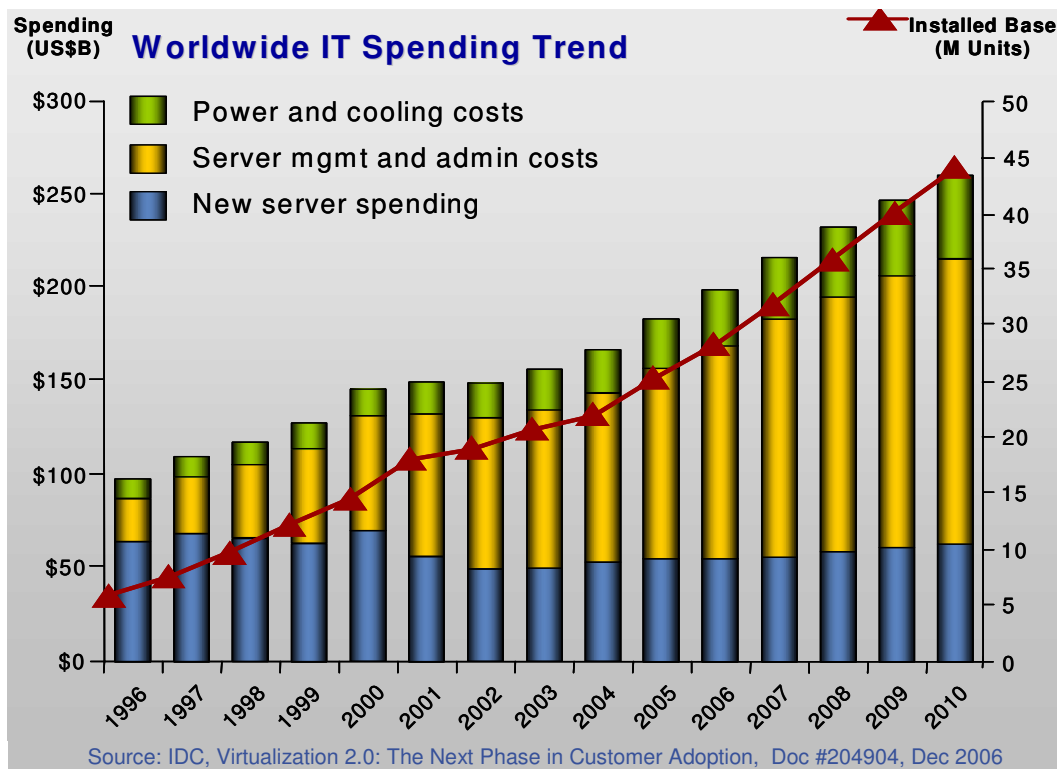
# Processor Performance and Scalability

***System z is no longer your father's mainframe system!!!***



# Cost effective alternative to masses of commodity servers

- *Management/Administration costs are already exceeding new server spending*
- *Power and energy costs growing dramatically*



**2000 –**  
Raw processing  
“horsepower” is the primary goal, while the infrastructure to support it is assumed ready

**2010 –**  
Raw processing  
“horsepower” is a given, but the infrastructure to support deployment can be a limiting factor

# Cost Benefits of System z

## Reduce Operational Costs

- **Save on storage using DB2 for z/OS System z hardware based data compression without compromising performance**
  - Typical savings in the 50-60% range
  - V9 adds index compression
- **Save on labor, energy, and floor space through consolidation to System z**
  - Save up to 85% on floor space\*\*
  - Save up to 80% on energy costs\*\*
  - Save up to 50% on labor resource\*\*

## Reduce overall Risk Exposure

- **Highly sensitive personal or financial data is secure and auditable**
- **DB2 for z/OS data sharing and GDPS provide unparalleled availability**
- **Mixed workloads benefit from the capability of the world class System z workload manager**
- **Reduced latency and duplication provide a sound basis for timely and consistent business analysis**

## Position for Growth

- **Application growth on System z scales well, while cost savings increase**
- **Parallel Sysplex and DB2 for z/OS data sharing enable incremental growth**
- **System z provides a non-disruptive upgrade path; System z tech dividends make that upgrade path more cost effective**

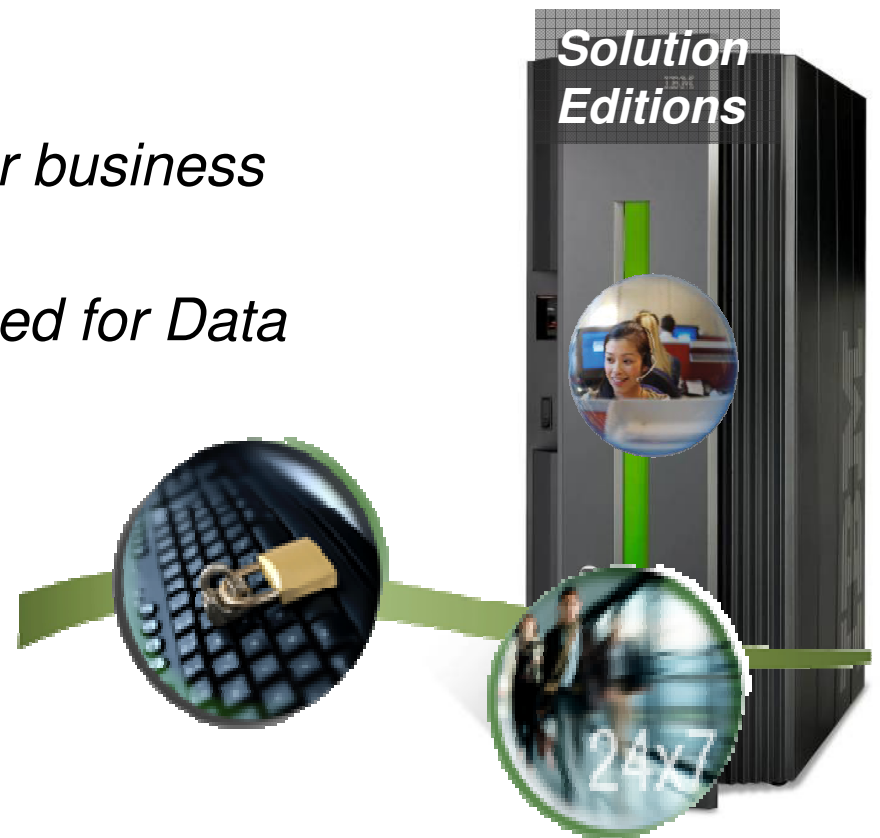
\*\* Based on calculations of savings realized in internal IBM System z consolidation projects



# Solution Edition for Data Warehousing –

*Legendary System z quality, priced aggressively to fit within a shrinking budget*

- **Complete Offering:**
  - *Solution – based offering*
  - *Customized hardware to meet your business requirements*
  - *Pre-determined software stack tuned for Data Warehousing workloads*
  - *Maintenance for the hardware*
- *For a highly available, energy efficient*
- *Data Warehousing infrastructure designed for growth*



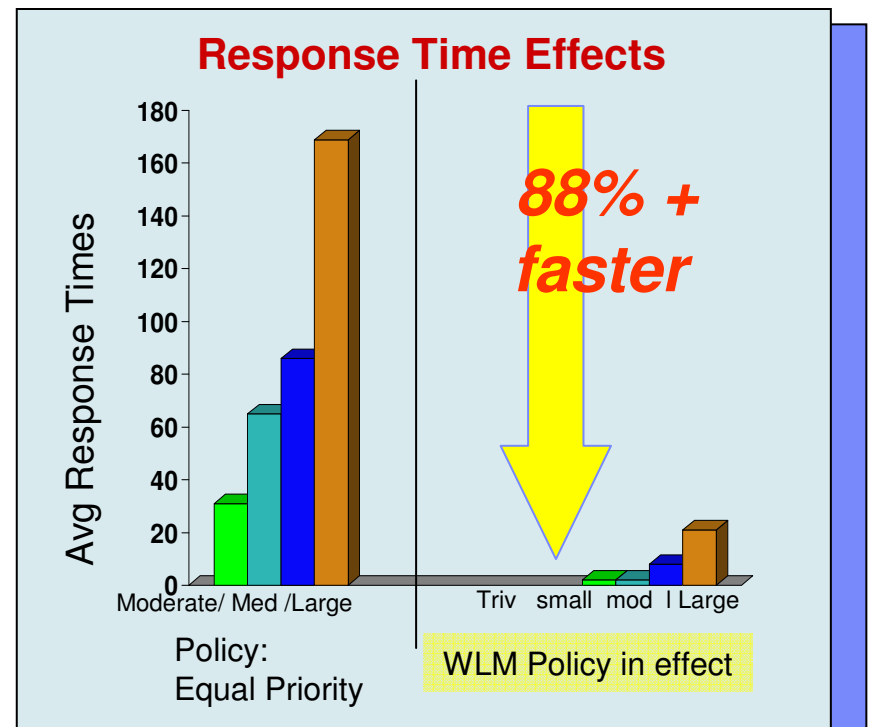
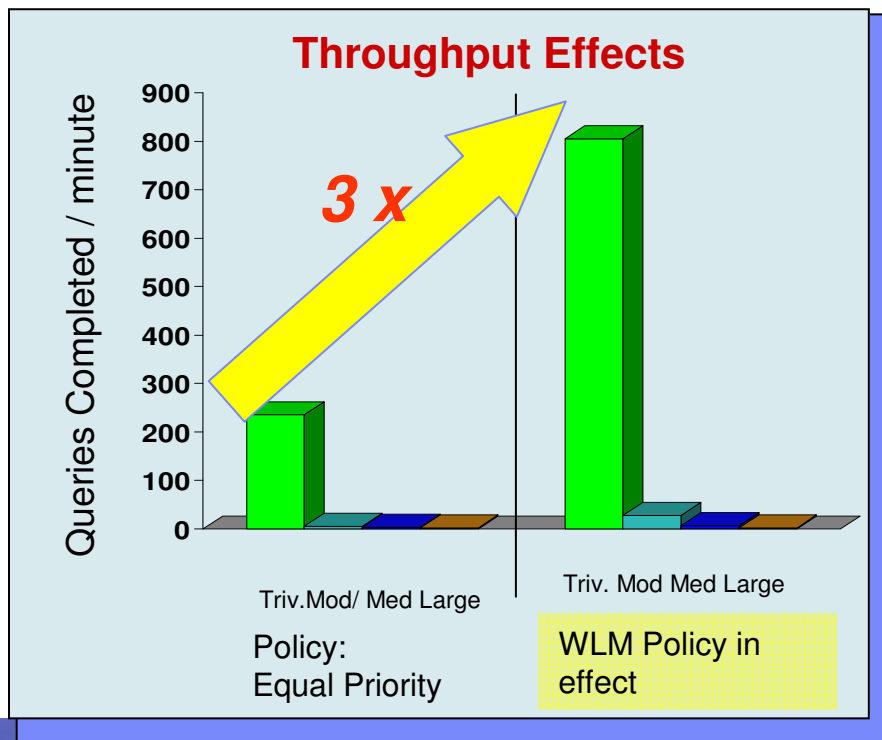
## The Changing Data Warehouse Market

- The data warehouse is a mission-critical system, with data warehouses serving in an increasingly mixed workload capacity, including as a data source for online applications. "Deep mining" analysts and business analysts are running less-structured but equally complex queries and fast running tactical queries, each with differing service-level expectations. These differing workloads are all competing for CPU, memory and disk access. At the same time, data latency continues to progress from batch to continuous loading demands.

# Proven Ability to Handle Mixed Workloads

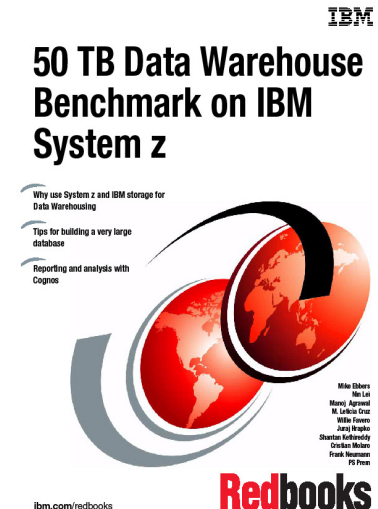
*The System z WLM is the most mature and sophisticated workload manager available today*

*Granular WLM policies provide improved performance for mixed workloads*

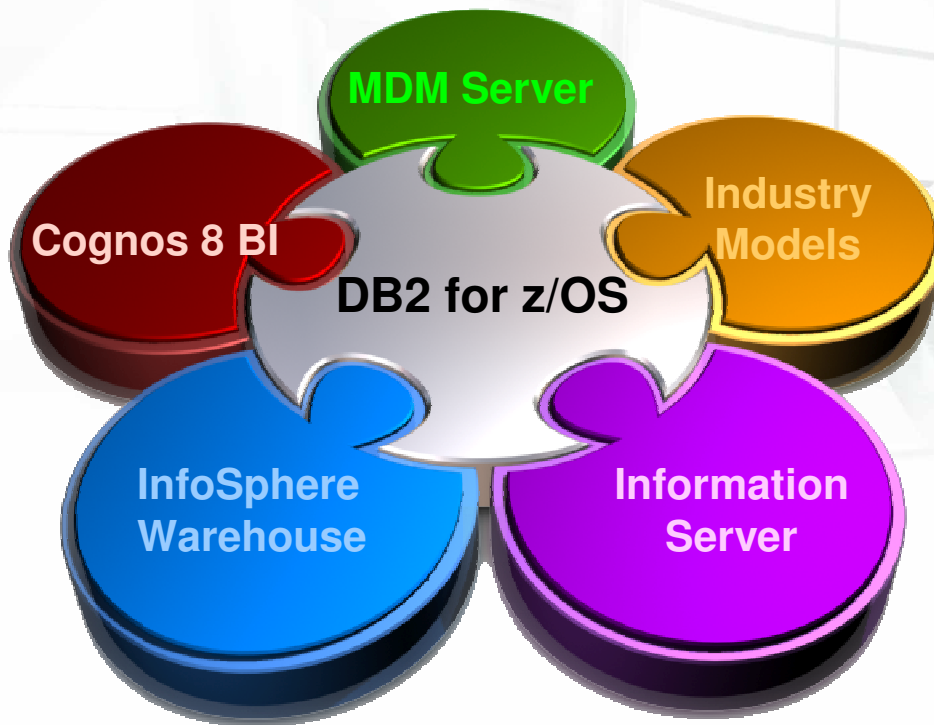


# Poughkeepsie 50TB Study (Redbook)

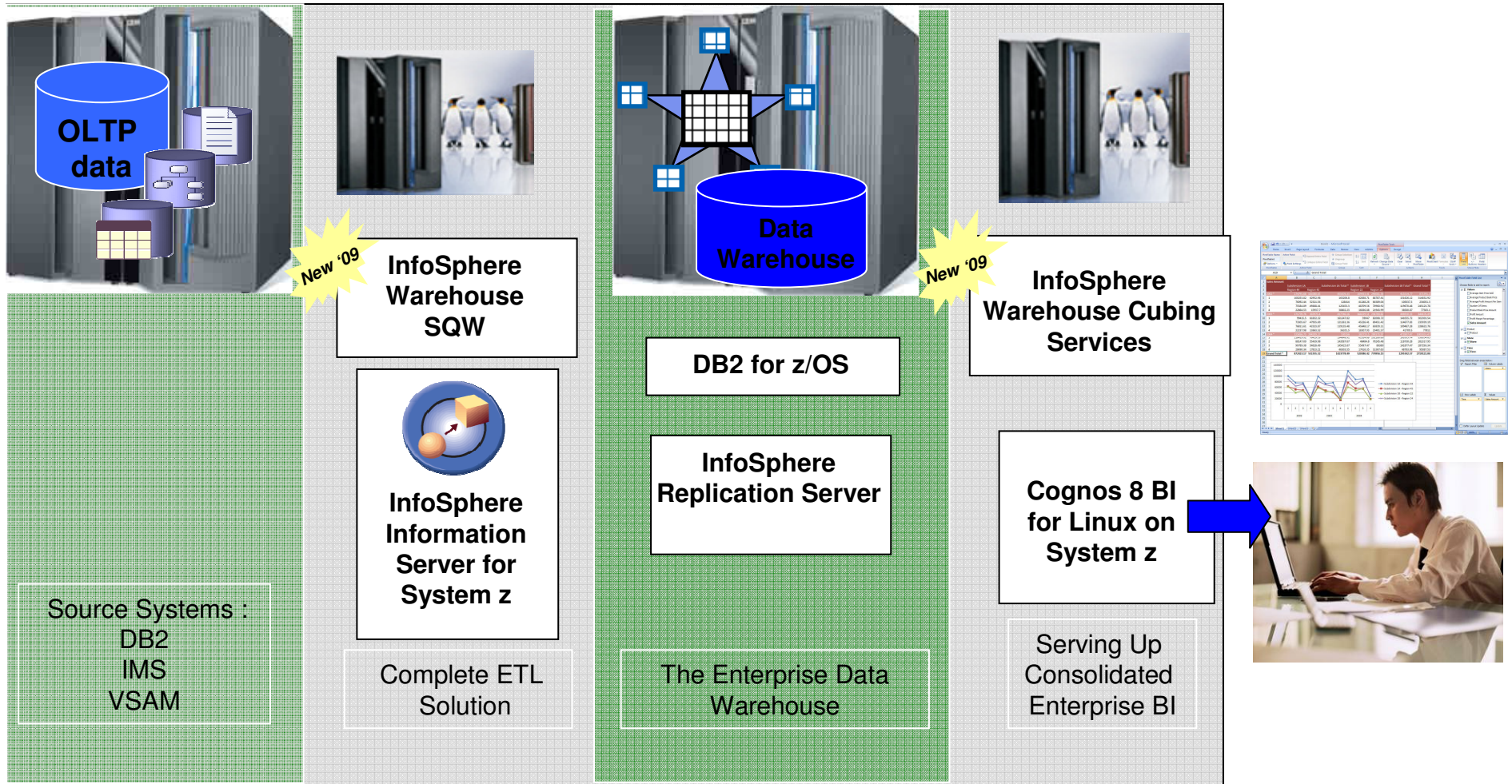
- **Deliver proof points of System z scalability in Business Intelligence environment**
- **Establish capability of System z to scale to larger volumes**
- **Develop best practices of managing large data warehouses**
- **Drive unique value of System z**
  - Workload Manager: capabilities to manage mixed workloads
  - Operational BI: large volumes of users, smaller queries
  - zIIP: lower cost BI solution
  - Data compression: minimal overhead with hardware compression



# *An Introduction to Data Warehouse on System z*



# The Data Warehouse and BI Solution on System z



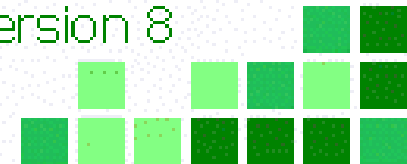
***Combining the Reliability and Availability of DB2 for z/OS with Cost Effective Applications running on Linux for System z***

## Warehouse/BI Highlights of DB2 for z/OS V8

- Materialized Query Tables
  - **Improved performance for query**
- Multi-row insert and fetch
  - **Improved speed of warehouse ETL and query**
- Online Schema evolution
  - **Improves availability and efficiency**
- Longer Table and Column names
  - **Increased compatibility with ETL and BI tools**
- Increased SQL vocabulary
  - **Increased compatibility with ETL and BI tools**
- 64 bit support
  - **Expanded size capability**

DB2 UDB for z/OS

Version 8



## Warehouse/BI Highlights of DB2 9 for z/OS

- Index over Expression
  - **Improved performance on ETL and BI tools**
- Index Compression
  - **Improved use of space (up to 50%)**
- Not Logged Tablespace
  - **More efficient for temporary tables such as Staging tables**
- Universal Table Space – partition by growth
  - **Easier to manage growth**
  - **Best of segmented tables and partitioning**
- Utility Improvements
  - **More online utility operations, reduced CPU**
- Overall DB2 9 for z/OS– reduced CPU





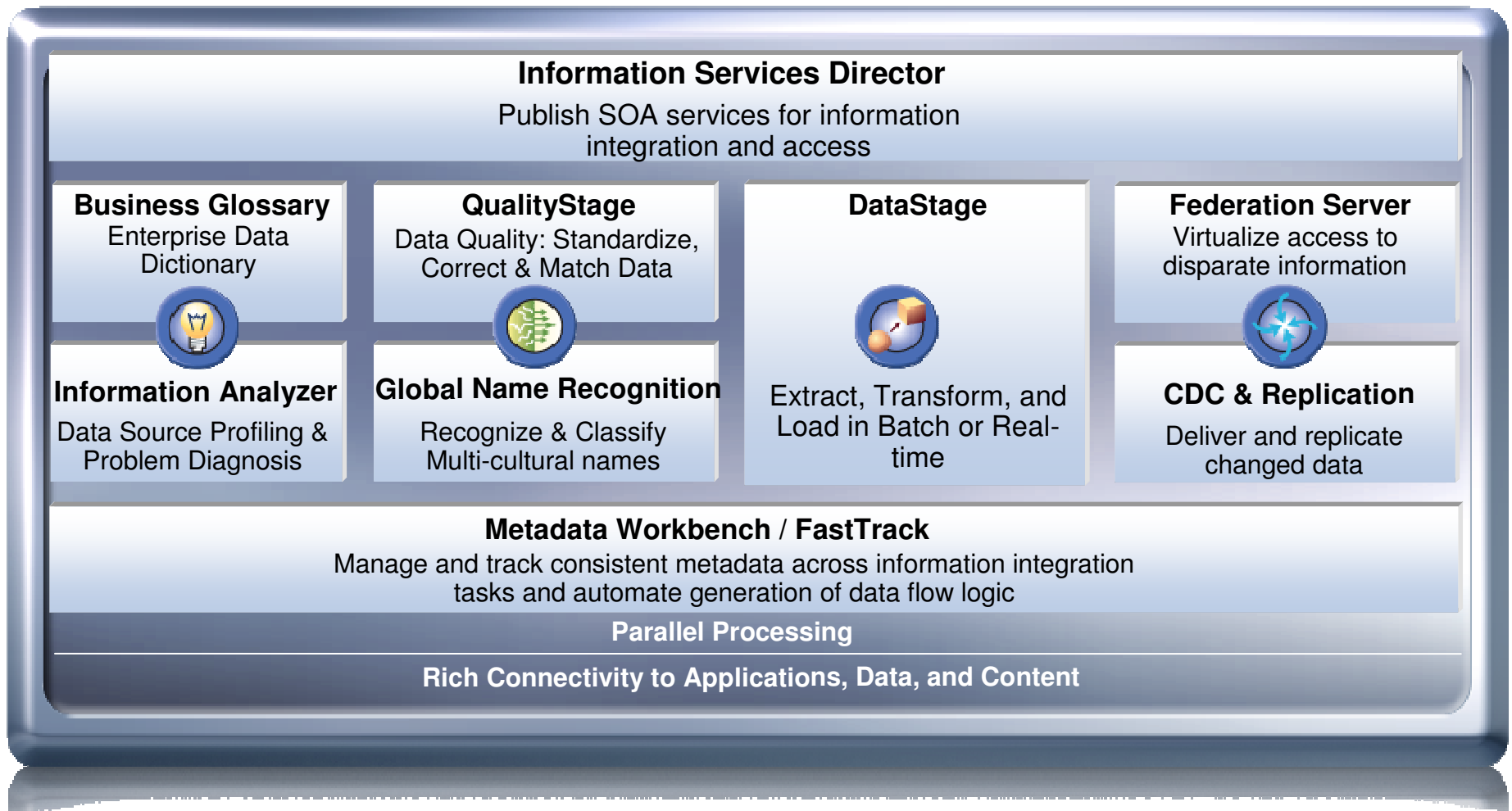
# DB2 X Query Enhancements

- **CPU time reductions for queries, batch, & transactions**
- **SQL enhancements: Moving Sum, Moving Average, ...**
- **pureXML improvements**
- **Access improvements: Index include columns, Hash**
- **Optimization techniques & REOPT(AUTO)**
  - Access paths more parallel, increased zIIP, reducing restrictions
  - In-memory techniques for faster query performance
  - Access Path Stability
- **Analysis: instrumentation & Data Studio & Optim Tuner**
- **Advanced query acceleration techniques**
  - IBM Smart Analytics Optimizer

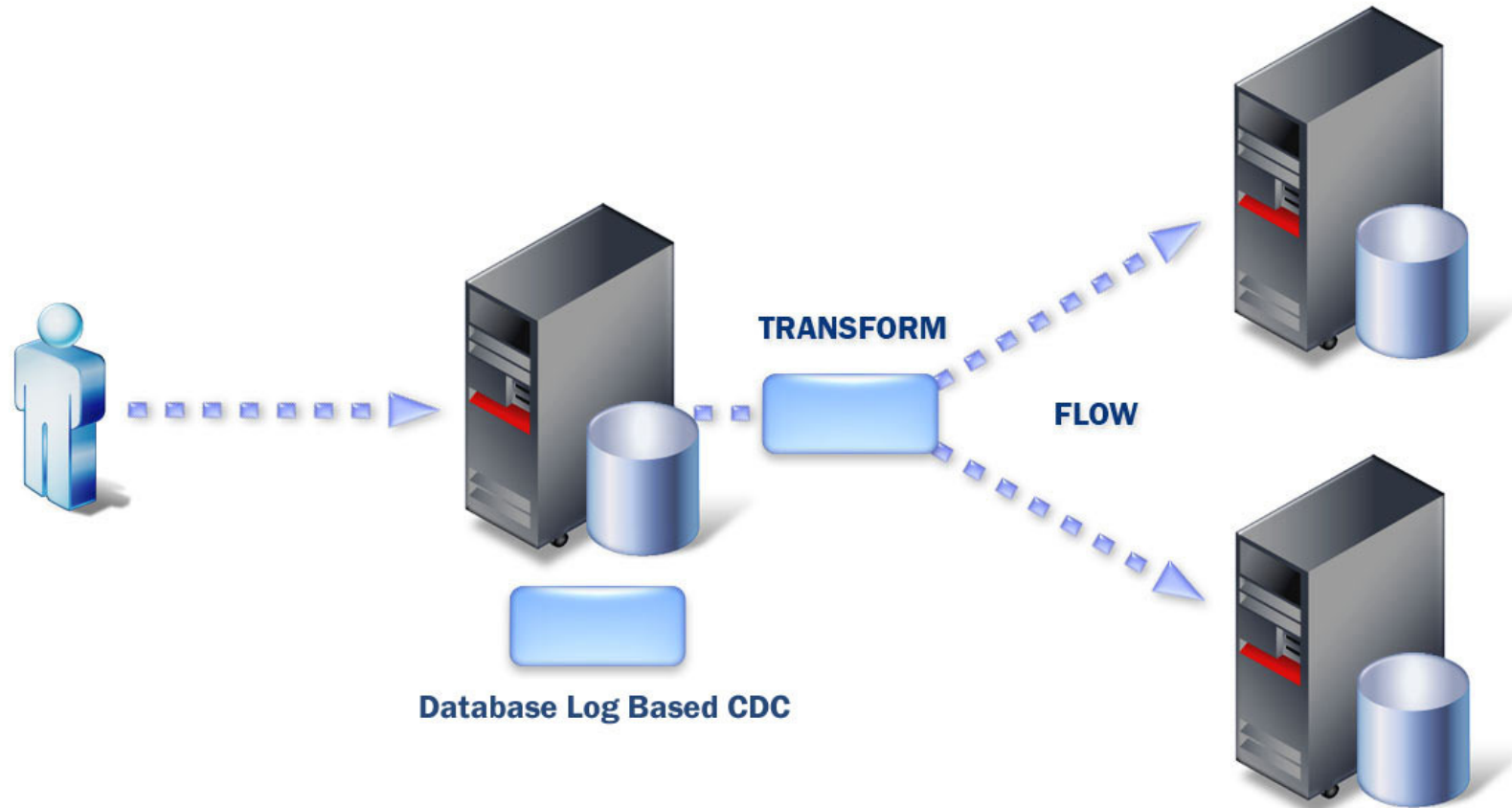
# DB2 X for z/OS At a Glance

<p>Application Enablement</p>	<ul style="list-style-type: none"> <li>• Versioned data or Temporal</li> <li>• pureXML enhancements</li> <li>• Last Committed reads</li> <li>• SQL improvements that simplify porting</li> </ul>
<p>RAS, Performance, Scalability, Security</p>	<ul style="list-style-type: none"> <li>• Wide range of performance improvements</li> <li>• Hash access to data</li> <li>• More online schema changes</li> <li>• Catalog restructure for improved concurrency</li> <li>• Row and column access control</li> <li>• Administrator privileges with finer granularity</li> </ul>
<p>Simplification, Reduced TCO</p>	<ul style="list-style-type: none"> <li>• 10 times more threads per DB2 image</li> <li>• Auto statistics</li> <li>• Data compression on the fly</li> <li>• Query stability enhancements</li> <li>• Reduced need for REORG</li> <li>• Utilities enhancements</li> </ul>
<p>Dynamic Warehousing</p>	<ul style="list-style-type: none"> <li>• Moving sum, moving average</li> <li>• Many query optimization improvements</li> <li>• Query parallelism improvements</li> <li>• Advanced query acceleration</li> </ul>

# InfoSphere Information Server on System z



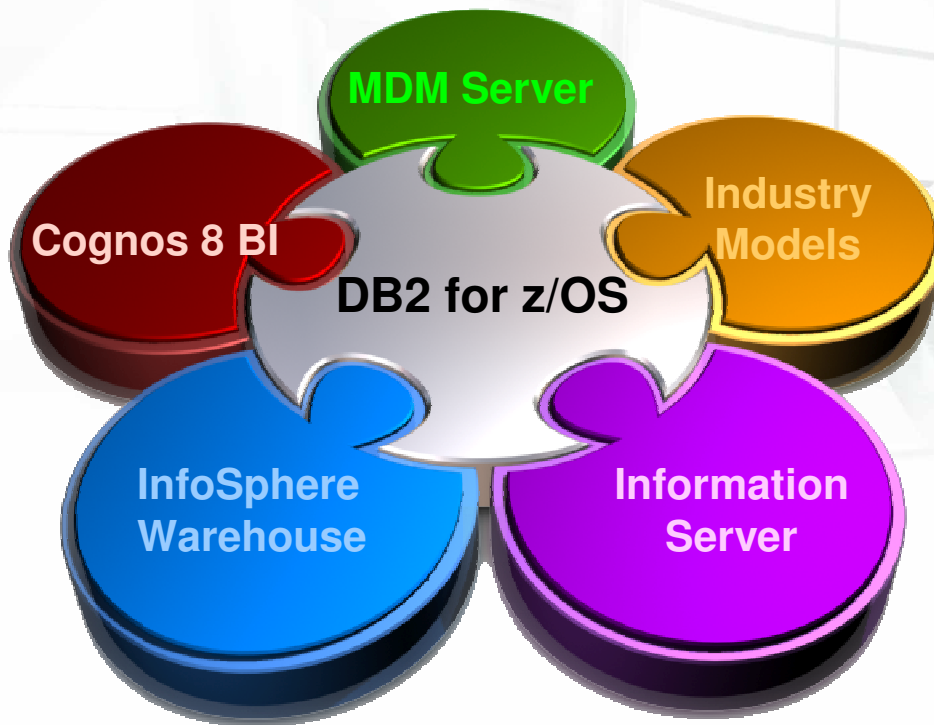
# Maintaining a Low Latency Warehouse or ODS



Replication, InfoSphere CDC, Event Publisher, Classic Event Publisher

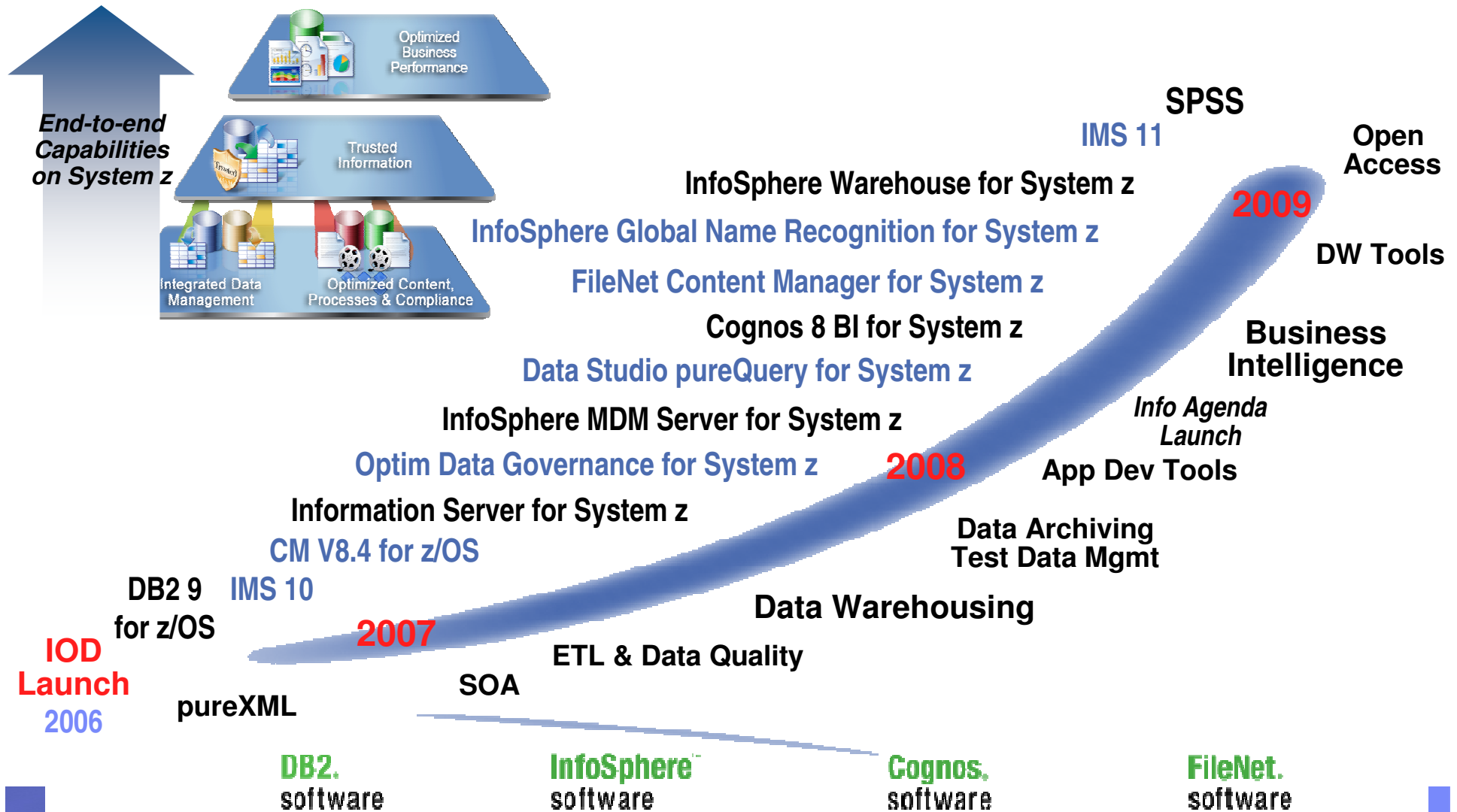
Provides incremental near real time feed to InfoSphere DataStage from IMS, VSAM, IDMS, Adabas, DB2 all platforms, Oracle, MS SQL Server, Sybase

# Business Intelligence and Data Analytics on System z

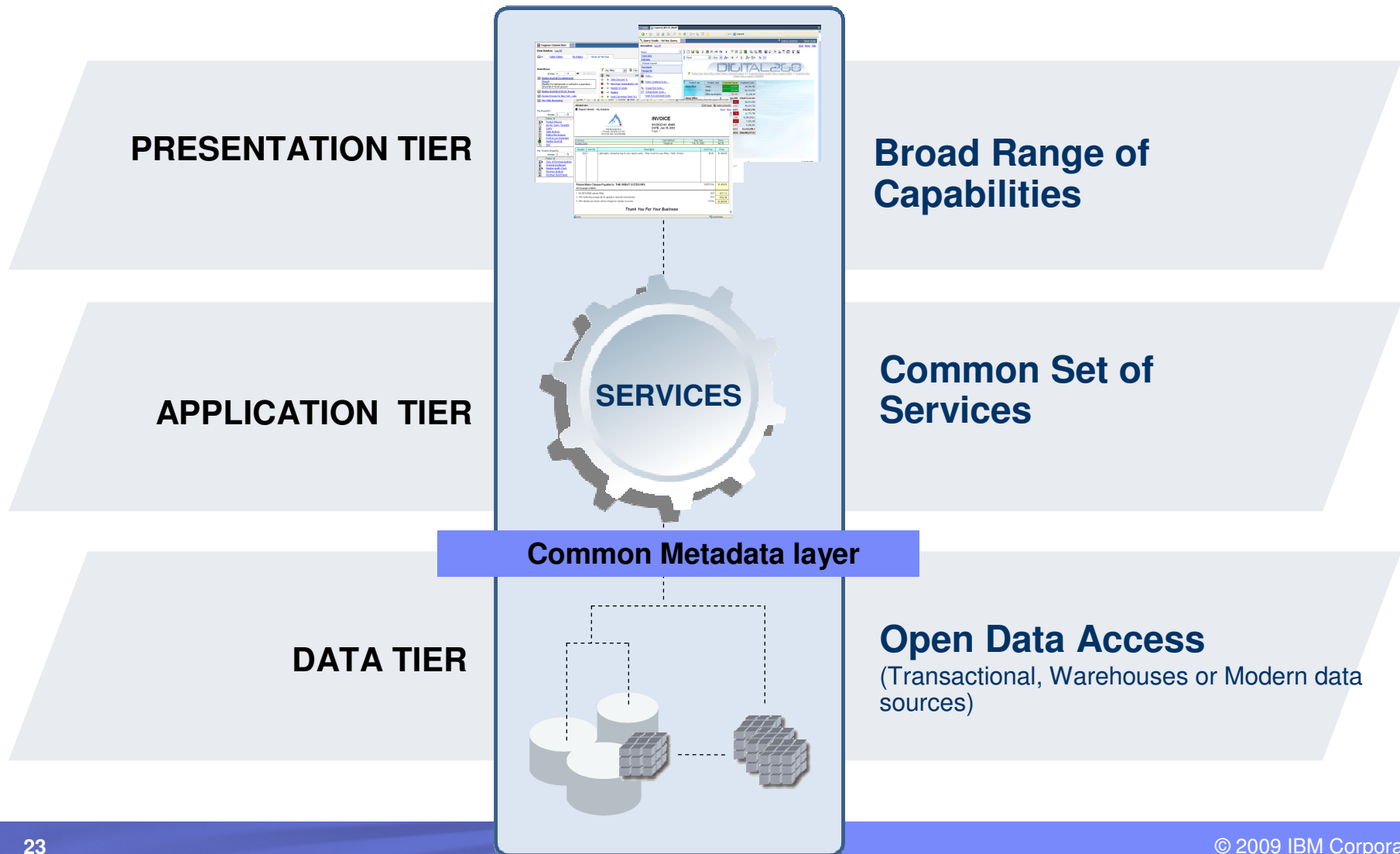


# Leveraging System z for Information on Demand

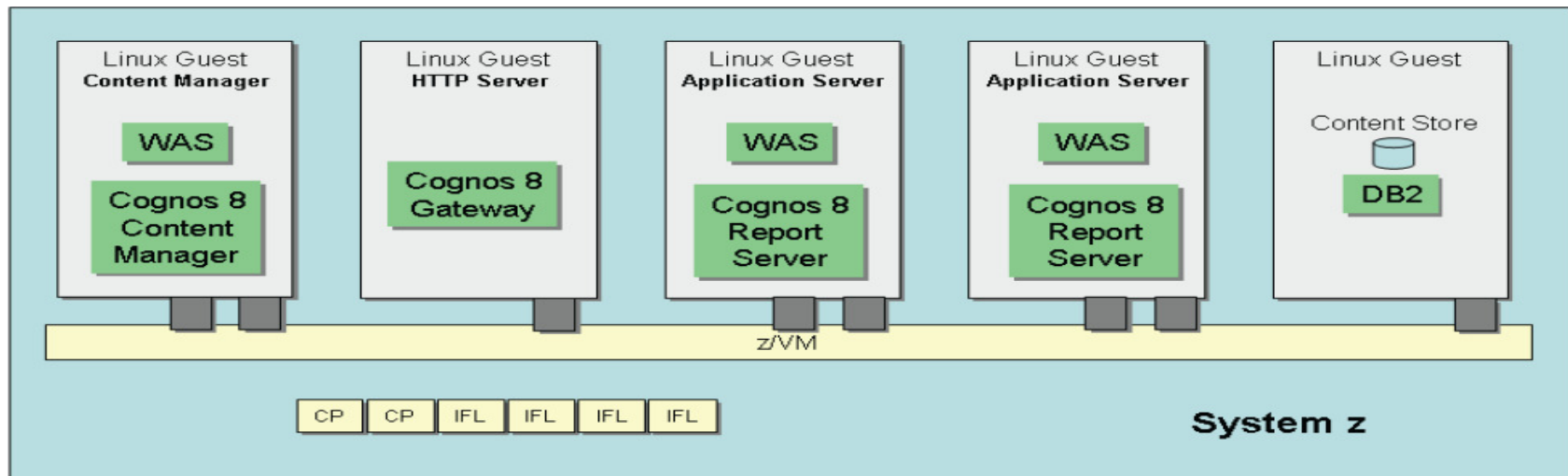
More new capabilities delivered in the past 3 years than at any point in history



# Cognos 8 BI for Linux on System z



# Cognos 8 BI – 10 TB Performance Study



## Linear scalability

**Use of z/VM Highly Recommended – slight performance gap from native, but greatly improved flexibility**

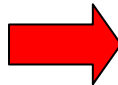


# OLAP – Online Analytical Processing

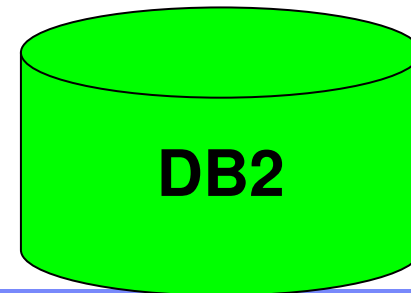
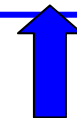
BI analyses most often involve aggregates and calculations such as ...

“I want to see all departments where the total salary is greater than \$xx and if they exceed the previous year by 15% then ...”

Let's start with a simple example ... just the sum of the salaries ...



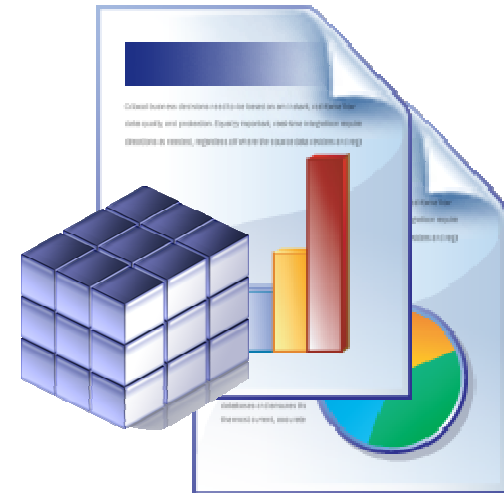
```
SELECT DEPT,(SUM) SALARY AS SUMSAL
FROM BIG_PERSONNEL_TABLE
GROUP BY DEPT
```



DB2 will do a vast scan of the table  
... form the aggregate values  
... and present the result set to the end user or application  
...and frequently many business users are looking at the same data over and over

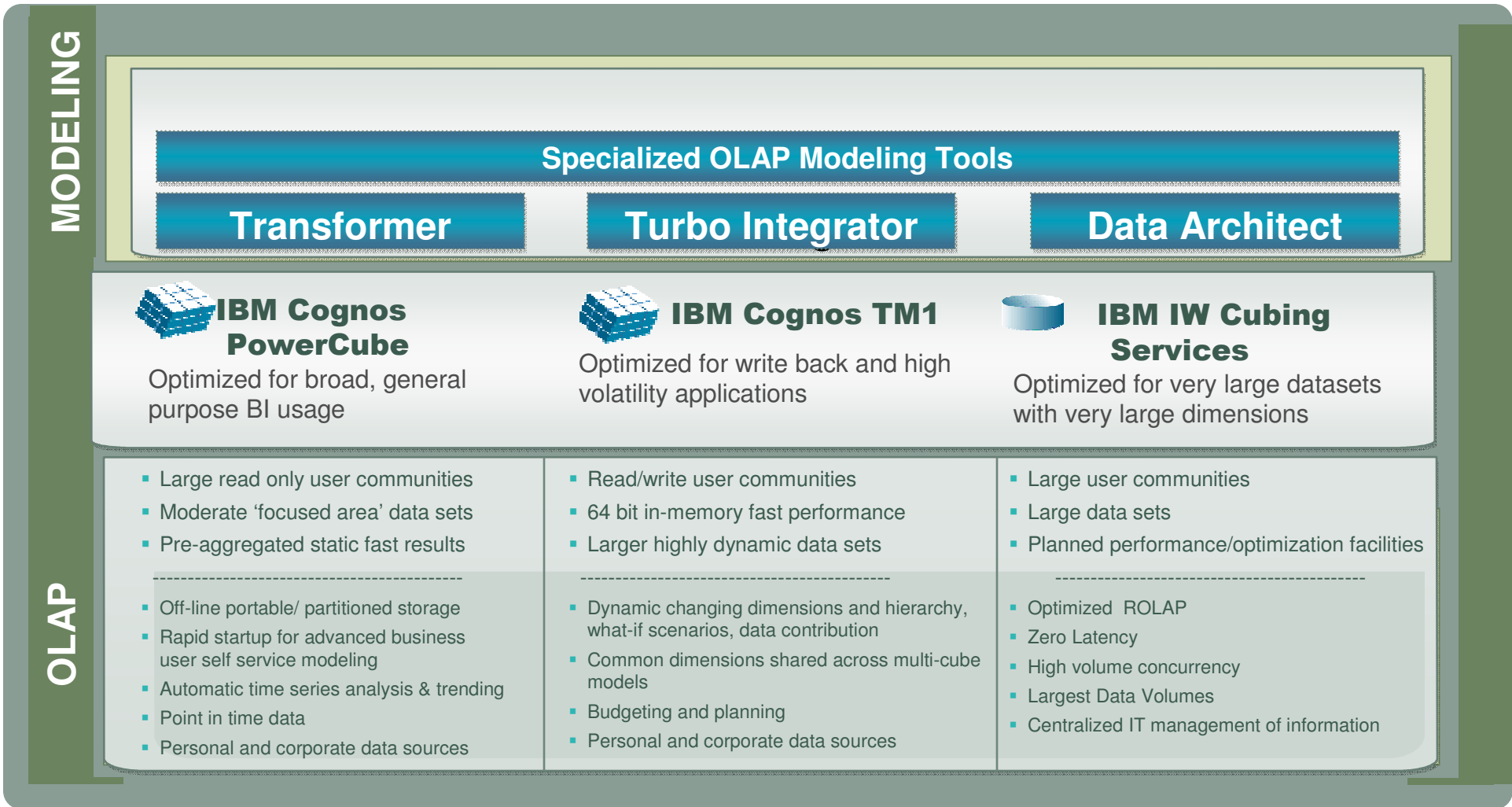
# Multidimensional Analytics

- **Multidimensional Analytics (aka OLAP)**
  - Multidimensional view of the business: customers, products, time and measures
  - Analysis of business metrics/measures over time, previous to “current”
  - Maps well to Star Schema concepts (Dimensions, Facts, Aggregates)
  
- **Key Aspects**
  - Dimensions, Hierarchies, Measures, Summarizations
  - Cross dimensional calculations
  - Time Series, Parallel Period analysis
  - Dimensional navigation – slice, dice, drill, pivot



# OLAP Portfolio from IBM

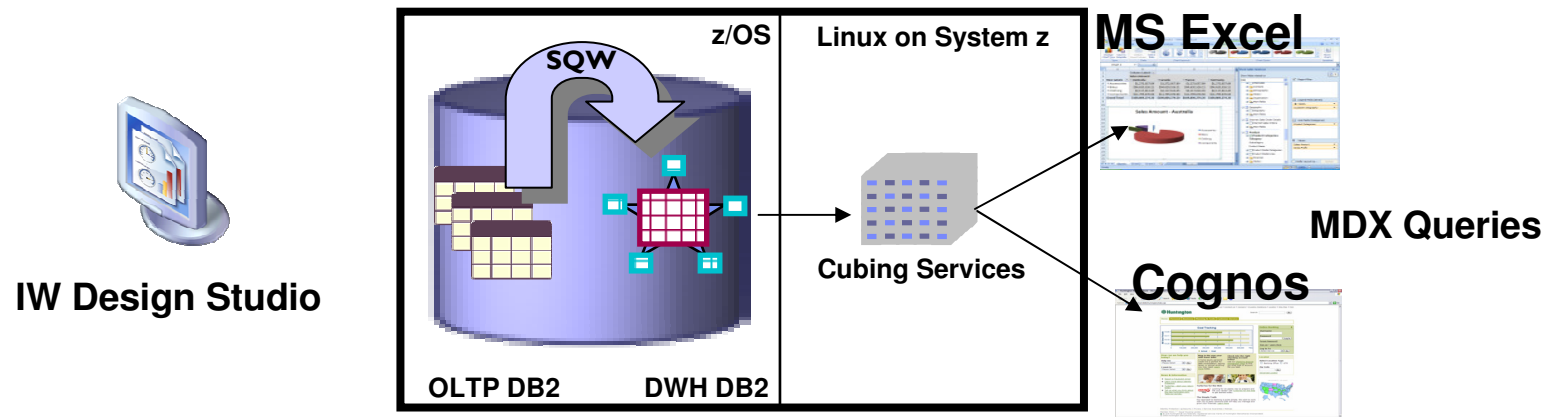
## *Different methods for different uses*



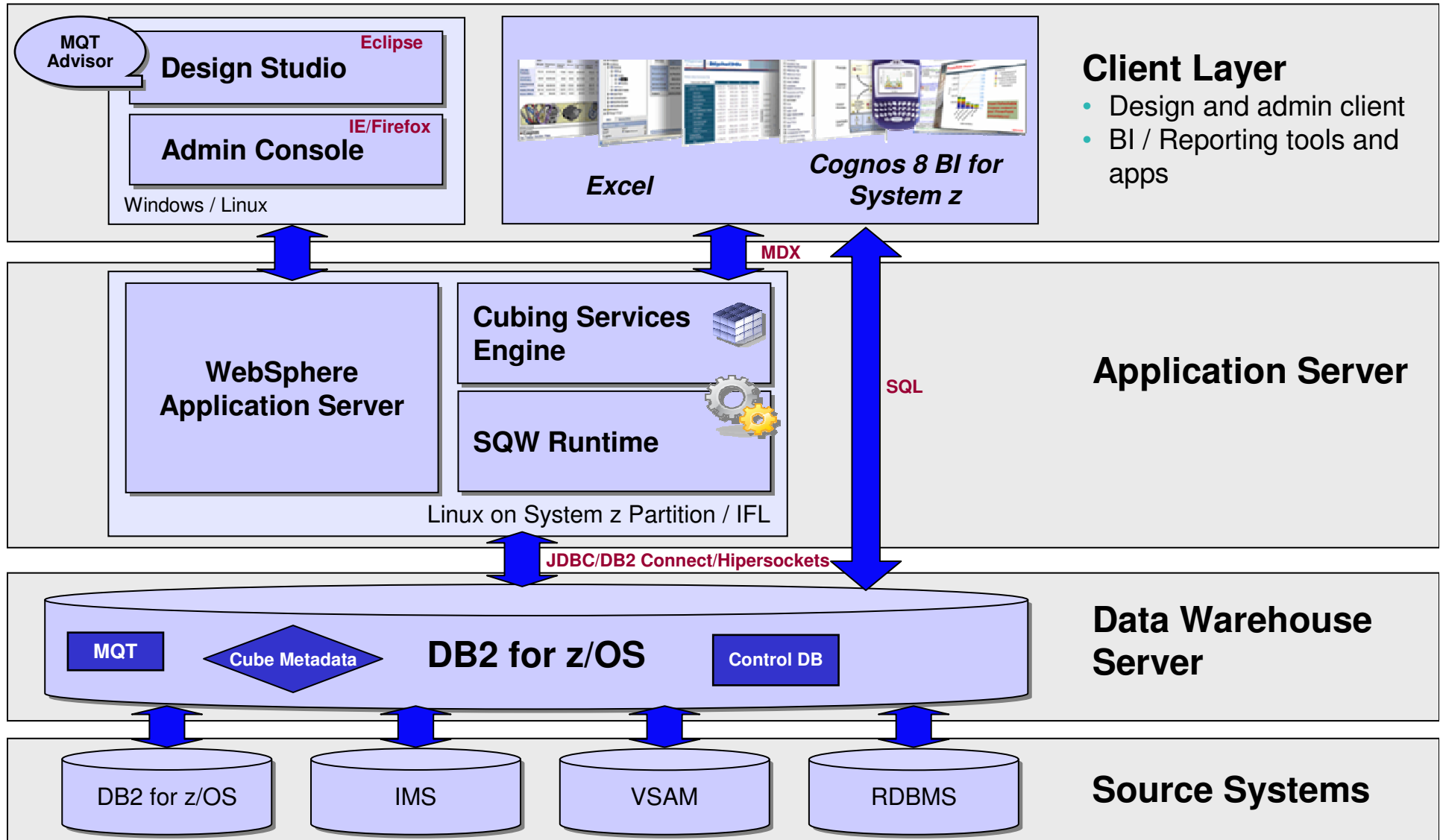
# InfoSphere Warehouse on System z

**Adds core data warehouse and analytics capability to DB2 for z/OS:**

- Advanced physical database **modeling and design**
- In-database data movement and manipulation capabilities of **SQL Warehouse Tool (SQW)**
- Optimize multidimensional reporting and analysis of data with **Cubing Services**



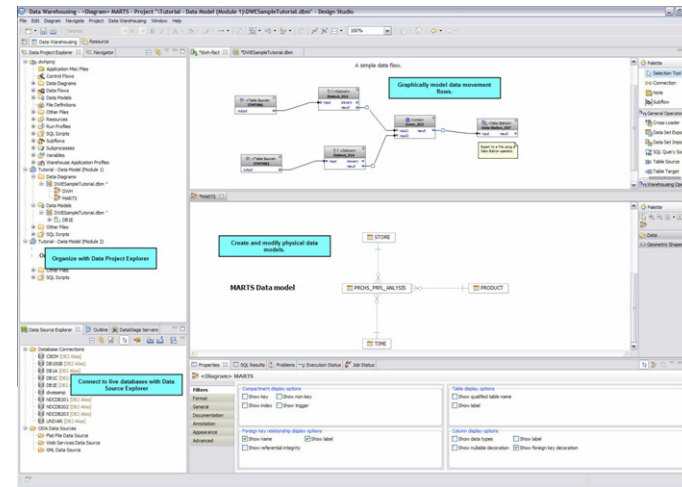
# InfoSphere Warehouse on System z - Architecture



# InfoSphere Warehouse on System z Tooling

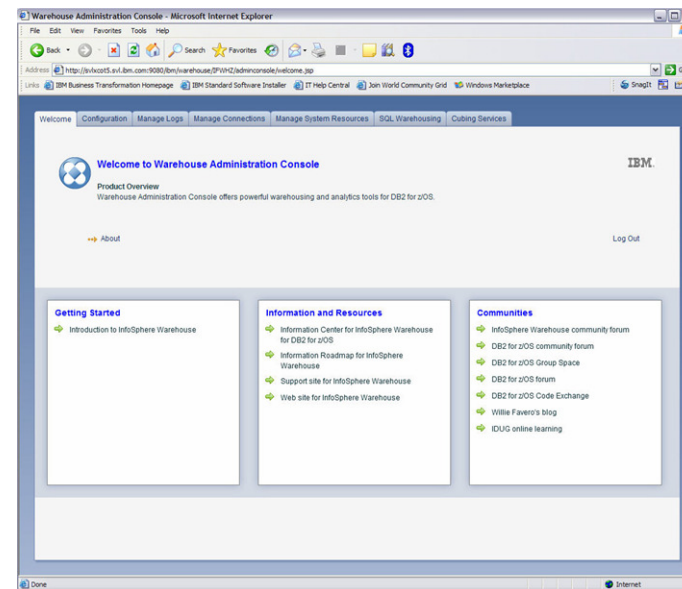
## – Design Studio – development IDE

- Develop physical data model
- Develop data movement flows
- Develop OLAP Model
- Prepare for deployment



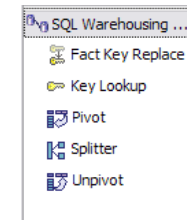
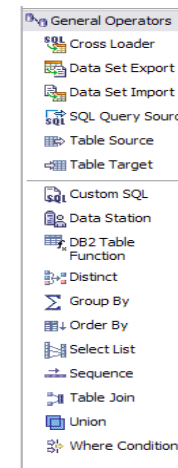
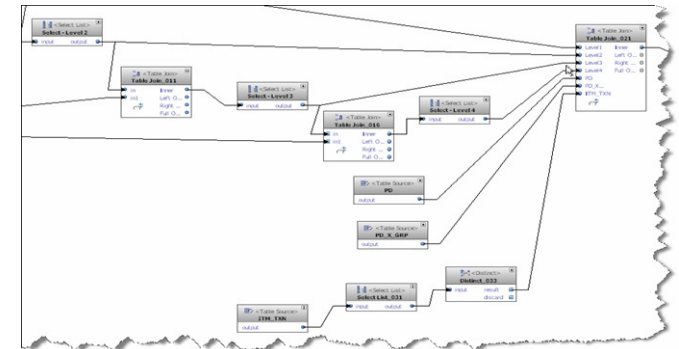
## – Administration Console – manage the runtime environment

- Deploy data movement applications
- Schedule, Execute, Monitor flows
- Define and manage cube servers
- Manage OLAP Metadata
- Assign cubes to cube servers

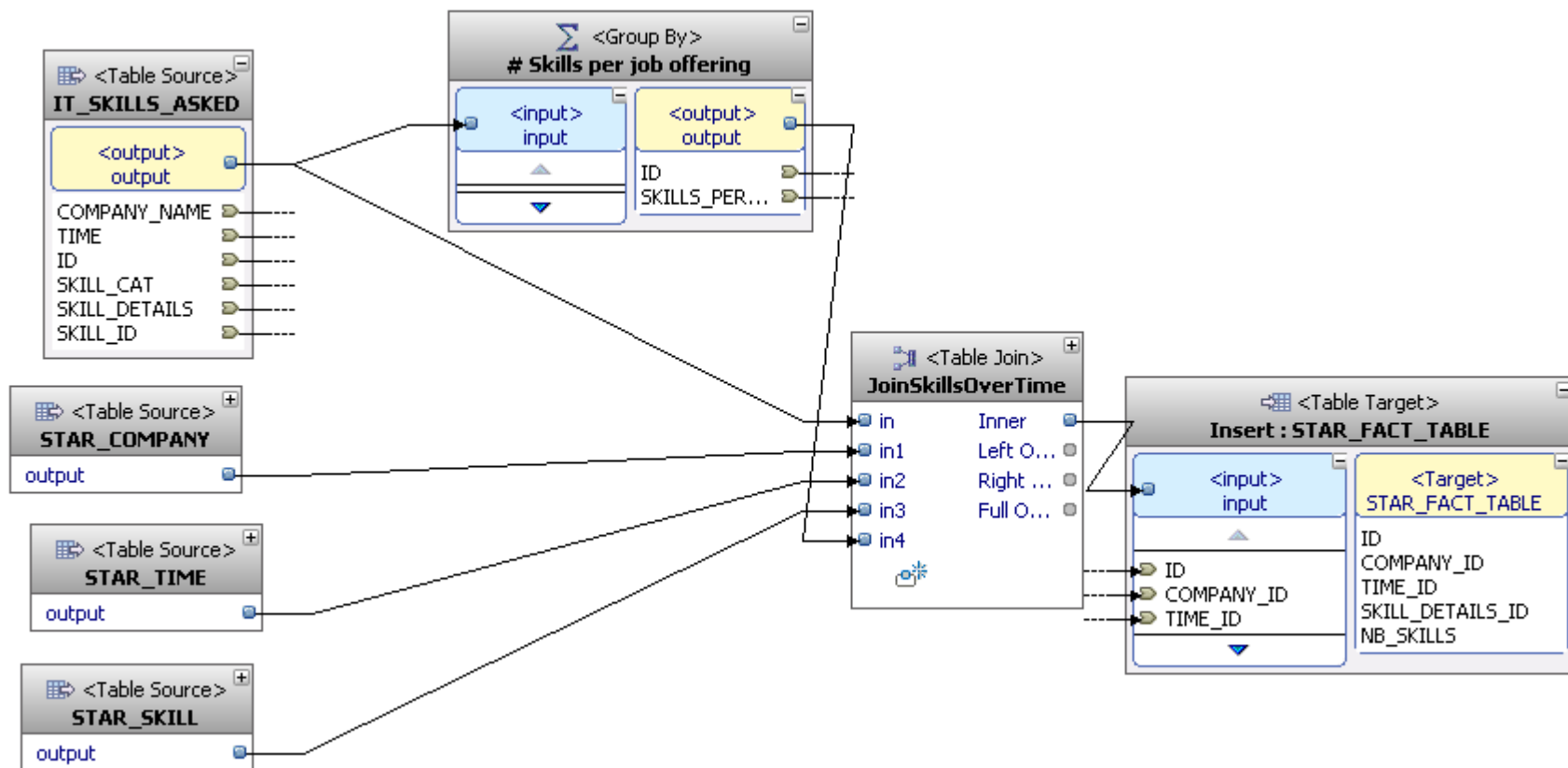
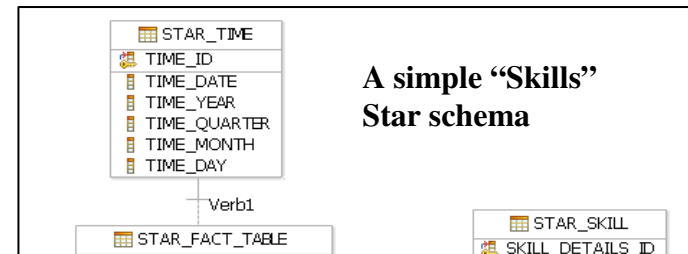


# Data Flows for in-database data movement and transformation

- Eclipse tool that generates SQL instructions, optimized for DB2, to move and transform data within the database
  - Ease of use, drag and drop UI for fast warehouse building
  - Data flow for doing transformations: 30+ SQL operators optimized for DB2
    - Leverages DB2 functions for transformations 100+
    - Customize with User-Defined Functions or custom SQL
  
- Perform warehouse building in homogeneous “database” environment
  - Leverage existing DBA skills, reuse custom SQL code
  - Native integration for fast BI application deployment
    - Database modeling and administration
    - Data Prep for in-database Report & OLAP
    - Analyze impact from tables to BI applications



# Describe your logic, graphically ....





... and have  
optimized SQL  
generated for  
you.

```

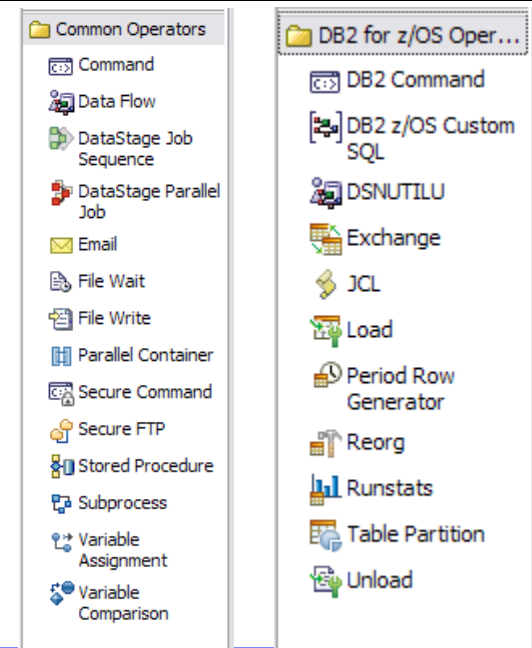
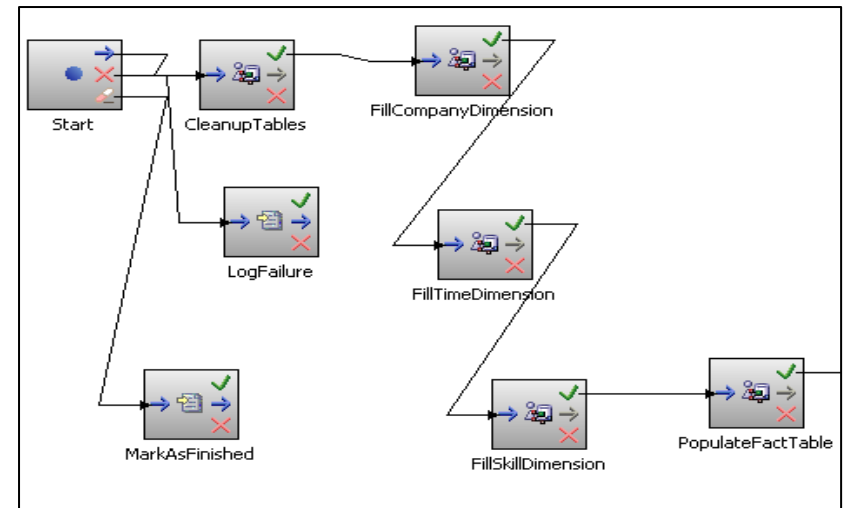
INSERT INTO OLAPANL.STAR_FACT_TABLE
(ID, COMPANY_ID, TIME_ID, SKILL_DETAILS_ID, NB_SKILLS)
WITH INPUT_04 (COMPANY_NAME, TIME, ID, SKILL_CAT, SKILL_DETAILS,
SKILL_ID)
AS (
SELECT
COMPANY_NAME AS COMPANY_NAME,
TIME AS TIME,
ID AS ID,
SKILL_CAT AS SKILL_CAT,
SKILL_DETAILS AS SKILL_DETAILS,
SKILL_ID AS SKILL_ID
FROM
  TXTANL.IT_SKILLS_ASKED INPUT_0281),
IN4_07 (ID, SKILLS_PER_OFFER)
AS (
SELECT
INPUT_04.ID AS ID,
COUNT(*) AS SKILLS_PER_OFFER
FROM
  INPUT_04
GROUP BY INPUT_04.ID)
SELECT
IN_07.ID AS ID,
IN1_07.COMPANY_ID AS COMPANY_ID,
IN2_07.TIME_ID AS TIME_ID,
IN3_07.SKILL_DETAILS_ID AS SKILL_DETAILS_ID,
DOUBLE(1) / DOUBLE(IN4_07.SKILLS_PER_OFFER) AS NB_SKILLS
FROM
  TXTANL.IT_SKILLS_ASKED IN_07,
  OLAPANL.STAR_COMPANY IN1_07,
  OLAPANL.STAR_TIME IN2_07,
  OLAPANL.STAR_SKILL IN3_07,
  IN4_07
WHERE (IN_07.SKILL_DETAILS = IN3_07.SKILL_DETAILS AND
IN_07.COMPANY_NAME = IN1_07.COMPANY_NAME AND
IN_07.TIME = IN2_07.TIME_DATE AND
IN_07.ID = IN4_07.ID)

```

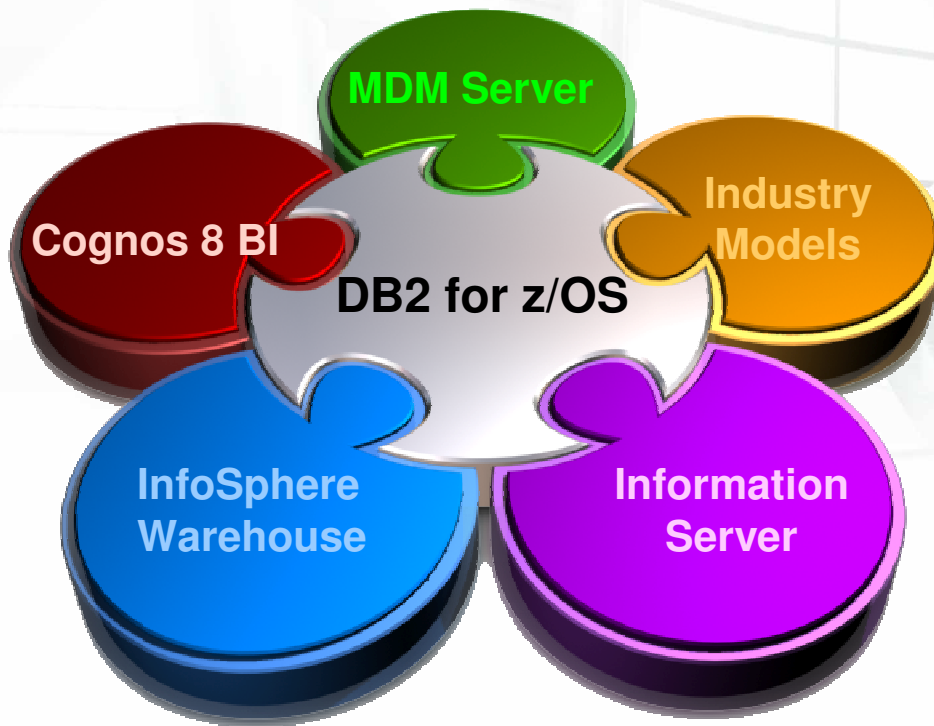
# Control Flows: Control and Sequencing

## Key Features:

- Control flows to coordinate the execution of several data flows and other activities
- Support for execution conditions: On Success, on Failure, always, Variable Comparison
- Support for non-database activities like:
  - Data flows
  - Datastage jobs
  - Executables, FTP
  - SQL Scripts, JCL jobs, Runstats, Reorg, Unload
  - E-mail, Logging
  - Iterators
  - Stored Procedures
- Execute activities in Parallel
- Modularize using Subprocesses
- Generate deployable packages (Warehouse Applications)



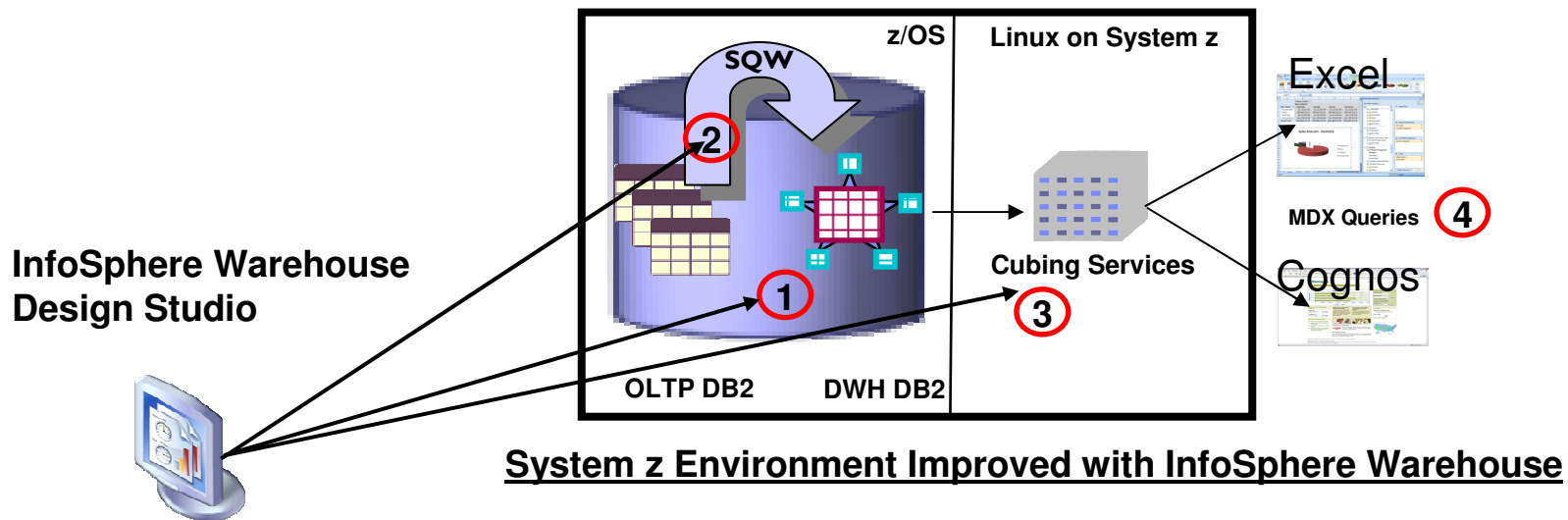
# Best Practices and User Experiences



# Modernize an Existing Reporting Server

## *Transform it into a DB2 for z/OS Warehouse*

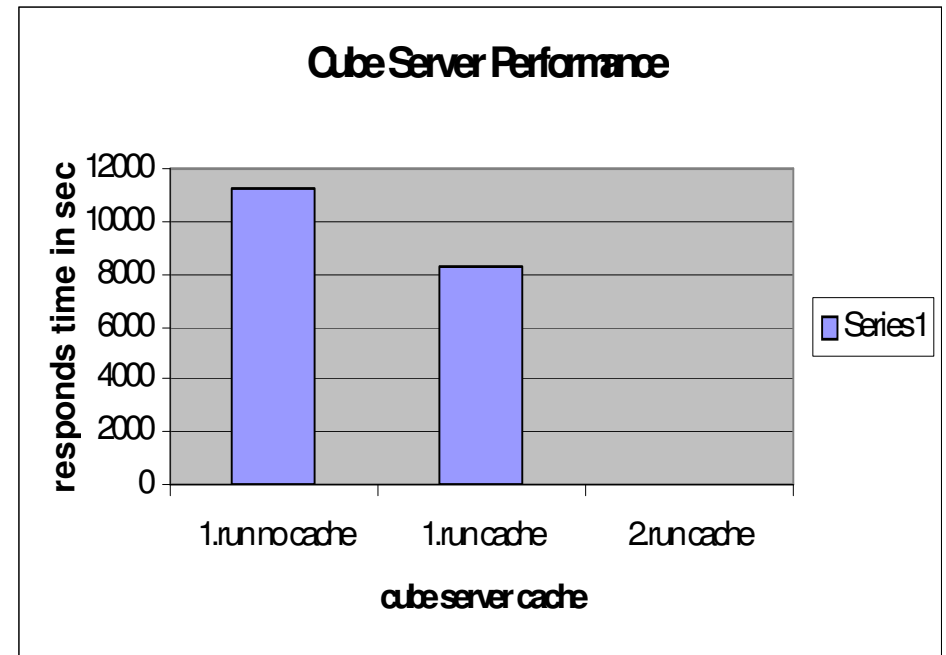
1. **Model your data for OLAP access using the InfoSphere Warehouse Design Studio**
2. **Populate your star schema using InfoSphere Warehouse SQW Scripts**
3. **Build one or more Cube structures using the InfoSphere Warehouse Design Studio**
4. **Access data through Excel or Cognos MDX queries**



# Modernize an Existing Reporting Server

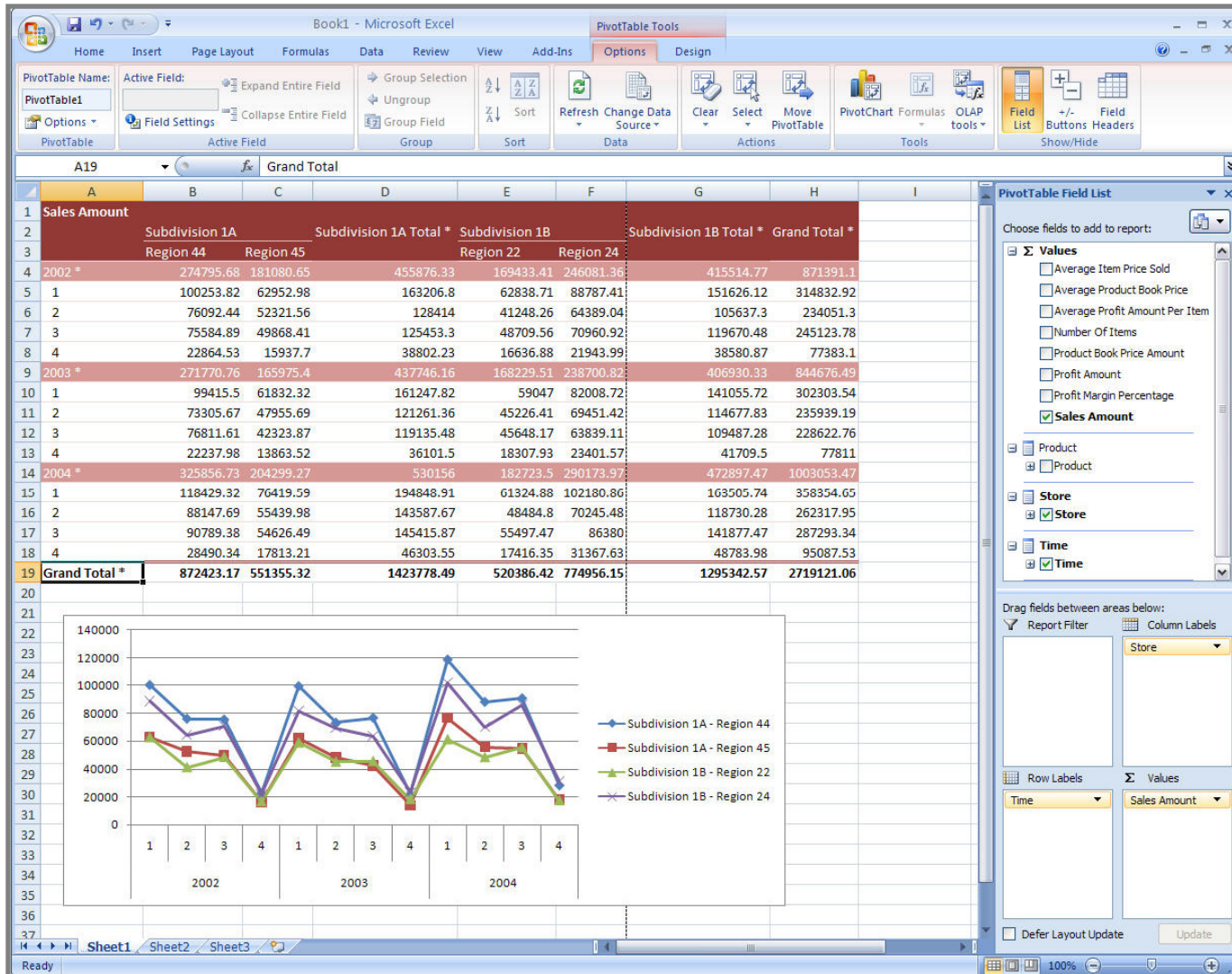
*Add Cubing Services to enhance business user experience*

- **Data can remain in one single repository – DB2 for z/OS**
  - **Increasing ease of data compliance, security and governance**
- **Analysis can be performed with high speed through in-memory caching**
- **CPU on System z can be reduced when more queries can be answered through the in-memory caching**



workload contains 30 MDX queries

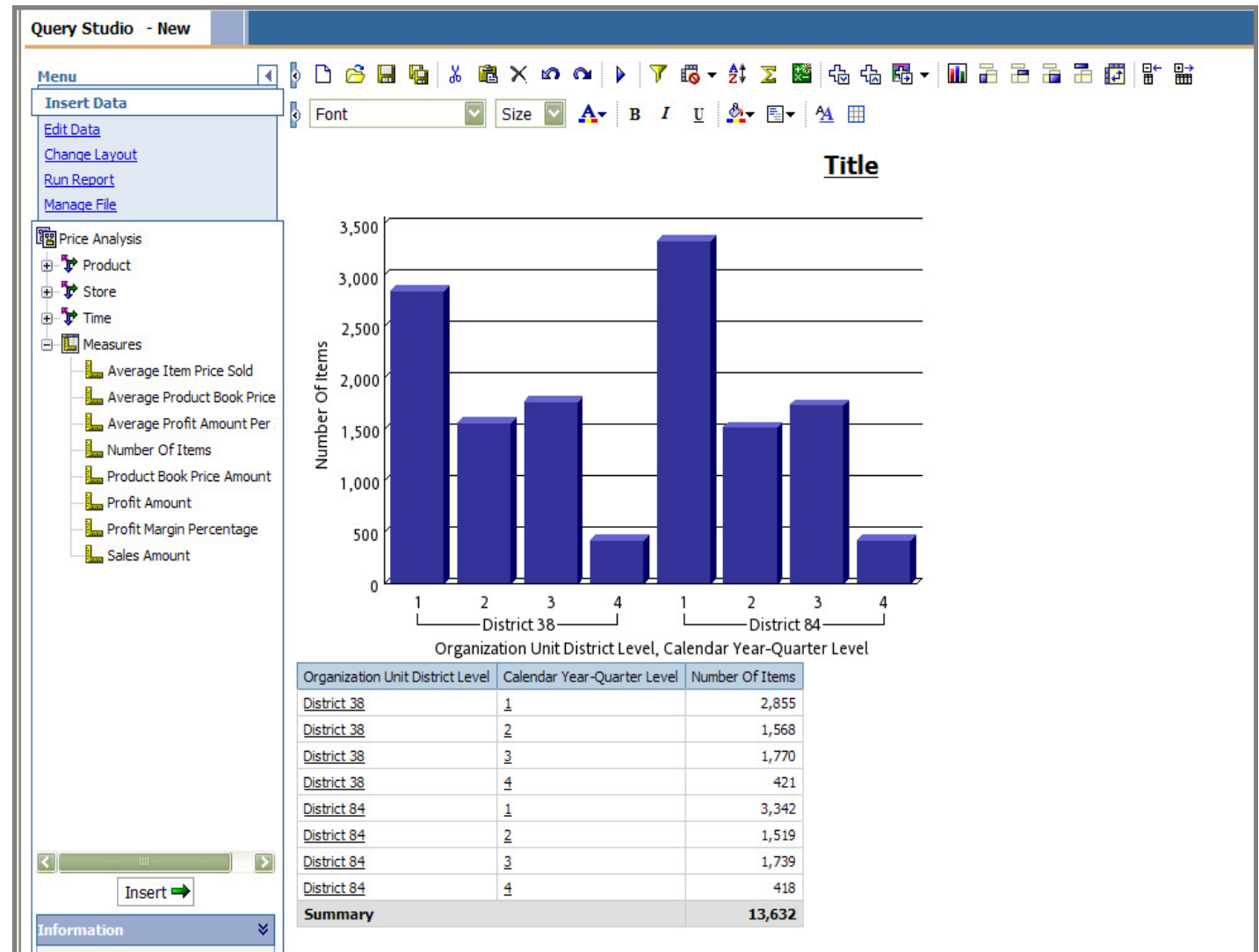
# Cubing Services Supports Microsoft Excel 2007



- **IBM OLE DB Provider for OLAP**
- **Uses Excel Pivot Table Services**

# Cubing Services Supports Cognos

- **All Cognos Studios can access Cubing Services**
  - Query Studio
  - Report Studio
  - Analysis Studio
- **Integrate cube data into Cognos Dashboards and Reports**
- **Cognos 8 BI uses metadata directly from Cubing Services – no duplicate metadata development**



# Consider Consolidation

- ***Application consolidation can provide some amazing efficiencies:***
  - ***Excellent scale-out***
  - ***Easier maintenance***
  - ***Dramatically improved upgrades***
    - ***Think days not months***
    - ***Think upgrade versus “rip and replace”***
  
- ***Data can be federated or consolidated, but again data consolidation on DB2 for z/OS can provide strong benefits:***
  - ***Security***
  - ***Availability***
  - ***Ease of data governance***
  - ***Data can be more current***

## ***IBM internal customer results:***

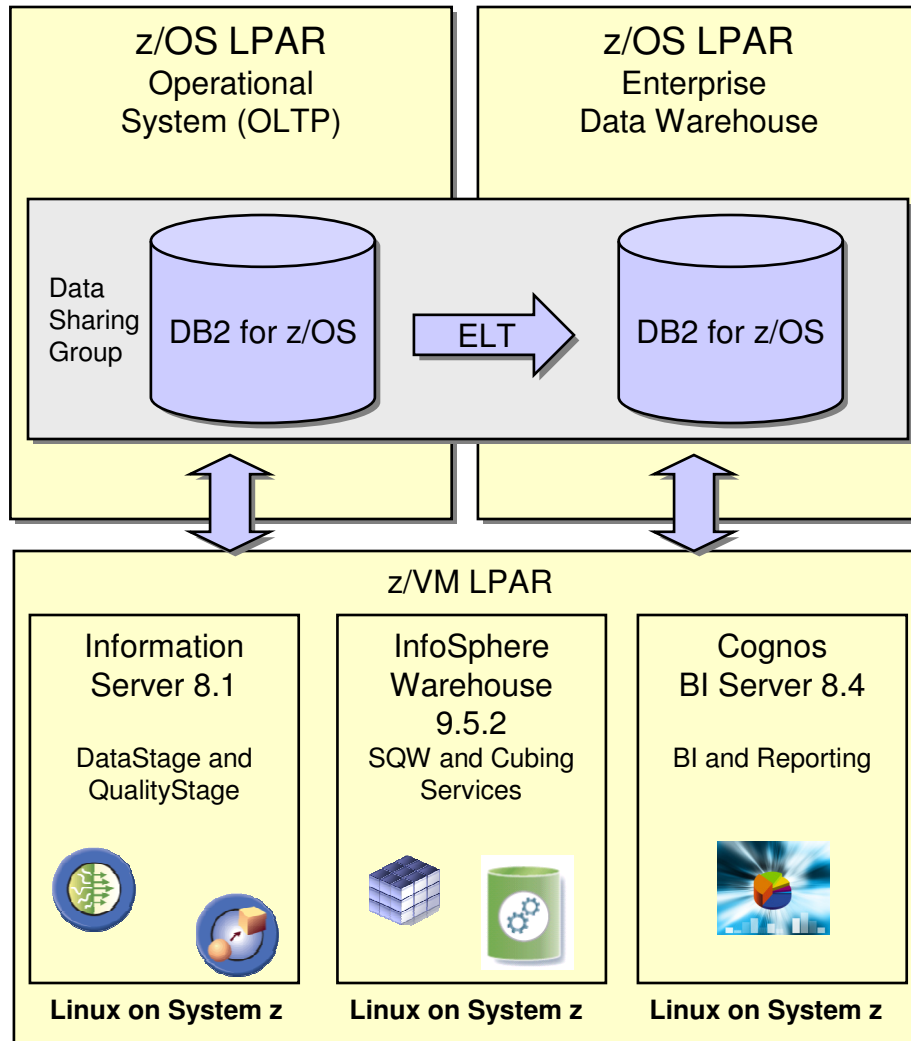
- Consolidating +20 multi-product, departmental BI deployments to Cognos 8 BI on System z
- Deploying a private cloud to support +200,000 named users across our global workforce
- Realizing value from +60 data sources across IBM



## Other Best Practices to Consider

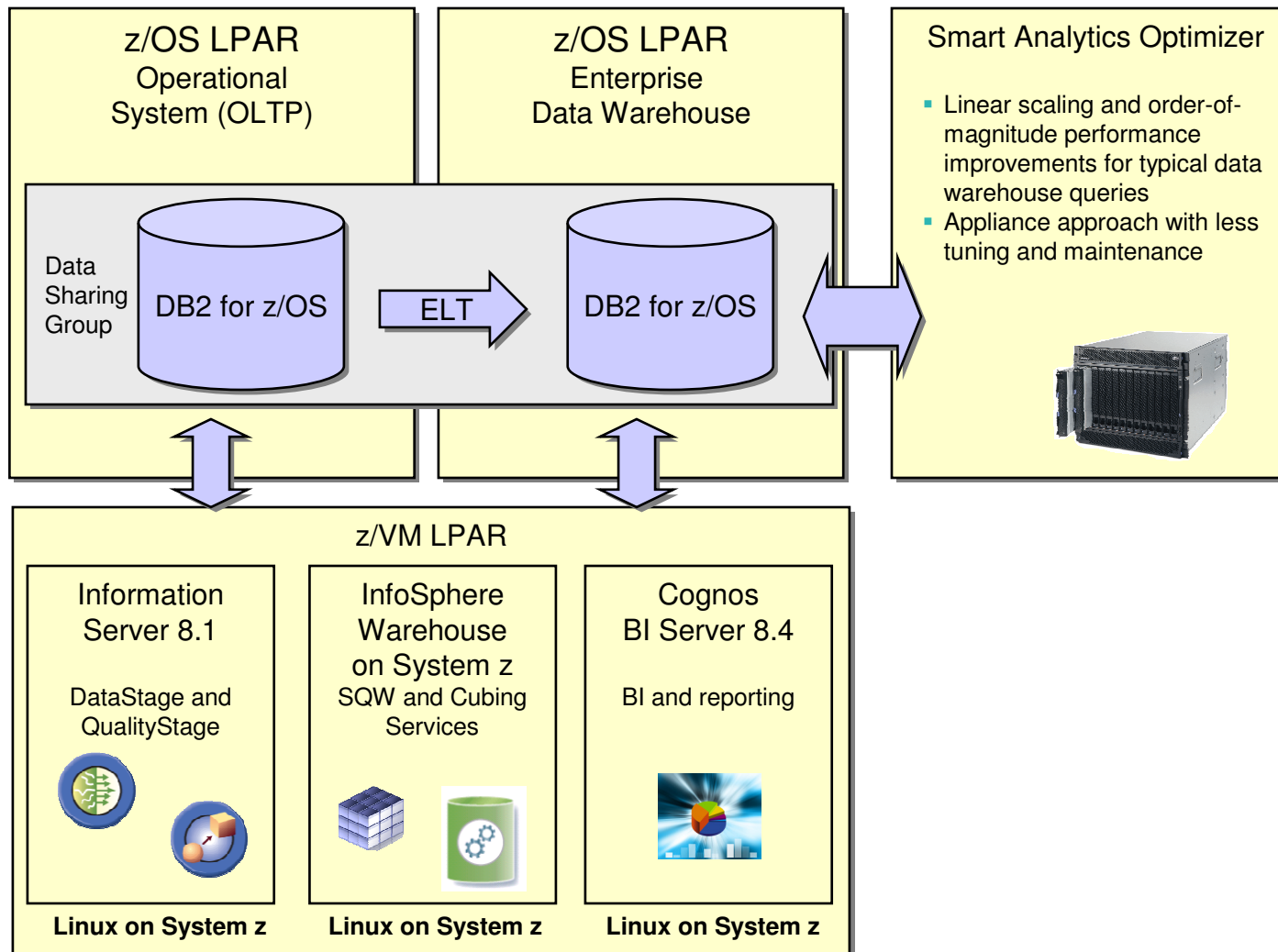
- **Think about how zIIPs can lower the cost of your DB2 for z/OS Data Warehouse implementation**
  - Parallel query
  - ETL utilities
  - DRDA - Business Intelligence queries
- **Design for mixed workload and implement granular workload manager policies that reflect the right balance of OLTP short running queries, longer running OLAP query workload, and power user interrupts**
- **Consider strategies for a trickle feed warehouse instead of full extract/load feeds**

# Ultimate Consolidation Opportunity



- Consolidation of mission-critical data on System z
- Leveraging existing environment, high availability, backup and governance procedures as well as skills
- Efficient data movement within a data sharing group (no network)
- Performance and TCO improvements through cubing services (data marts) and DB2 enhancements
- Complex transformations and data quality are driven from Linux on System z with Information Server

# Adding SAO to the System z Solution Stack



# Reference Materials

- **Relevant System z DW/BI Redbooks:**

- <http://www.redbooks.ibm.com/abstracts/sg247637.html>
- <http://www.redbooks.ibm.com/redpieces/abstracts/sg247674.html>
- <http://www.redbooks.ibm.com/redpieces/abstracts/sg247679.html>
- <http://www.redbooks.ibm.com/abstracts/redp4514.html>
- 3 more currently in progress: Cognos on System z, Data Collocation, InfoSphere Warehouse

- **Cognos 8 BI Performance Whitepaper:**

- <http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101437>

- **WLM Magazine Article:**

- <http://www.ibmssystemsmag.com/mainframe/novemberdecember09/tipstechniques/26908p1.aspx>

- **More info:**

- <http://www-01.ibm.com/software/data/businessintelligence/systemz/>

THANK YOU

# Disclaimer

© Copyright IBM Corporation [current year]. All rights reserved.

*U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.*

**THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM’S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.**

IBM, the IBM logo, ibm.com, DB2, InfoSphere, Cognos, and InfoSphere Warehouse on System z are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)

Other company, product, or service names may be trademarks or service marks of others.

Beth Hamel  
hameleb@us.ibm.com