Data Warehousing on System z:
SQL news for Application Developers & Meeting the need for real-time data for Business Analytics

Copenhagen, 15.03.2011

Patric Becker

Data Warehousing on System z CoE

Disclaimer

Agenda

Data Warehouse Challenges

Smart Analytics System 9600

Data Warehousing with DB2 9 and 10 for z/OS

Accelerating BI workloads with the Smart Analytics Optimizer

Additional thoughts and Summary

Agenda

Data Warehouse Challenges

Smart Analytics System 9600

Data Warehousing with DB2 9 and 10 for z/OS

Accelerating BI workloads with the Smart Analytics Optimizer

Additional thoughts and Summary

Market Dynamics are shifting

- **Business Analytics is now <u>mission critical</u>**

  – *Need to support broader users*

  – *Users are more intense with increasing data access demands*

  – *Requirements for high scalability, availability & performance*

- **Asked to do more with less (IT & Business)**

  – *Better access to relevant information*

  – *Need economies of scale*

  – *Consolidation with reduced complexity*

- **Corporate regulatory compliance**

  – *Driving intense scrutiny of data security policies*

- **Environmental concerns still top of mind**
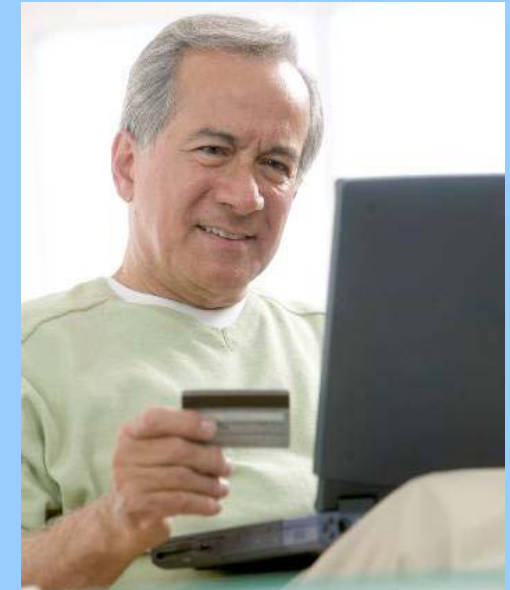
By 2014, externalizing BI will increasingly become an expected aspect of most companies' relationships with customers and partners.

Externally focused BI programs will frequently go beyond information dissemination by facilitating a collaborative decision-making process that breaks through the firewall to involve stakeholders from organizations in a broad ecosystem

Source : Gartner, Prepare for Customer-Facing Business Intelligence, Kurt Schlegel ,October 2010

What matters is changing

Results of New Intelligence Enterprise Survey of nearly 3,000 executives

Historic trend analysis and forecasting

Standardized reporting

Data visualization

Analytics applied within Business processes

Simulations and Scenario development

Clustering and segmentation

Regression analysis, discrete choice modeling and mathematical optimization

Data visualization

Simulations and Scenario development

Analytics applied within Business processes

Regression analysis, discrete choice modeling and mathematical optimization

Historic trend analysis and forecasting

Clustering and segmentation

Standardized reporting

> Respondents were asked to identify the top three analytic techniques creating value for the organization and predict which three would be creating the most value in 24 months.

Source: MIT Sloan Management Review,10 Data Points: Information and Analytics at Work, N Kruschwitz and R Shockley, Fall 2010

All departments, all users, in all roles across the organization need access to business insights

Executives | Business Managers | Line of Business Manager | Business Users | Business Analyst | Financial Analyst

How are we doing?

Why?

What should we do next?

SCORECARDING | REPORTING | QUERY | ANALYSIS | CONTENT ANALYTICS | PREDICTIVE MODELING | PLANNING
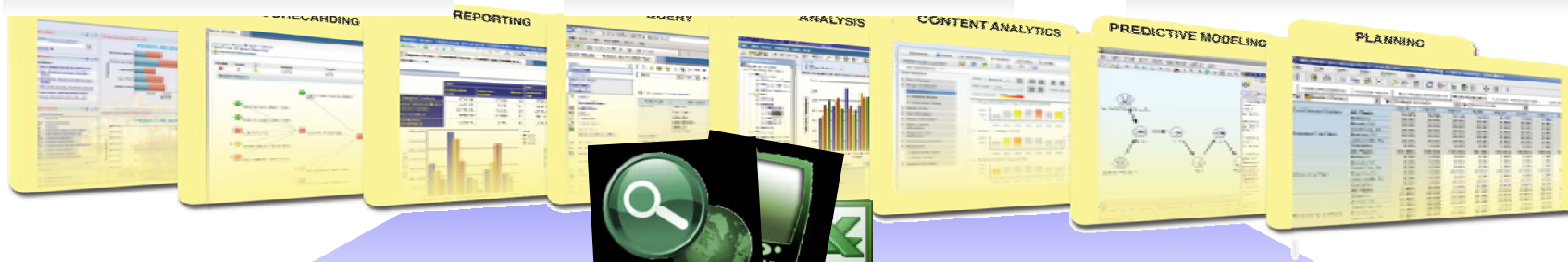
Real-time or historical; operational or strategic

Guided or self-service access and exploration…

Foresight using Statistical, and Predictive Analytics…

Common Business Model

Message Sources | Relational Sources | Application Sources | OLAP Sources | Modern and Legacy Sources

Information is at the Center…

44x

as much Data and Content Over Coming Decade

- Velocity
- Variety
- Volume

2020
35 zettabytes

2009
800,000 petabytes
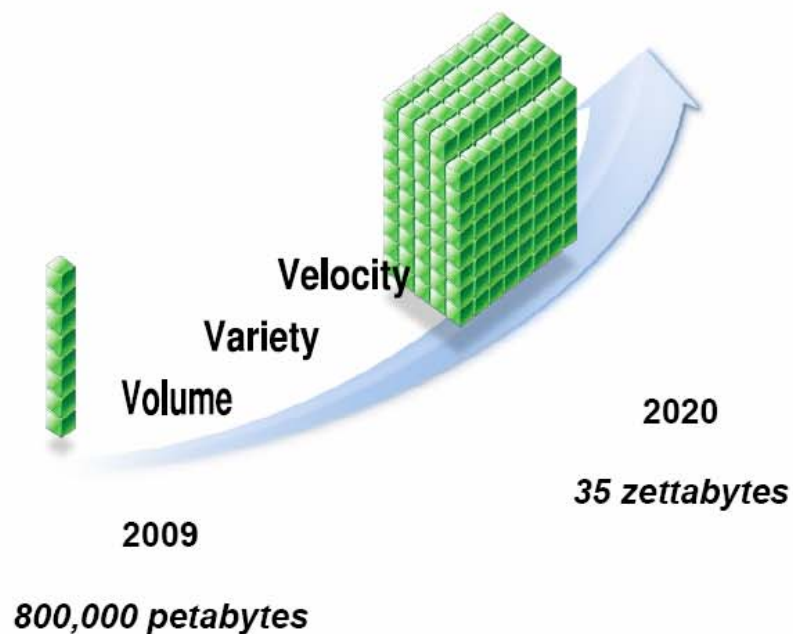
Sources:
- The Guardian, May 2010
- IBM Institute for Business Value, 2009
- IBM CIO Study 2010
- TDWI: Next Generation Data Warehouse Platforms Q4 2009

… And Organizations Need Deeper Insights

1 in 3 Business leaders frequently make decisions based on information they don't trust, or don't have

1 in 2 Business leaders say they don't have access to the information they need to do their jobs

83% of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

35% of businesses will look to replace their current warehouse with a pre-integrated Warehouse solution in the next 3 years; only 14% have today

© 2011 IBM Corporation

Today, many business users are not getting to the information they need, when they need it

60%+ of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions

27% of managers time is spend searching for information

50% of the information they obtain has no value to them

Key components for Business Analytics success are being implemented & managed in isolation

Business Analytics
(Business Intelligence & Predictive Analytics)

Data Warehousing
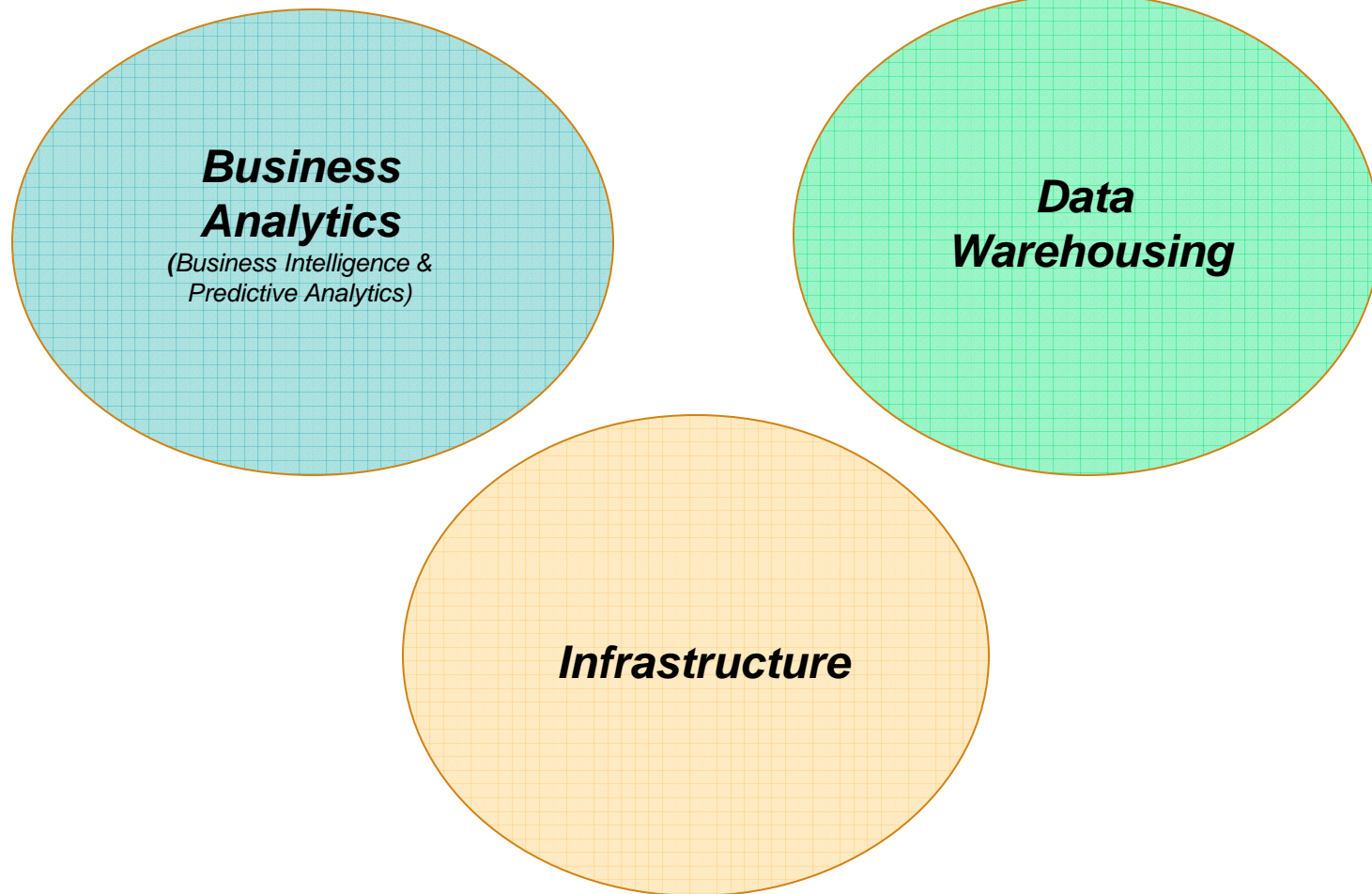
Infrastructure

Today's Traditional Infrastructure – *a siloed approach*

Division 'A'
Marketing & R&D Users

BA Solution

Data Warehouse / Data Mart

DW & BA Administrators & Report Authors

Division 'B'
Marketing & R&D Users

BA Solution

Data Warehouse / Data Mart

DW & BA Administrators & Report Authors

Executive Management

Operations

BA Solution

Data Warehouse / Data Mart

DW & BA Administrators & Report Authors

Sales

BA Solution

Data Warehouse / Data Mart

DW & BA Administrators & Report Authors

Finance

BA Solution

Data Warehouse / Data Mart

DW & BA Administrators & Report Authors

Data Warehouses have become isolated

- Information to drive the business is known but not available to the decision makers

- Information in the DW is limited to a small number of people in the organization

- Little to no interactivity with other systems

- Not built with the same criteria as the operational systems

- Difficult to manage and maintain multiple servers and copies of the data

- Minimal control over who is accessing the data

*"**Nearly 70%** of data warehouses experience performance-constrained issues of various types."*

- Gartner 2010 Magic Quadrant

Single version of the truth

- Single version of the truth is an essential element for Analytics and Business Intelligence

- Data redundancy elimination

- What deviation can be tolerated to make business decisions ?
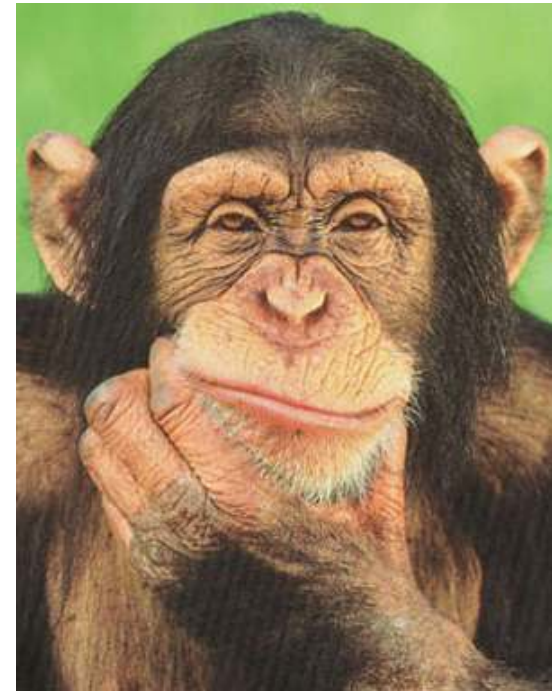
More than 98% human

Agenda

Data Warehouse Challenges

Smart Analytics System 9600

Data Warehousing with DB2 9 and 10 for z/OS

Accelerating BI workloads with the Smart Analytics Optimizer

Additional thoughts and Summary

Simplicity, Flexibility, Choice
IBM Data Warehouse & Analytics Solutions

Appliances

**Flexible
Integrated Systems**

**Custom Software,
Hardware & Services**

IBM Netezza

IBM Smart Analytics System

IBM InfoSphere Warehouse

Warehouse Accelerators

Information Management Entry Points
(Information Server, MDM, Streams, etc)

Simplicity *The right mix of simplicity and flexibility* **Flexibility**

IBM Smart Analytics System 9600
High Value Data Warehousing

Cognos 8 BI

System z

InfoSphere Warehouse

Cubing Services

Data Warehouse

Powercube Services

ELT

DB2 for z/OS VUE (option for MLC)

Implementation Services

DB2 Utilities Suite
Image Copy, LOAD, UNLOAD, REORG, etc

Operational Source Systems
Structured/ Unstructured Data
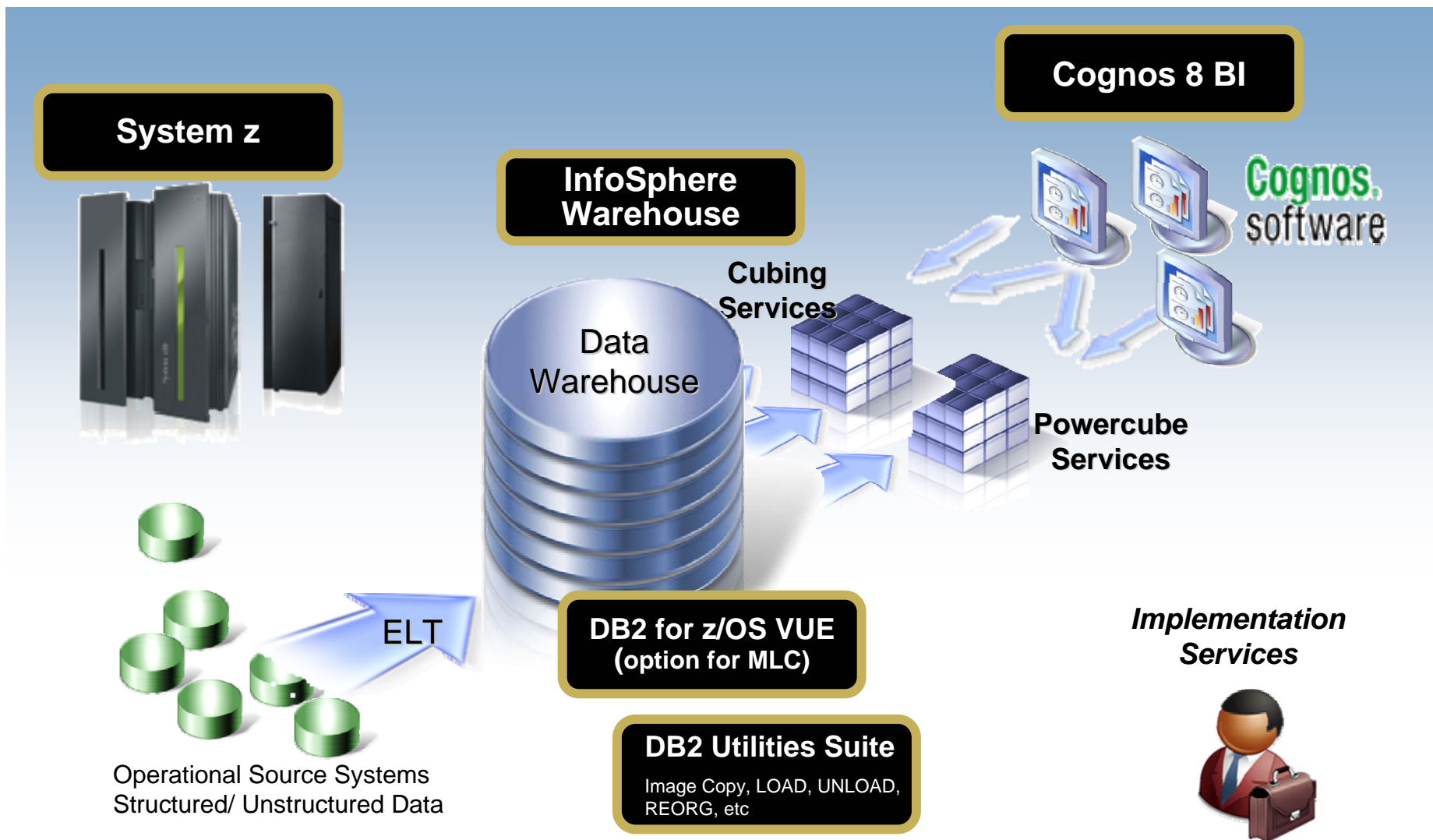
Smart Analytics System 9600
High Value, Dynamic Warehousing Solution…

…mix and match
user roles…

…in tested, ready to deploy
configurations at a Bottom
Line Price

✓ Hardware

✓ Software

✓ Services

Data Capacity Sizing

Amt Disk*	DB2 Base Amt User Data*
250TB	100TB
125TB	50TB
62TB	25TB
30TB	12TB
10TB	4TB

BI Capacity Sizing

	Cognos Workload	
	Named Users	Max Concurrent Users
	10,000	100
		50
	5	25

Choose from Each

Hardware: A single centralized BI environment

Building an end-to-end BI environment on System z

End-user functionality
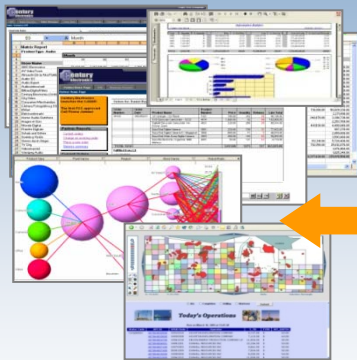
BI tooling in z-Linux LPAR:

- InfoSphere Warehouse
- Cognos
- Optional:
  - DataQuant/QMF
  - IBI WebFocus, SAS
  - Business Objects

z/OS foundation LPAR for enterprise BI:
- **DB2 for z/OS** – highly available, secure, query prioritization, concurrent utilities/queries
  - **z/OS:** Highly secure, Efficient, Trusted, Responsive

IBM Smart Analytics System 9600 Software

Deeply Optimized by IBM Experts… Flexible Growth…
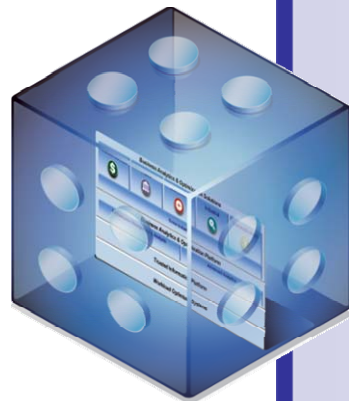
Powerful Data Warehouse and BI Software
- ❑ DB2 for z/OS Value Unit Edition (primary) V9
  - *Option for MLC*
- ❑ DB2 Utilities Suite V9
- ❑ InfoSphere Warehouse on System z V9.5.2
- ❑ Cognos 8 BI for Linux on System z V8.4
- ❑ z/OS Operating System Stack V11

Optional Value Priced Add-ons
- ❑ Tivoli OMEGAMON for DB2 Performance Expert
- ❑ DB2 Connect
- ❑ Tivoli Directory Server
- ❑ InfoSphere Information Server
- ❑ InfoSphere Replication Server
  - ❑ Q-Rep, CDC and Event Publisher eligible
- ❑ InfoSphere Federation Server plus Classic Federation on System z
- ❑ SPSS
- ❑ Tivoli ITCAM, ITUAM
- ❑ Cognos Now! For Linux on System z
- ❑ Cognos Blueprints for Healthcare, Banking and others…
- ❑ BI User on-boarding application (as proposed for Smart Analytics Cloud)

A Broad Range of Analysis, Dashboards & Reporting

Pre-optimized Business Intelligence Software
triples out of the box performance*

Hardware and Software Services

- **Services Components as part of the turn-key solution:**
  - Project Management
  - Installation of Hardware
  - Installation of Software Stack
  - Testing & Validation of system in a table ready state
  - Information Transfer
  - Accelerated Value Program Advocate ongoing support

- **Duration:**
  - Overall Software services engagement:
    - Average 5 weeks effort without the Cognos configuration
    - Cognos Configuration is 2-3 weeks
  - Advocate Services for one-year following installation

- **Timeline example for installation in Texas:**

  Order to delivery: 4 weeks

  Hardware installation / Software installation: 1 day / 2 to 3 weeks

  Table ready and Ongoing support – 1 year

IBM Smart Analytics System – 9600 foundation

System z10 - Server Options:

- Choose - dedicated z10 server or bolt-on LPARs to existing server
  - Includes general processor, zIIP and IFL engines to meet capacity requirements
  - Supporting memory, channels, OSA card
  - Fixed configurations based on amount of data, number of concurrent Cognos users
  - Upgradeable as requirements grow
- Lab Services will define TWO LPARs
  - **z/OS LPAR** - for DB2 for z/OS database delivering highly available, secure environment
  - **z/VM LPAR** – multiple Linux virtual images – InfoSphere Warehouse, Cognos

Now on zEnterprise

InfoSphere Warehouse
Cognos 8.4 BI
Linux

DB2 for z/OS
z/OS

Storage Subsystem:

DS8700 Storage Subsystem

- Included in the 9600 configuration:
- **DEFAULT** - Installs with fiber channel drives
- **UPGRADEABLE** with 2 additional disk options:
  - Optional:
    - Solid State Disk
    - OR slower, more economical SATA drives
    - mix and match for workload
  - Automatic Data Management:
    - DB2 Online Reorg automatically moves data to SSD
    - EASY TIER automatically aligns data to disk types

IBM System z: System Design Comparison

System I/O Bandwidth
288 GB/Sec*

Balanced System
CPU, nWay, Memory,
I/O Bandwidth*

172.8 GB/sec*

96 GB/sec

24 GB/sec

Memory
3 TB*

PCI for
1-way
1202

1.5 TB**   512 GB   256 GB   64 GB   300   450   600   920

16-way

32-way

54-way

64-way

80-way
Processors

Legend:
- z196
- z10 EC
- z9 EC
- zSeries 990
- zSeries 900
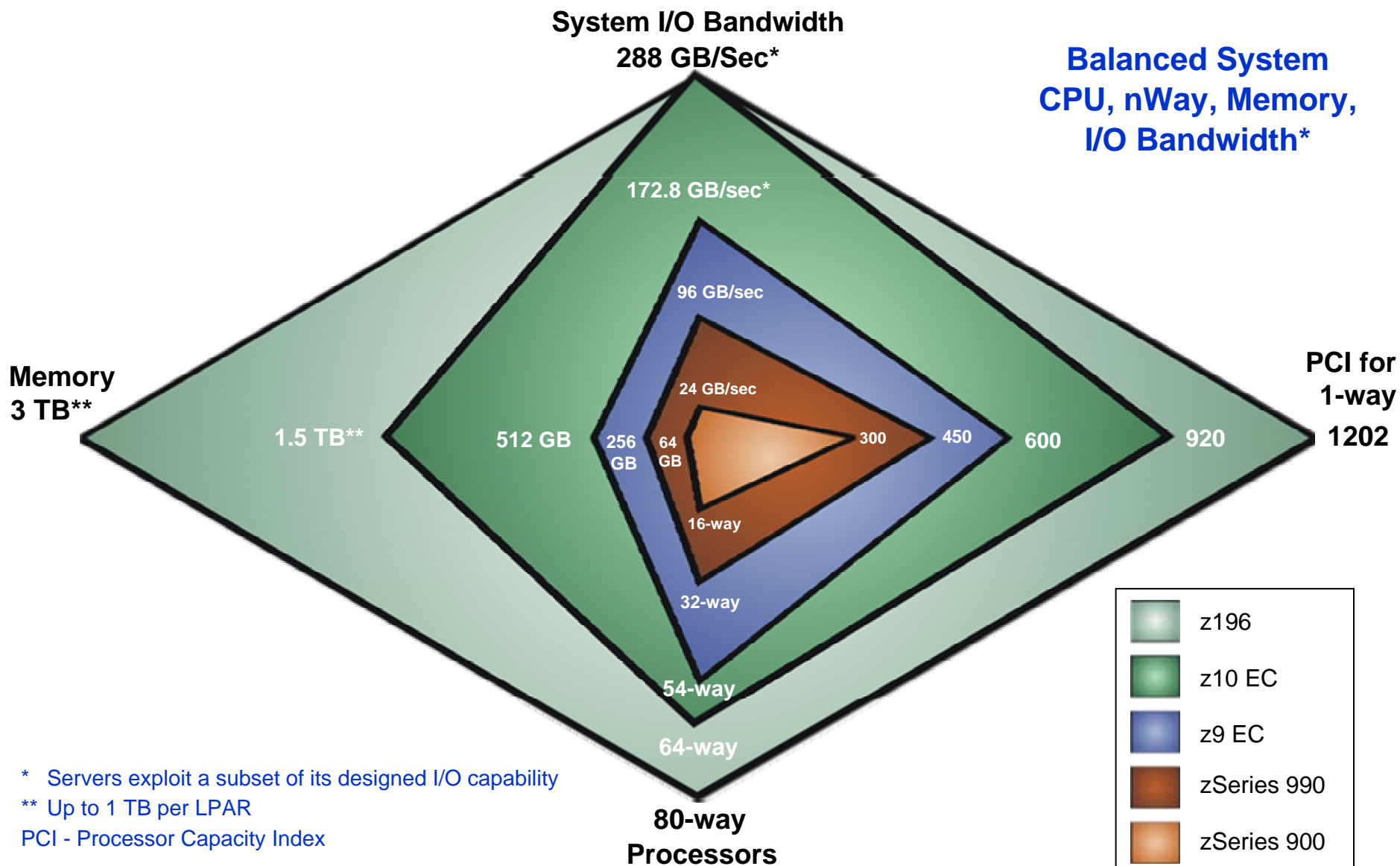
* Servers exploit a subset of its designed I/O capability
** Up to 1 TB per LPAR
PCI - Processor Capacity Index

z196 – The Core

- New 5.2 GHz Quad Core Processor Chip boosts hardware price/performance

  - 110+ new instructions – improvements for CPU intensive, Java™, and C++ applications

  - Over twice as much on-chip cache as System z10 to help optimize data serving environment

  - Out-of-order execution sequence gives significant performance boost for compute intensive applications

  - Significant improvement for floating point workloads

- Performance improvement for systems with large number of cores – improves MP ratio

- Data compression and cryptographic processors right on the chip

- Chip Area – 512.3mm^2
  - 23.5mm x 21.8mm
  - 13 layers of metal
  - 3.5 km wire
  - 1.4 Billion Transistors

Agenda

Data Warehouse Challenges

Smart Analytics System 9600

Data Warehousing with DB2 9 and 10 for z/OS

Accelerating BI workloads with the Smart Analytics Optimizer

Additional thoughts and Summary

Information Management for System z
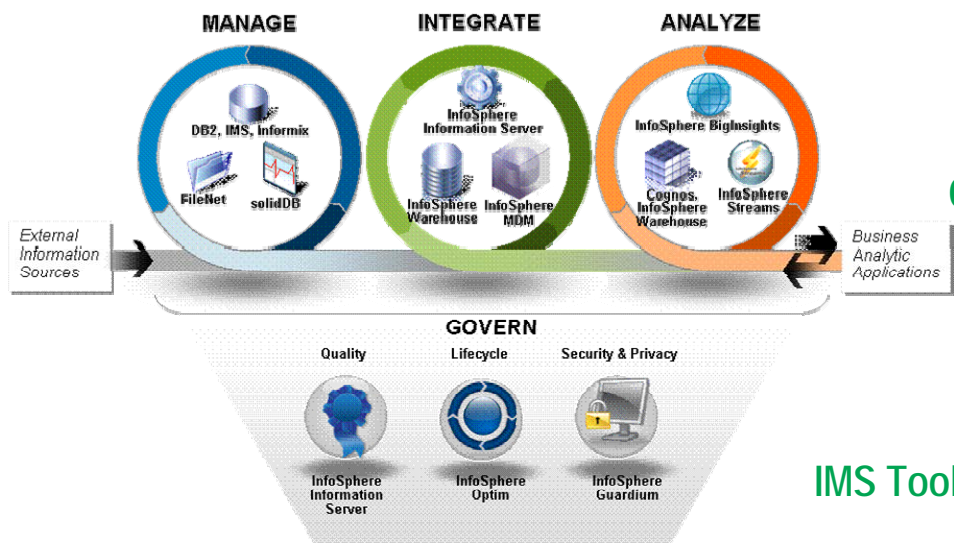
Smart Analytics Optimizer
OMEGAMON DB2 end–to-end monitoring   **2011**
DB2 10 for z/OS
DB2 Sort for z/OS
Guardium for System z
SPSS for System z
ISAS 9600 for System z
IMS Tools Solution Packs for z/OS   **Analytics**
IMS 11 for z/OS   **2010**
InfoSphere Discovery for z/OS
InfoSphere Warehouse for System z   *Information-Led*
Transformation
Cognos 8 BI for System z
InfoSphere MDM Server for System z   **Business Intelligence**
Optim Data Governance for System z   **2009**
Information Server for System z

Information Agenda

2008

Information On Demand   **Data Warehousing**

DB2 9 for z/OS – Data Warehousing was a major topic…

- SHRLEVEL(REFERENCE) for REORG of LOB tablespaces
- Online RENAME COLUMN
- Online RENAME INDEX
- Online CHECK DATA and CHECK LOB
- Online REBUILD INDEX
- Online ALTER COLUMN DEFAULT
- More online REORG by eliminating BUILD2 phase
- Faster REORG by intra-REORG parallelism
- Renaming SCHEMA, VCAT, OWNER, CREATOR
- LOB Locks reduction
- Skipping locked rows option
- Tape support for BACKUP and RESTORE SYSTEM utilities
- Recovery of individual tablespaces and indexes from volume-level backups
- Enhanced STOGROUP definition
- Conditional restart enhancements
- Histogram Statistics collection and exploitation
- WS II OmniFind based text search
- DB2 Trace enhancements
- WLM-assisted Buffer Pools management

- Global query optimization
- Dynamic index ANDing
- Generalizing sparse index and in-memory data caching method
- Optimization Service Center
- Autonomic reoptimization
- Logging enhancements
- LOBs network flow optimization
- Faster operations for variable-length rows
- Index on expressions
- Universal Tablespaces
- Partition-by-growth tablespaces
- APPEND option at insert
- Autonomic index page split
- Different index page sizes
- Support for optimistic locking
- Faster and more automatic DB2 restart
- RLF improvements for remote application servers such as SAP
- Preserving consistency when recovering individual objects to a prior point in time
- CLONE Table: fast replacement of one table with another
- Index compression
- Index key randomization

- DECIMAL FLOAT
- BIGINT
- VARBINARY, BINARY
- TRUNCATE TABLE statement
- MERGE statement
- FETCH CONTINUE
- ORDER BY and FETCH FIRST n ROWS in sub-select and full-select
- ORDER OF extension to ORDER BY
- INTERSECT and EXCEPT Set Operations
- RANK, DENSE
- Instead of triggers
- Various scalar and built-in functions
- Cultural sort
- LOB File Reference support
- XML support in DB2 engine
- Enhancements to SQL Stored Procedures
- SELECT FROM UPDATE/DELETE/MERGE
- Enhanced CURRENT SCHEMA
- IP V6 support
- Trusted Context
- Database ROLEs
- Automatic creation of database objects
- Temporary space consolidation
- zIIP enabled SQL Procedures
- . . .
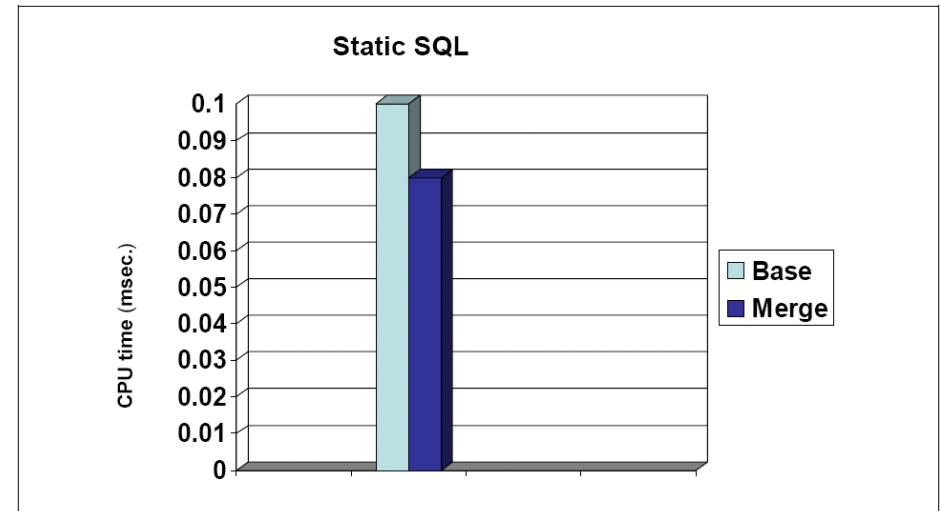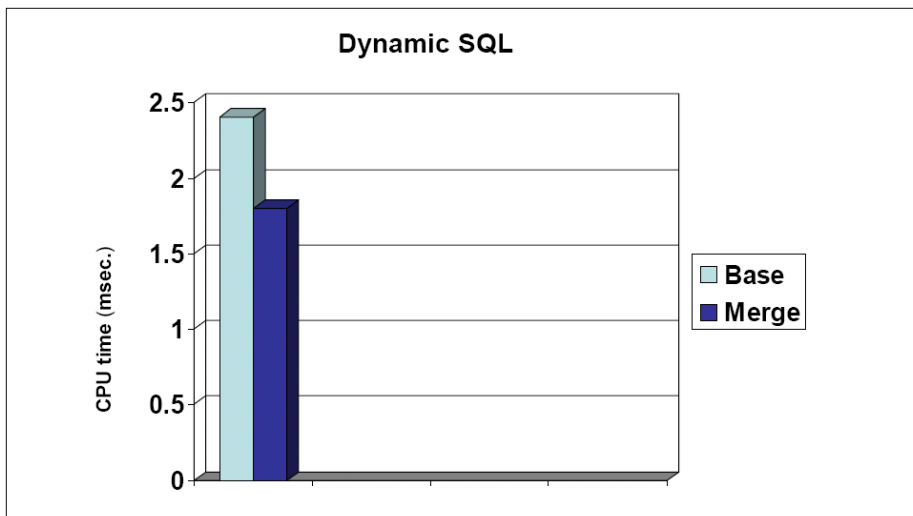
28

DB2 SQL 2009

z   z/OS 9
common
luw Linux, Unix & Windows 9.7

z
Multi-row INSERT, FETCH & multi-row cursor UPDATE, Dynamic Scrollable Cursors, GET DIAGNOSTICS, Enhanced UNICODE SQL, join across encoding schemes, IS NOT DISTINCT FROM, VARBINARY, FETCH CONTINUE, MERGE, SELECT from MERGE

common
Inner and Outer Joins, Table Expressions, Subqueries, GROUP BY, Complex Correlation, Global Temporary Tables, CASE, 100+ Built-in Functions including SQL/XML, Limited Fetch, Insensitive Scroll Cursors, UNION Everywhere, MIN/MAX Single Index, Self Referencing Updates with Subqueries, Sort Avoidance for ORDER BY, and Row Expressions, 2M Statement Length, GROUP BY Expression, Sequences, Scalar Fullselect, Materialized Query Tables, Common Table Expressions, Recursive SQL, CURRENT PACKAGE PATH, VOLATILE Tables, Star Join Sparse Index, Qualified Column names, Multiple DISTINCT clauses, ON COMMIT DROP, Transparent ROWID Column, Call from trigger, statement isolation, FOR READ ONLY KEEP UPDATE LOCKS, SET CURRENT SCHEMA, Client special registers, long SQL object names, SELECT from INSERT, UPDATE or DELETE, INSTEAD OF TRIGGER, Native SQL Procedure Language, BIGINT, file reference variables, XML, FETCH FIRST & ORDER BY in subselect & fullselect, caseless comparisons, INTERSECT, EXCEPT, not logged tables, OmniFind, spatial, range partitions, data compression, session variables, DECIMAL FLOAT, optimistic locking, ROLE, **TRUNCATE, index & XML compression, created temps**

luw
Updateable UNION in Views, GROUPING SETS, ROLLUP, CUBE, **more** Built-in Functions, SET CURRENT ISOLATION, multi-site join, MERGE, MDC, XQuery, XML enhancements, array data type, global variables, **even more vendor syntax, LOB & temp table compression, inline LOB, administrative privileges, implicit casting, date/time changes, currently committed**

MERGE

- MERGE instead of UPDATE and INSERT on condition

- Standard approach can be part of an incremental updating strategy while fixing potential data inconsistencies at the same time

- Simplifies SQL coding

Dynamic SQL

CPU time (msec.)
- 2.5
- 2
- 1.5
- 1
- 0.5
- 0

Legend:
- Base
- Merge

Static SQL

CPU time (msec.)
- 0.1
- 0.09
- 0.08
- 0.07
- 0.06
- 0.05
- 0.04
- 0.03
- 0.02
- 0.01
- 0

Legend:
- Base
- Merge

TRUNCATE TABLE

- Delete all rows from a table

  - Issues with today's solutions: LOAD REPLACE, DELETE

  - Can be used to remove a not-logged TS from the LPL and to reset RECP

  - Applicable for Simple TS, Segmented TS, Partitioned TS and Universal TS

  - The performance ?

  - TRUNCATE TABLE tablename

DB2 10 for z/OS – More Data Warehousing topics

- Compress on the fly on INSERT
- Auto-statistics
- Access path stability and hints enhancements
- Access path lock-in and fallback for dynamic SQL
- Automatic checkpoint interval
- Automated installation, configuration & activation of DB2 supplied stored procedures & UDFs
- Data set FlashCopy in COPY & inline copy
- Inline image copies for COPY YES indexes
- UNLOAD from FlashCopy backup
- REORG enhancements
- Reduce need for reorganizations for indices
- CPU reductions
- Hash access path
- Versioned data

- Numerous optimizer enhancements, paging through result sets
- Parallel index update at insert
- Faster single row retrievals
- Inline LOBs
- LOB streaming between DDF and rest of DB2
- Faster fetch and insert, lower virtual storage consumption
- DEFINE NO for LOBs and XML
- MEMBER CLUSTER for UTS
- Query parallelism enhancements: lifting restrictions
- Dynamic Index ANDing Enhancements
- Option to avoid index entry creation for NULL value
- Index include columns
- Buffer pool enhancements
- Increased number of active threads
- Reducing latch contention

- Workfile spanned records, PBG support, and in-memory enhancements
- More online schema changes for table spaces, tables and indexes via online REORG
- Online REORG for LOBs
- Online add log
- Automatically delete CF structures before/during first DB2 restart
- Allow non-NULL default values for inline LOBs
- Loading and unloading tables with LOBs in stream
- Currently committed locking semantics
- More granular DBA privileges
- SQL compatibility
- Additional zIIP eligible workloads

DB2 SQL 2010

z   z/OS 10
common
luw Linux, Unix & Windows 9.8

z Multi-row INSERT, FETCH & multi-row cursor UPDATE, Dynamic Scrollable Cursors, GET DIAGNOSTICS, Enhanced UNICODE SQL, join across encoding schemes, IS NOT DISTINCT FROM, VARBINARY, FETCH CONTINUE, MERGE, SELECT from MERGE, **data versioning, access controls**

common Inner and Outer Joins, Table Expressions, Subqueries, GROUP BY, Complex Correlation, Global Temporary Tables, CASE, 100+ Built-in Functions including SQL/XML, Limited Fetch, Insensitive Scroll Cursors, UNION Everywhere, MIN/MAX Single Index, Self Referencing Updates with Subqueries, Sort Avoidance for ORDER BY, and Row Expressions, 2M Statement Length, GROUP BY Expression, Sequences, Scalar Fullselect, Materialized Query Tables, Common Table Expressions, Recursive SQL, CURRENT PACKAGE PATH, VOLATILE Tables, Star Join Sparse Index, Qualified Column names, Multiple DISTINCT clauses, ON COMMIT DROP, Transparent ROWID Column, Call from trigger, statement isolation, FOR READ ONLY KEEP UPDATE LOCKS, SET CURRENT SCHEMA, Client special registers, long SQL object names, SELECT from INSERT, UPDATE or DELETE, INSTEAD OF TRIGGER, Native SQL Procedure Language, BIGINT, file reference variables, XML, FETCH FIRST & ORDER BY in subselect & fullselect, caseless comparisons, INTERSECT, EXCEPT, not logged tables, OmniFind, spatial, range partitions, data compression, session variables, DECIMAL FLOAT, optimistic locking, ROLE, TRUNCATE, index & XML compression, created temps, **inline LOB, administrative privileges, implicit casting, date/time changes, currently committed, moving sum & avg.**

luw Updateable UNION in Views, GROUPING SETS, ROLLUP, CUBE, more Built-in Functions, SET CURRENT ISOLATION, multi-site join, MERGE, MDC, XQuery, XML enhancements, array data type, global variables, even more vendor syntax, LOB & temp table compression,

Running a Large Number of Active Threads

Today

Coupling Technology

LPAR1	LPAR2	LPAR3
DB2A (500 thds)	**DB2B** (500 thds)	**DB2C** (500 thds)
DB2D (500 thds)	**DB2E** (500 thds)	**DB2F** (500 thds)
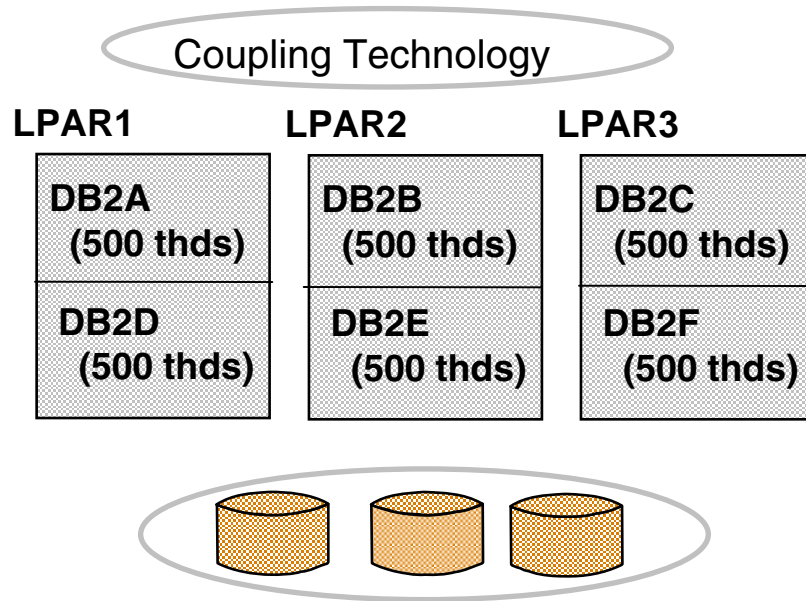
- Data sharing and sysplex allows for efficient scale-out of DB2 images
- Sometimes multiple DB2s per LPAR

DB2 10

Coupling Technology

LPAR1	LPAR2	LPAR3
DB2A (2500 thds)	**DB2B** (2500 thds)	**DB2C** (2500 thds)
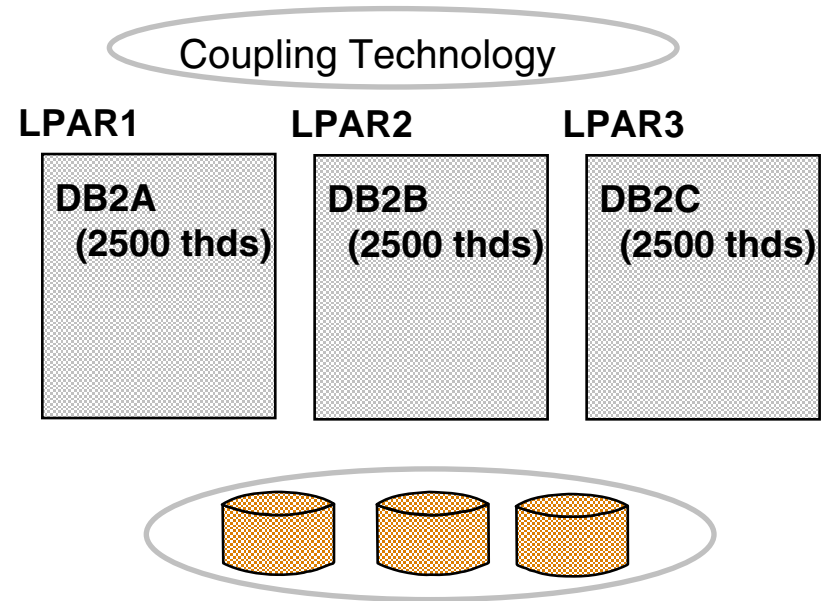
- More threads per DB2 image
- More efficient use of large n-ways
- Easier growth, lower costs, easier management
- Data sharing and Parallel Sysplex still required for very high availability and scale
- Rule of thumb: save ½% CPU for each member reduced, more on memory

Improved Availability – ALTER + REORG

Pending ALTER,
then online REORG
to make changes

Range-Partitioned
UTS PBR

ONE WAY

ONE WAY

Single-Table
Simple
Table Space

Classic Partitioned
Table Space

Single-Table
Segmented
Table Space

ONE WAY

ONE WAY

Page size
DSSIZE
SEGSIZE
MEMBER CLUSTER

Partition-By-Growth
UTS PBG

Hash

ONE WAY

INDEX page size
INCLUDE cols

ADD active log
BUFFERPOOL PGSTEAL NONE
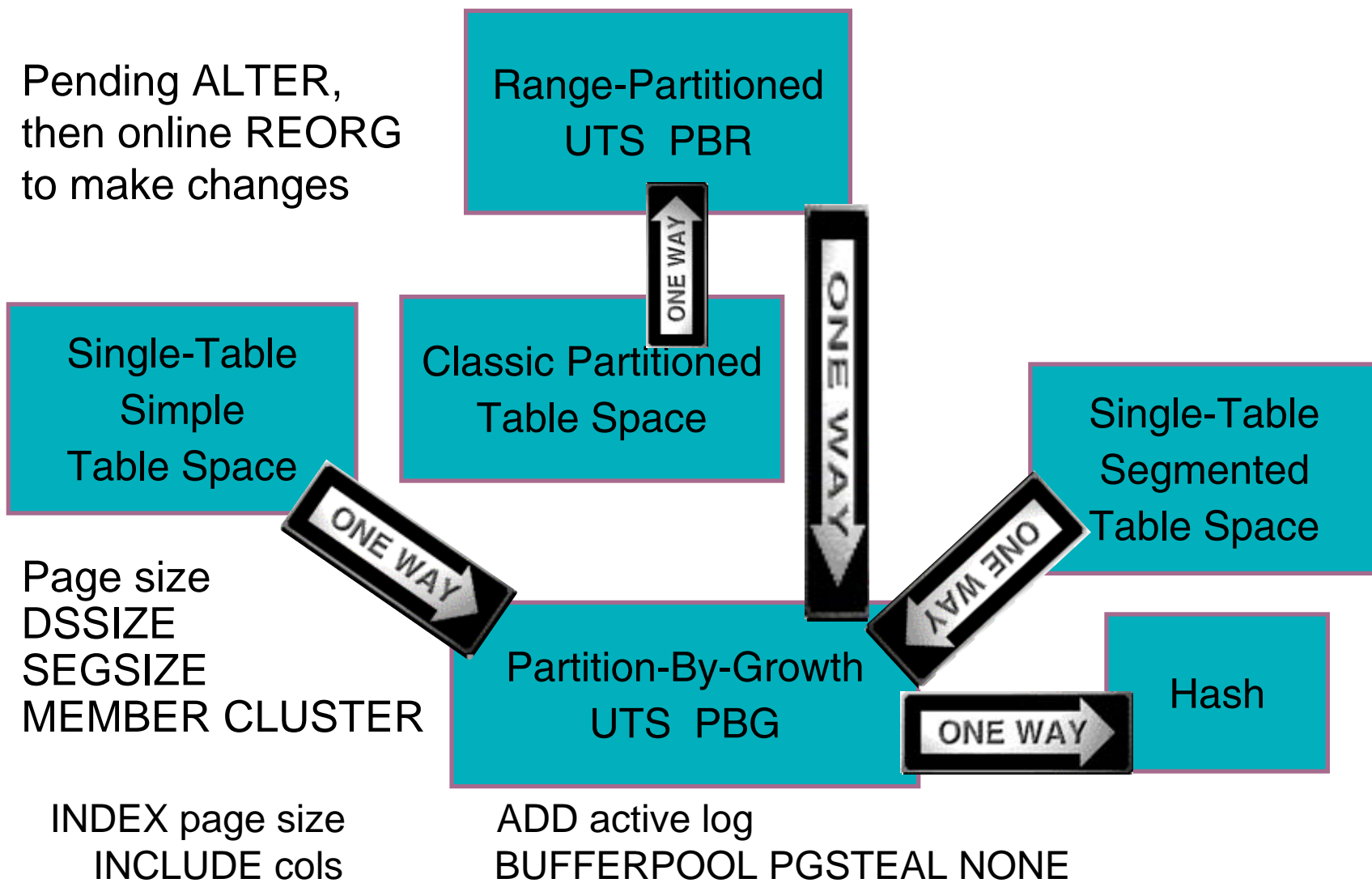
Compress on INSERT

- Data compression occurs when a dictionary exists

- **Prior to DB2 V10**

  - Dictionary not built on a table space with COMPRESS YES attribute until:
    - REORG or
    - LOAD utility was executed
  - For some customers, REORG or LOAD are not executed frequently

- **DB2 V10 NFM** allows for build of compression dictionary on:

  - INSERT

  - MERGE

  - LOAD SHRLEVEL CHANGE

- **No REORG or LOAD needed to get compressed data**

Safe Query Optimization

- **Bind/Prepare:**

  - Optimizer will evaluate the risk associated with each predicate
    - For example: WHERE BIRTHDATE < ?
      - Could qualify 0-100% of data depending on literal value used
  - As part of access path selection
    - Compare access paths with close cost and choose lowest risk plan

- **Runtime:**

  - If a RID limit is reached

    - Overflow RIDs to workfile and continue processing
    - Avoid fallback to table space scan as in V9.

  - Work-file usage may increase

    - Mitigate by increasing RID pool size (default increased in DB2 10).

Additional Non-Key Columns in an Index

- Index is used to enforce uniqueness constraint on table

- To achieve index only access on columns not part of the unique constraint during queries, additional indexes are often created for those columns not part of the unique constraint

  - Slower DB2 transaction time

  - Increased storage requirements

- Additional Non-key Columns can be defined in an unique index to reduce total amount of needed indexes

- Improves

  - insert performance as less indexes needs to be updated

  - space usage

  - Can stabilize access path as optimizer has less similar indexes to chose from

Additional Non-Key Columns in an Index

- **V9 definition**

  - CREATE UNIQUE INDEX i1 ON t1(c1,c2,c3)

  - CREATE (UNIQUE) INDEX i2 ON t1(c1,c2,c3,c4,c5)

- **Possible V10 definition**

  - CREATE UNIQUE INDEX ON t1(c1,c2,c3) INCLUDE (c4,c5)

  -     or

  - ALTER INDEX i1 ADD INCLUDE (c4,c5)   and DROP INDEX i2

- **The following restrictions will apply:**

  - INCLUDE columns are not allowed in non-unique indexes

  - Indexes on Expression will not support INCLUDE columns

  - Indexes with INCLUDEd columns can not have additional unique columns ALTER ADDed to the index

Versioned data - Current and History

Current SQL Application

Current

History
Generation →

Current ←

↔ Transparent/automatic Access to satisfy ASOF Queries

Auditing SQL Application Using ASOF

History

Jul 2008
Aug 2008
Sep 2008
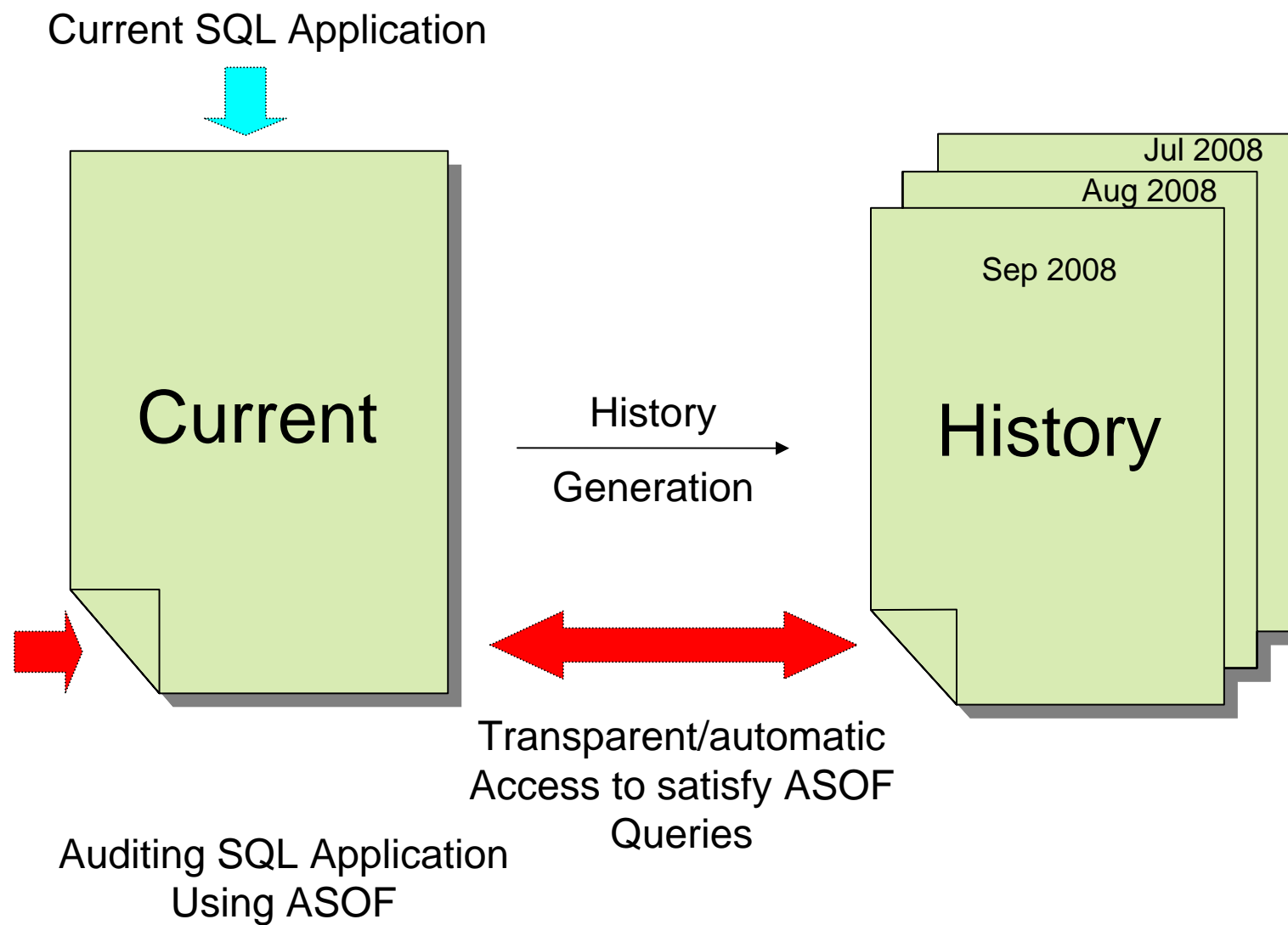
New SQL functions

- **Moving Sum, Moving Average**

- **Enhanced query parallelism technology for improved performance**

  - Remove query parallelism restrictions

- **In-memory techniques for faster query performance**

Moving Sum, Moving Average

- **OLAP Specification defined using a window.**
  - A window may specify 3 optional components

    - Partitioning – PARTITION BY clause

    - Ordering – ORDER BY clause

    - Aggregation group – ROWS/RANGE

Sales_History Table

Territory	Month	Sales
East	199810	10
East	199811	4
East	199812	10
East	199901	7
East	199902	10
West	199810	8
West	199811	12
West	199812	7
West	199901	11
West	199902	6

Compute the average sales over the current month and the preceding two months for each territory and month:

SELECT Territory, Month, Sales, AVG(Sales)

OVER(PARTITION BY Territory ORDER BY Month ROWS 2 PRECEDING)

AS Moving_Average

FROM Sales_History;

Territory	Month	Sales	Moving_Average
East	199810	10	10
East	199811	4	7
East	199812	10	8
East	199901	7	7
East	199902	10	9
West	199810	8	8
West	199811	12	10
West	199812	7	9
West	199901	11	10
West	199902	6	8

Compute the average sales over the current month, the preceding
Month, and the following month for each territory and month:

SELECT Territory, Month, Sales, AVG(Sales)

OVER(PARTITION BY Territory

      ORDER BY Month

      ROWS BETWEEN 1 PRECEDING AND 1 FOLLOWING)

AS Moving_Average

FROM Sales_History;

Territory	Month	Sales	*Moving_Average*
East	199810	10	7
East	199811	4	8
East	199812	10	7
East	199901	7	9
East	199902	10	8
West	199810	8	10
West	199811	12	9
West	199812	7	10
West	199901	11	8
West	199902	6	8

Compute the cumulative sum of sales in each territory, ordered by month:

SELECT Territory, Month, Sales, SUM(Sales)

OVER(PARTITION BY Territory ORDER BY Month ASC

    ROWS UNBOUNDED PRECEDING)

AS Cumulative_Sum

FROM Sales_History;

Territory	Month	Sales	Cumulative_Sum
East	199810	10	10
East	199811	4	14
East	199812	10	24
East	199901	7	31
East	199902	10	41
West	199810	8	8
West	199811	12	20
West	199812	7	27
West	199901	11	38
West	199902	6	44

DB2 Workloads that leverage zIIP

Portions of the following DB2 for z/OS workloads will benefit from zIIP:

1. **ERP, CRM, Business Intelligence or other enterprise applications**

   - Via DRDA over a TCP/IP connection
     - Workloads using DB2 Connect, T4 JCC Universal driver, Data Server Client and CLI/ODBC, JDBC, SQLJ, .NET, pureQuery APIs
   - Remote native SQL procedures (DB2 9 for z/OS NFM)
   - XML parsing processing (DB2 9 for z/OS NFM)

2. **Data warehousing / Business Intelligence applications**

   - CPU intensive parallel queries, including star schema queries

3. **DB2 for z/OS utility functions used to maintain index structures and DFSORT sort for some utilities**

- **Usage of zIIP is transparent to applications**

   - No changes to applications

What is the zAAP on zIIP capability?

- **A new capability that can enable System z Application Assist Processor (zAAP) eligible workloads to run on System z Integrated Information Processors (zIIPs).**

  - For customers with no zAAPs and zIIPs

    - The combined eligible workloads may make the acquisition of a single zIIP cost effective.

  - For customers with only zIIP processors

    - Makes Java and z/OS XML System Services -based workloads eligible to run on existing zIIPs – maximizes zIIP investment.

- **Available September 25, 2009 with z/OS V1.11 and z/OS V1.9 and V1.10 (with PTF)**

  - This new capability is not available for z/OS LPARS if zAAPs are installed on the server.

How to enable the zAAP on zIIP capability

- **The capability ships default enabled with z/OS V1.11.**
  - Parameter in SYS1.PARMLIB(IEASYSxx) : ZAAPZIIP = YES (default in z/OS V1.11)
  - If you wish to disable the function for any reason, you must IPL with ZAAPZIIP=NO in the IEASYSxx Parmlib member.

- **Also available with z/OS V1.9 and V1.10**
  - With PTF for APAR OA27495, and
  - Enabled with ZAAPZIIP=YES in the IEASYSxx Parmlib (the default is NO)

- **This new capability does not remove the requirement to purchase and maintain one or more general purpose processors for every zIIP processor on the server.**

Agenda

Data Warehouse Challenges

Smart Analytics System 9600

Data Warehousing with DB2 9 and 10 for z/OS

Accelerating BI workloads with the Smart Analytics Optimizer

Additional thoughts and Summary

IBM Smart Analytics Optimizer

What is it?

The IBM Smart Analytics Optimizer is a workload optimized, appliance-like, add-on, that enables the integration of business insights into operational processes to drive winning strategies. It accelerates select queries, with unprecedented response times.

How is it different?

- **Performance:** Unprecedented response times to enable 'train of thought' analysis frequently blocked by poor query performance.

- **Integration:** Connects to DB2 through deep integration providing transparency to all applications.

- **Self-managed workloads:** queries are executed in the most efficient way

- **Transparency:** applications connected to DB2, are entirely unaware of the Smart Analytics Optimizer

- **Simplified administration:** appliance-like hands-free operations, eliminating many database tuning tasks

Breakthrough Technology Enabling New Opportunities

Extreme performance for complex queries

Game Changing Performance

✓ **Rapidly delivers information to decision makers through breakthrough technologies providing dramatic performance improvement.**

    ✓ Technology optimized for fast scans

    ✓ Highly compressed data

    ✓ Compressed data operations

    ✓ Hardware-optimized data analysis

    ✓ Massively parallel architecture

✓ **Enables decision makers to submit queries they never dared in the past, that analyze trends, predict outcomes, and produce better business results.**

One Beta customer asked us to repeat a query under lab conditions because he couldn't believe the acceleration. A query execution time was reduced from 13 minutes, 42 seconds to just one second (end to end)!
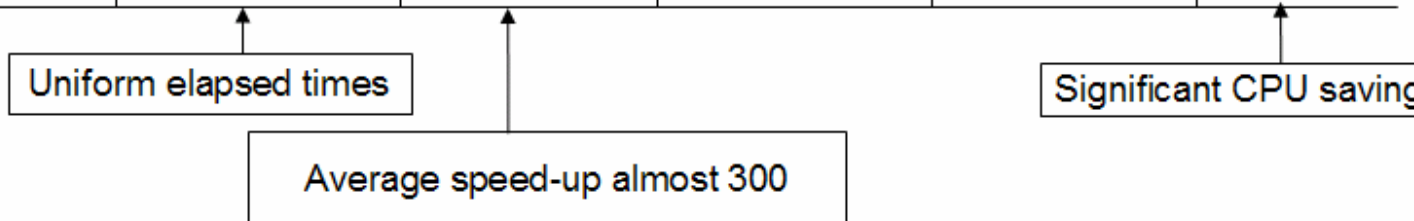
Test Results with real Customer Data

- The problem queries were provided by a customer
- Expert database tunning done on all the queries
    - Q1 – Q6 even after tuning run for too long and consume lots of resources
    - Q7 improved significantly – no IBM Smart Analytics Optimizer offload is needed
    - The provided workload was representing the IBM Smart Analytics Optimizer "Sweetspot" perfectly.
    - DB2 measurements were done on a System z9.

- The table shows elapsed and CPU times measured in DB2
  (**w/o Smart Analytics Optimizer**)

Query	Elsapsed time	CP	zIIP	Total CP time
Q1	0:02:43	0:03:52	0:02:39	0:06:31
Q2	0:38:31	0:11:52	0:36:10	0:48:02
Q3	0:00:25	0:00:04	0:00:15	0:00:19
Q4	0:26:33	0:13:43	0:20:50	0:34:33
Q5	0:00:35	0:00:09	0:00:29	0:00:38
Q6	1:30:35	5:53:30	1:29:56	7:23:26
Q7	0:00:02	0:00:02	0:00:00	0:00:02

Test Results with real Customer Data

Query	Query Elapsed Time			Query CPU Consumption on System z		
	DB2 only	Smart Analytics Optimizer	Speed-up	DB2only	Smart Analytics Optimizer	Saving
Q1	0:02:43.0	0:00:03.4	48	0:06:31.0	0.004495	~100%
Q2	0:38:31.0	0:00:04.5	511	0:48:02.0	0.004713	~100%
Q3	0:00:25.0	0:00:02.2	12	0:00:19.0	0.099702	99.48%
Q4	0:26:33.0	0:00:07.8	206	0:34:33.0	0.005174	~100%
Q5	0:00:35.0	0:00:08.3	4	0:00:38.0	0.520915	98.63%
Q6	1:30:35.0	0:00:03.8	1424	7:23:26.0	0.003979	~100%
Q7	0:00:02	0:00:02	1	0:00:02	2.0	N/A
Total	2:39:24.0	0:00:32.0	298	8:53:31.0	2.64	99.99%

Uniform elapsed times

Average speed-up almost 300

Significant CPU saving
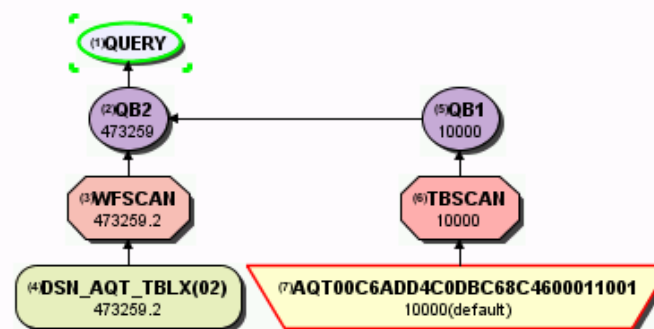
3/16/2011

© 2011 IBM Corporation
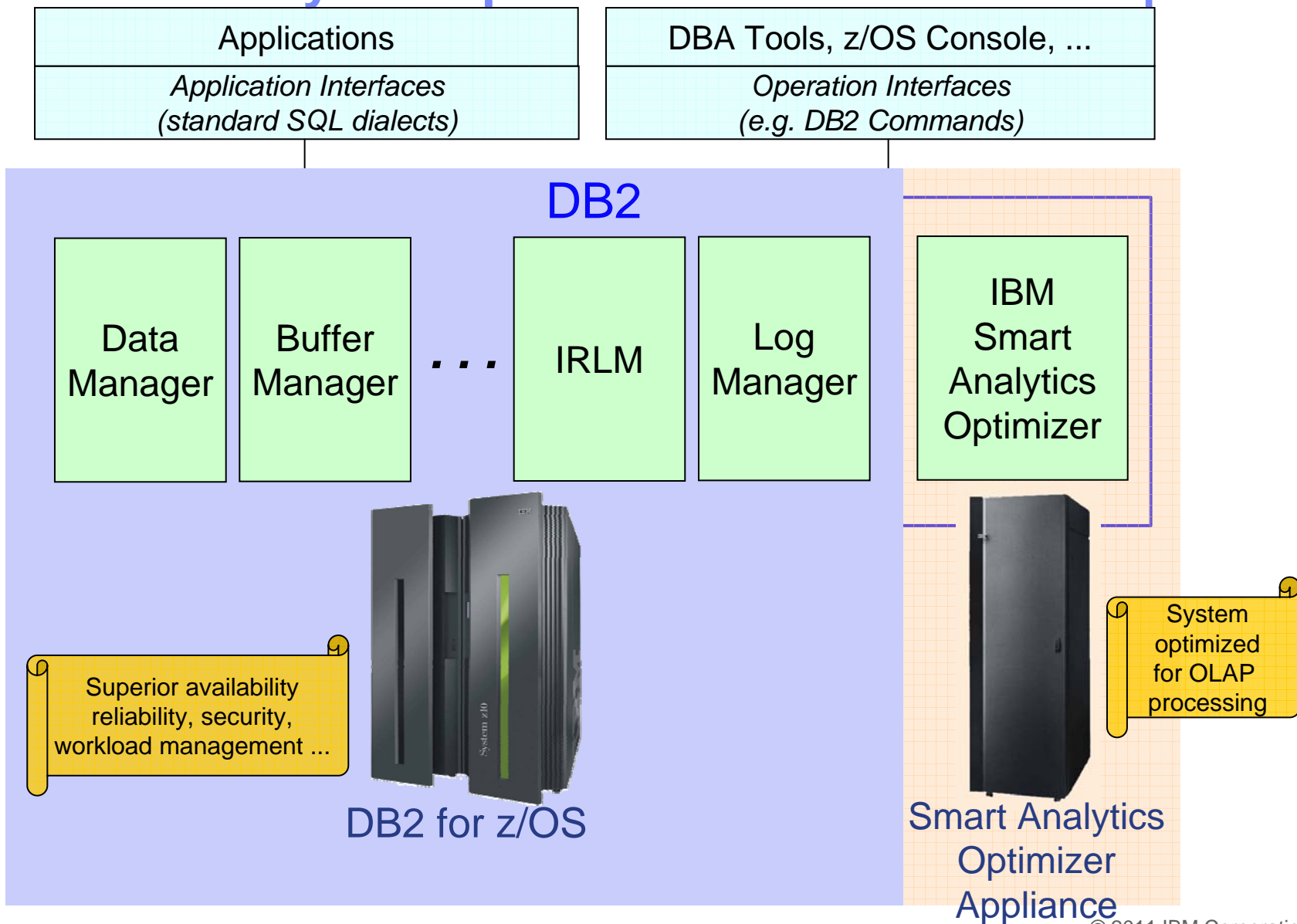
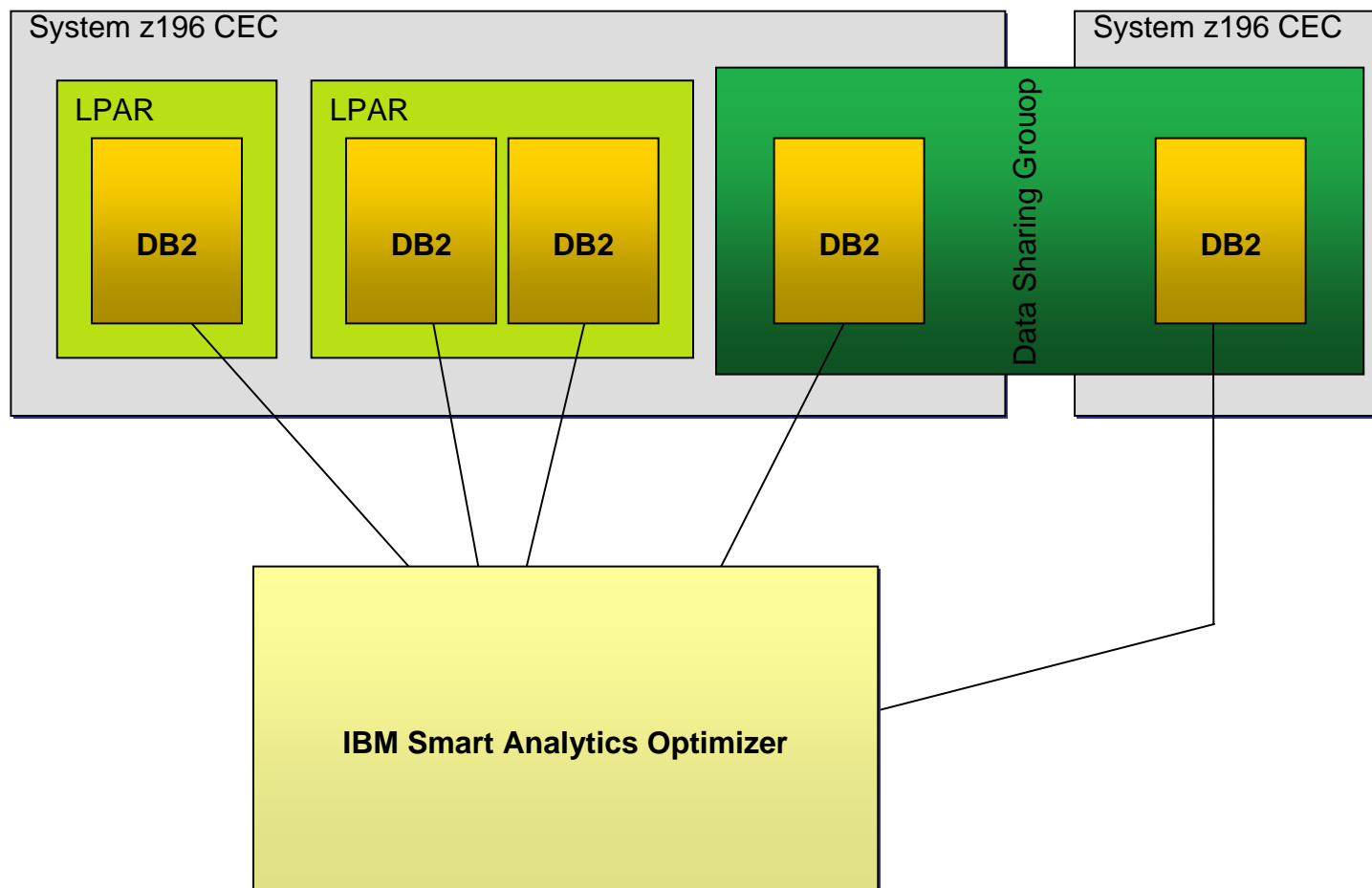# Access plan comparison

Access plan graph without IBM Smart Analytics Optimizer:
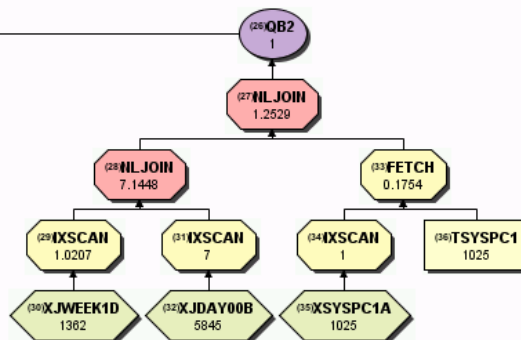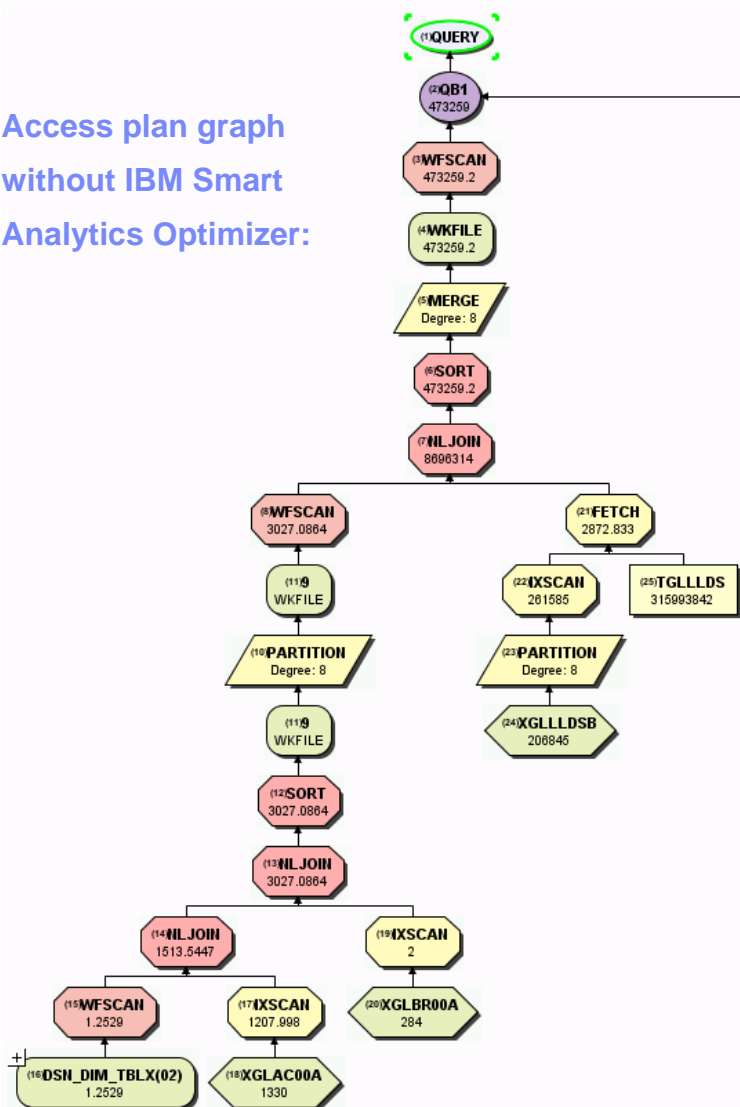
Access plan graph with IBM Smart Analytics Optimizer:

IBM Smart Analytics Optimizer - a virtual DB2 Component

Applications	DBA Tools, z/OS Console, ...
Application Interfaces (standard SQL dialects)	*Operation Interfaces (e.g. DB2 Commands)*

DB2

Data Manager	Buffer Manager	. . .	IRLM	Log Manager	IBM Smart Analytics Optimizer

Superior availability reliability, security, workload management ...

DB2 for z/OS

System optimized for OLAP processing

Smart Analytics Optimizer Appliance

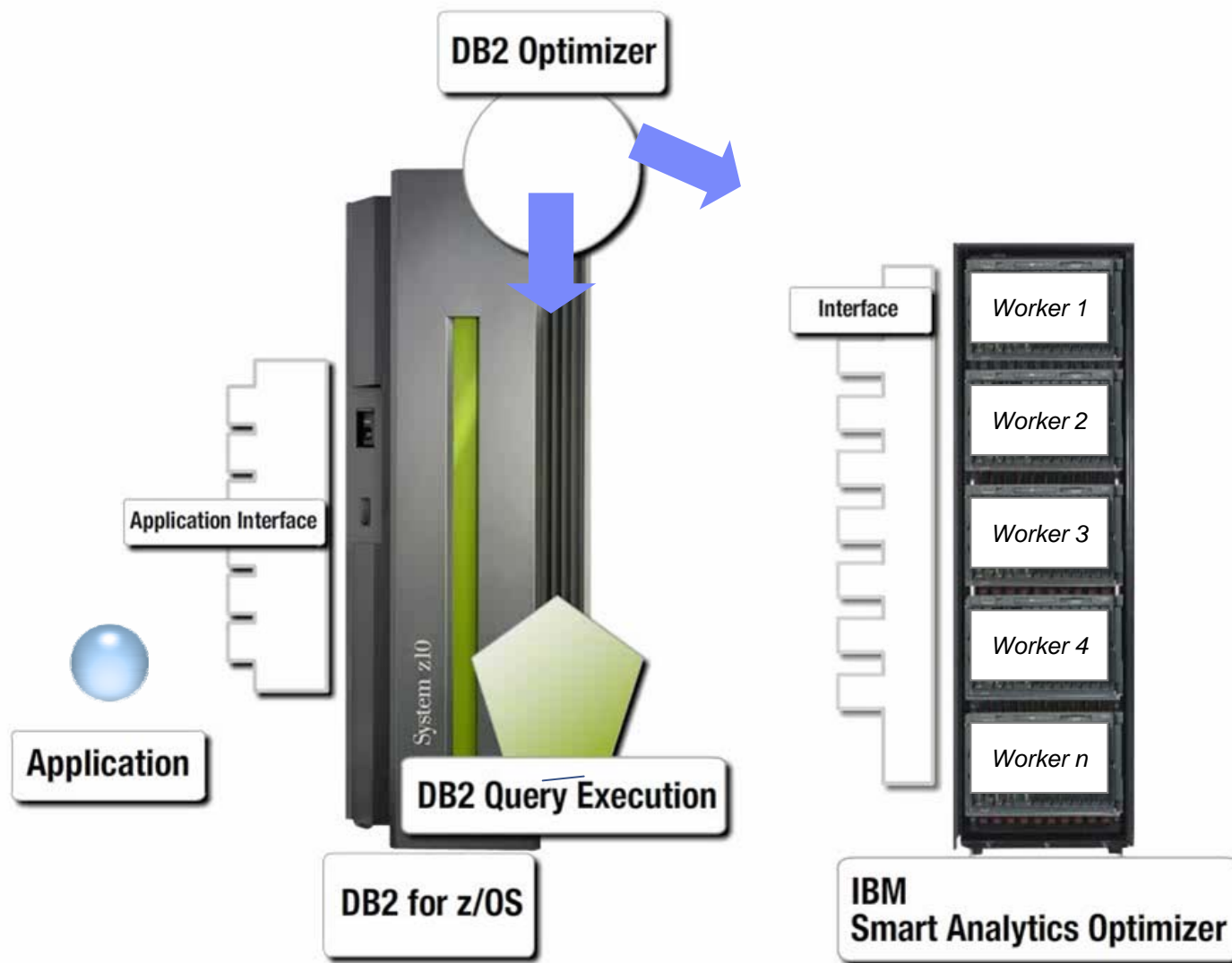# IBM Smart Analytics Optimizer with multiple DB2 subsystems

Routing Queries to the Optimizer

- **Routing of Queries to the Smart Analytics Optimizer is determined by the DB2 for z/OS Optimizer**

Query Execution Flow

Integration also into DB2 Commands: DISPLAY THREAD

```
-DIS THD(*) ACC(*) DETAIL

DSNV401I # DISPLAY THREAD REPORT FOLLOWS -
DSNV402I # ACTIVE THREADS -
NAME   ST A   REQ ID   AUTHID    PLAN      ASID TOKEN
BATCH  AC *    231 BI   ADMF001   DSNTEP2 0053 55
V666 ACC=ISAOPT1,ADDR=::FFFF:9.30.30.133..446:1076
```
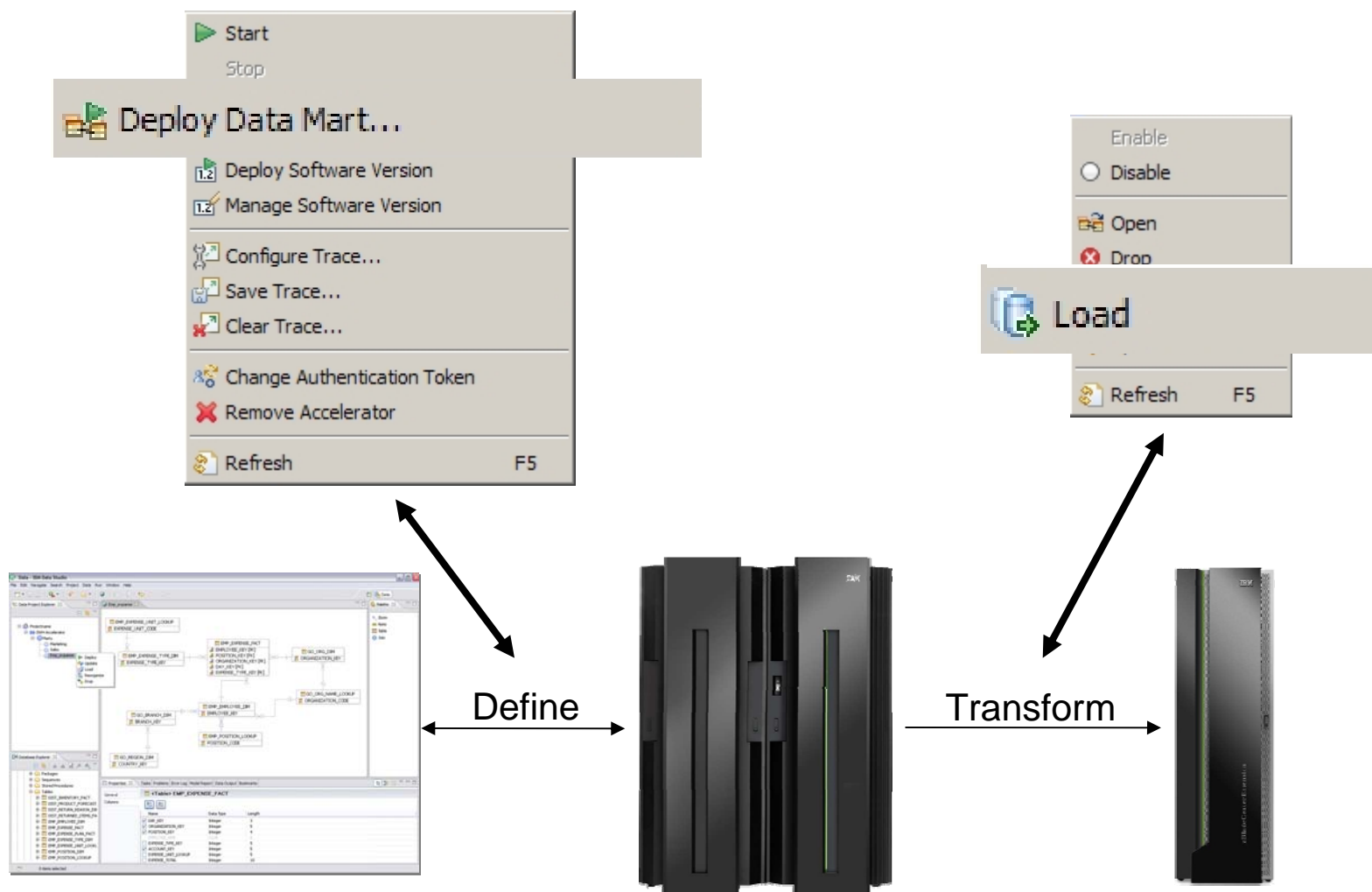
Defining tables for acceleration

Start
Stop

Deploy Data Mart...

1.2 Deploy Software Version
1.2 Manage Software Version

Configure Trace...
Save Trace...
Clear Trace...

Change Authentication Token
Remove Accelerator

Refresh   F5

Enable
○ Disable

Open
Drop

Load

Refresh   F5

Define

Transform

Comparing the Smart Analytics offerings in System z:

IBM Smart Analytics **System 9600**	IBM Smart Analytics **Optimizer**
GA June 2010	**Uses Services of Accelerator** → GA 4Q 2010
Complete, end-to-end environment for BI workload Processes ALL queries submitted by end-users **Software:** Includes z/OS, DB2 for z/OS, Linux, InfoSphere Warehouse, Cognos, DB2 Connect **Supports:** Data movement / ETL End user tools (Cognos) / BI queries/reports Data Storage (Data warehouse) Runs in z/OS-DB2 LPAR, Linux for System z LPAR for Tooling Is an all purpose environment to deploy any BI workload	**Workload optimized, appliance-like, add-on to a Data Warehouse on System z** *MUST connect TO a DB2 for z/OS environment that is running a BI workload* Will enhance a Smart Analytics System 9600 **Software:** Includes software running on optimized hardware and connects to DB2 for z/OS to enable query routing to the Smart Analytics Optimizer **Supports:** (Subset of) complex and long-running BI queries with access to data in a star or snowflake schema Order of magnitude acceleration of a SUBSET of queries that are selected/routed.
For acceleration of multidimensional star schema queries, use IBM Smart Analytics Optimizer as add on option	

Agenda

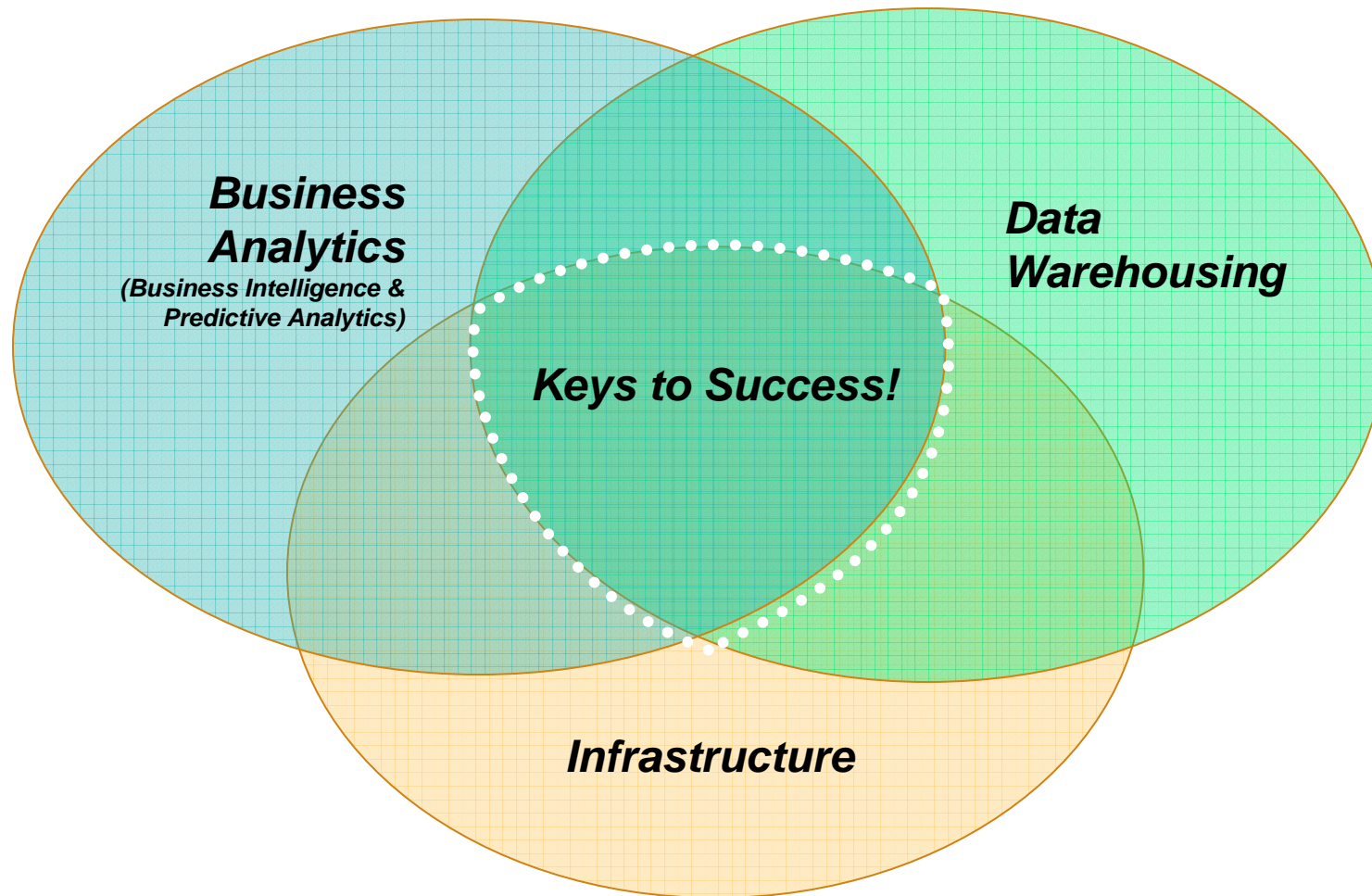Data Warehouse Challenges

Smart Analytics System 9600

Data Warehousing with DB2 9 and 10 for z/OS

Accelerating BI workloads with the Smart Analytics Optimizer

Additional thoughts and Summary

A successful strategy requires the integration of all key components

Business Analytics *(Business Intelligence & Predictive Analytics)*

Data Warehousing

Keys to Success!

Infrastructure

ETL Solution Accelerates with Automation & Reuse
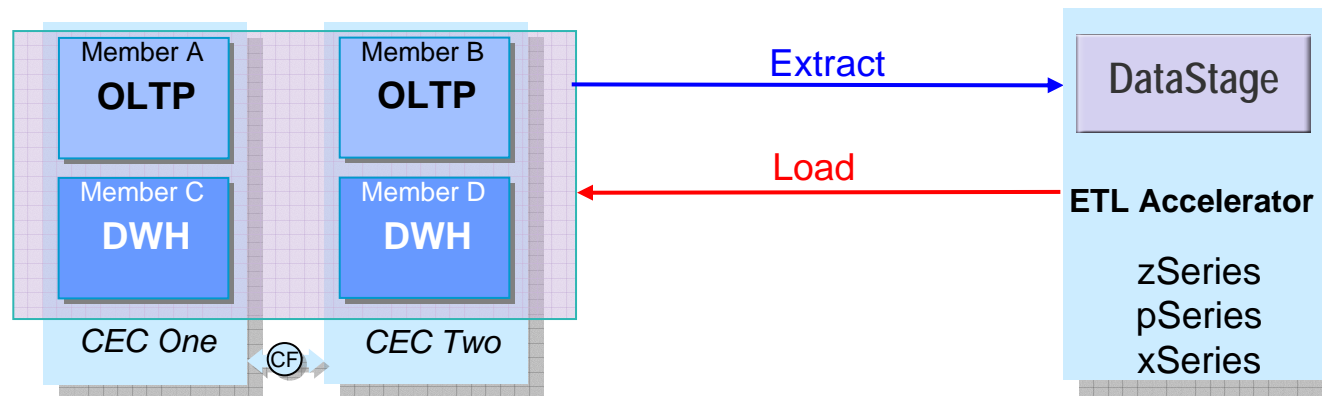
- ETL is done using an ETL Accelerator like **DataStage.**

- *E* – the data is extracted from the OLTP information sources

- *T* – 100s of embedded transforms & custom built transforms are applied

- *L* – Aggregated and transformed data is loaded into the warehouse tables

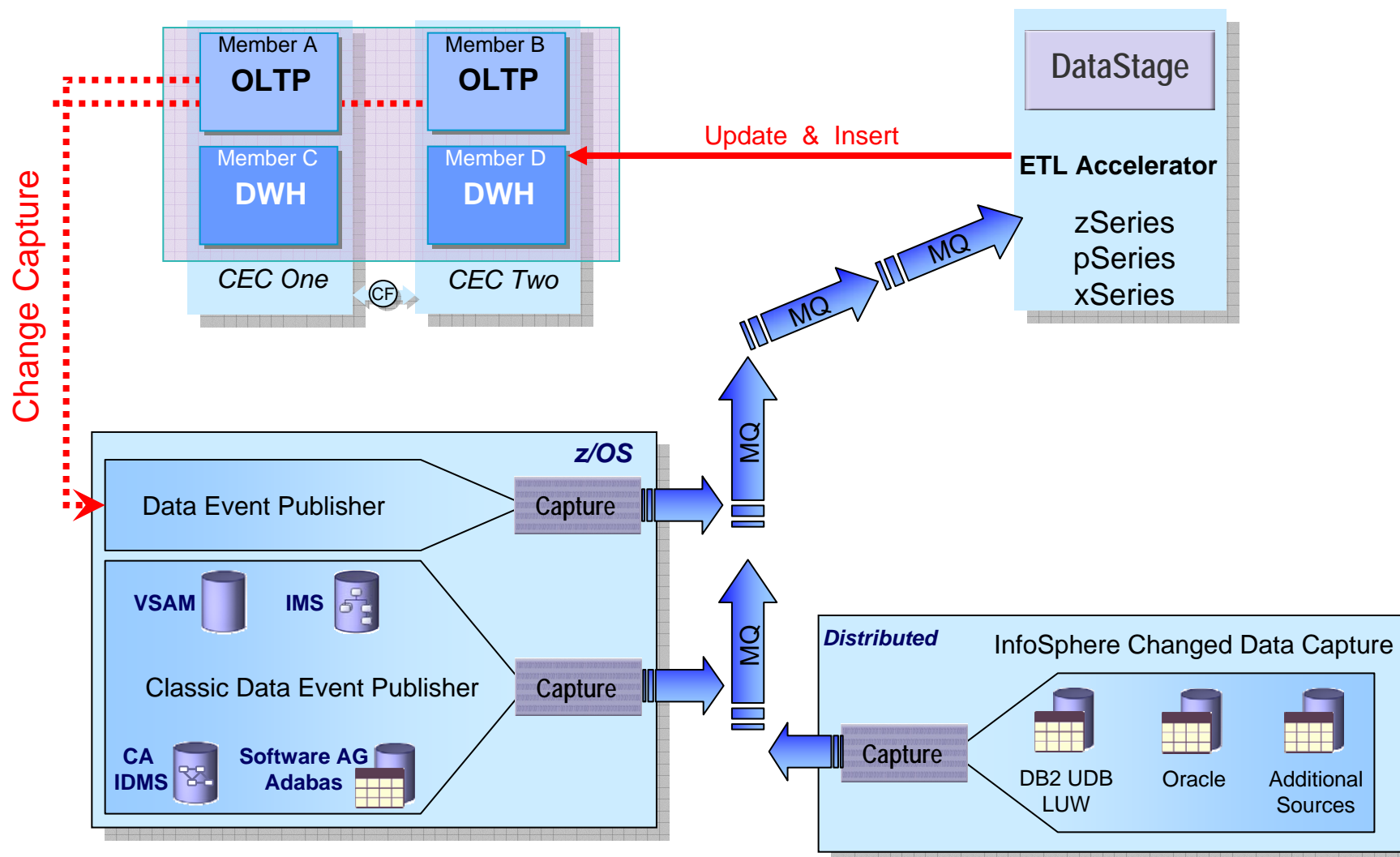# Maintenance : Incremental Updating is Essential

Member A
OLTP

Member B
OLTP

Member C
DWH

Member D
DWH

CEC One   CF   *CEC Two*

Change Capture

DataStage

Update & Insert

ETL Accelerator

zSeries
pSeries
xSeries

MQ   MQ   MQ

z/OS

Data Event Publisher   Capture

VSAM   **IMS**

Classic Data Event Publisher   Capture

CA IDMS   **Software AG Adabas**

MQ

MQ

Distributed   InfoSphere Changed Data Capture

Capture

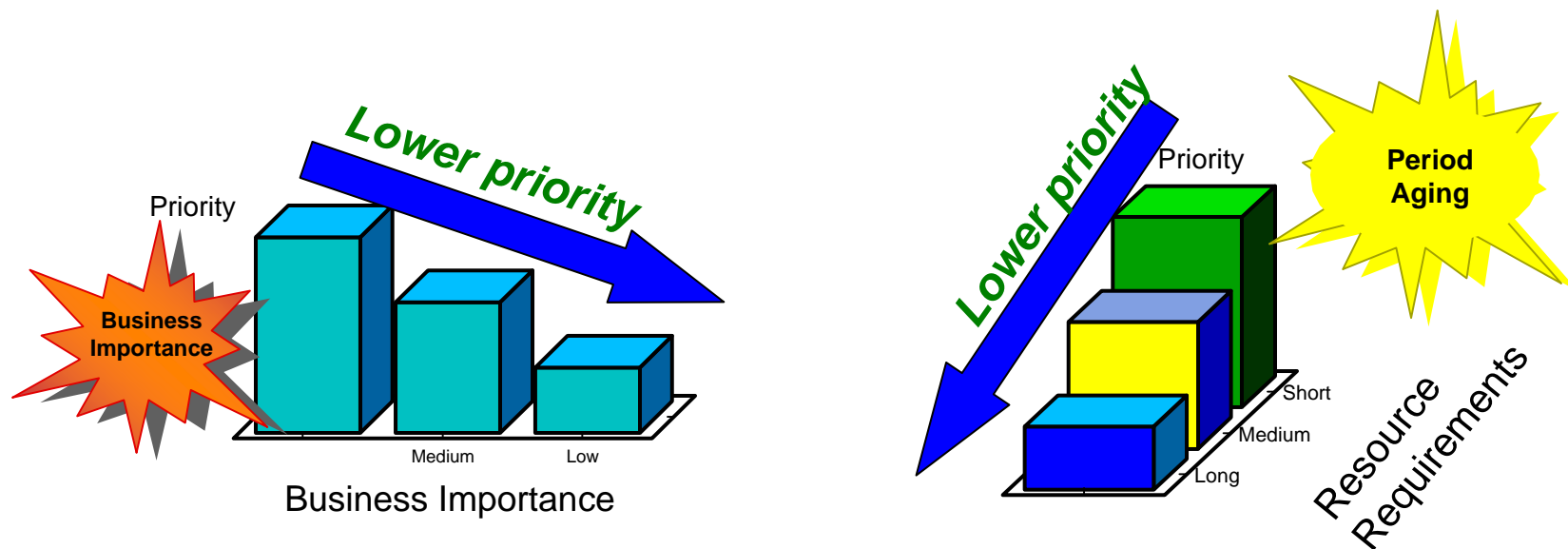DB2 UDB LUW   Oracle   Additional Sources

Workload Management

- Traditional workload management approach:

  - Screen queries before they start execution

  - Time consuming for DBAs.

  - Some large queries slip through the crack.

  - Running these queries degrade system performance.

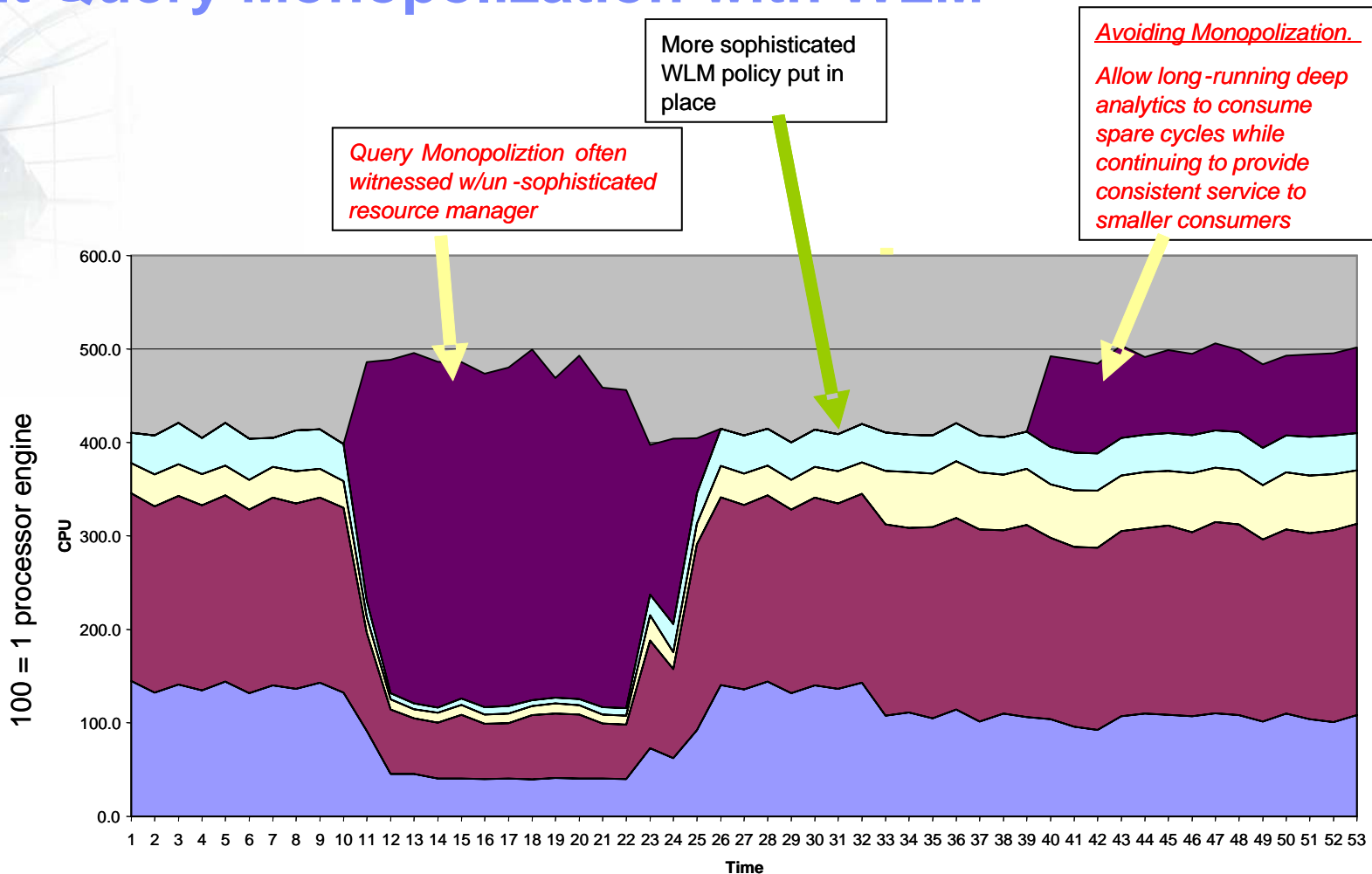 - Cancellation of the queries wastes CPU cycles.

Workload Manager

Work is managed according to it's Business Importance and Goals.

- You can prioritize work based on
  - Business requirements
  - Resource requirements – by allowing system to automatically lower priority of queries as they consume system resources: period aging
  - Limits the impact of unexpected long running queries

Lower priority

Priority

Business Importance

Medium   Low

Business Importance

Lower priority

Priority

Period Aging

Short

Medium

Long

Resource Requirements

Prevent Query Monopolization with WLM

More sophisticated WLM policy put in place

Query Monopoliztion often witnessed w/un-sophisticated resource manager

Avoiding Monopolization.

Allow long-running deep analytics to consume spare cycles while continuing to provide consistent service to smaller consumers

When you have unexpected spikes in workload, will you be able to handle it?

… Or do you have to micro-manage the system ?

- Business Analyst Large
- Business Analyst Medium
- Business Analyst Small
- Opertnl BI
- Data-Mining_Q

IBM zEnterprise 196: The heart of the new machine
*The industry's **fastest and most scalable** enterprise system*

Infrastructure

5.2 GHz superscalar processor

Up to 96 Cores, 1 to 80 configurable for client use

Up to 3 TB RAIM memory

Over 100 new instructions

1.5MB L2 Cache per core, 24MB L3 Cache per processor chip

Cryptographic enhancements

Optional water cooling

Dramatic improvement over IBM System z10™:

For Linux

Up to **60%**

Improvement in performance

for **35%**

Less cost

For z/OS

Up to **40%**

Improvement in performance

with **60%**

More capacity

- **With no increase in energy consumption**
- **And even better performance with new software**

IBM Smart Analytics Optimizer

Capitalizing on the best of relational and columnar databases

Data Warehousing

What is it?

The IBM Smart Analytics Optimizer is a workload optimized, appliance-like add-on that enables the integration of business insights into operational processes to drive winning strategies. It accelerates select queries, with unprecedented response times.

How is it different

- **Performance:** Unprecedented response times to enable 'train of thought' analyses frequently blocked by poor query performance.

- **Integration:** Connects to DB2 through deep integration, providing transparency to all applications.

- **Self-managed workloads:** queries are executed in the most efficient way

- **Transparency:** applications connected to DB2 are entirely unaware of the accelerator

- **Simplified administration:** appliance-like hands-free operations, eliminating many database tuning tasks

Breakthrough Technology Enabling New Opportunities

IBM InfoSphere Warehouse

- Adds core DW and analytics capability to DB2 for z/OS
- Advanced physical database modeling and design
- In-database data movement and manipulation capabilities
- Multidimensional reporting and analysis of data with Cubing Services
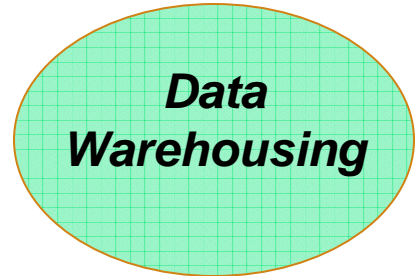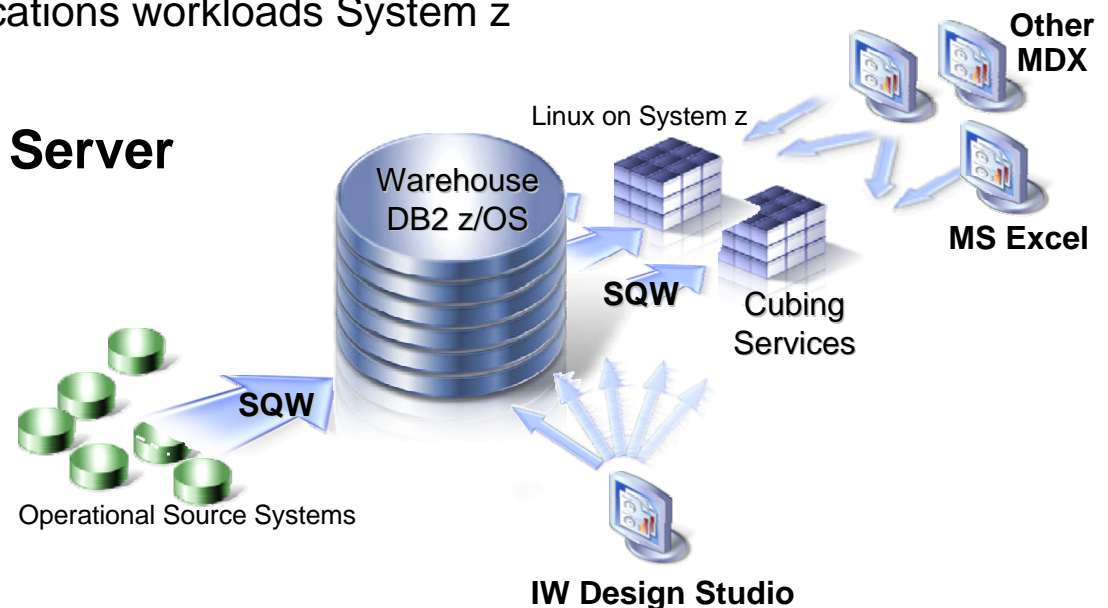
IBM DB2 Value Unit Edition

- Offers the same robust DB2 for z/OS data server at a one-time charge price
- Available for eligible net new applications workloads System z

IBM InfoSphere Information Server

- Profiles, cleanses and integrates information from heterogeneous sources to drive greater business insight faster, at lower cost

Data Warehousing

Other MDX

Linux on System z

Warehouse DB2 z/OS

MS Excel

SQW

Cubing Services

SQW

Operational Source Systems

IW Design Studio

IBM Cognos Business Intelligence for Linux on System z

Business Analytics

- **Full range of BI capabilities**
  - Query, reporting, analysis, dashboarding, realtime monitoring

- **Delivers information where, when and how it is needed**
  - Self-service reporting and analysis
  - Automated delivery of information in context
  - Author once, consume anywhere

- **Purpose-built SOA platform**
  that fits client environments and scales easily

• IBM SPSS

• Full breadth of predictive analytics

- Data collection, statistics, data mining, predictive modeling, deployment services…

• Putting prediction in hands of the business

- Decision Management

• Driving better business outcomes

- Attract and retain more profitable customers
- Detect and prevent fraud
- Improve resource allocation

Business Analytics

PREDICTIVE ANALYTICS SOFTWARE

Publish • Automate

DEPLOYMENT   MODELING

DATA COLLECTION   STATISTICS

Share • Protect
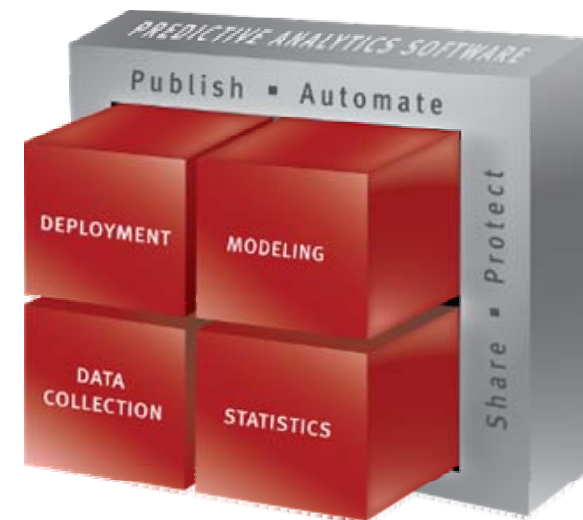
Some key Redbooks

- **Enterprise Data Warehousing with DB2 9 for z/OS**
  - http://www.redbooks.ibm.com/abstracts/sg247637.html

- **Co-Locating Transactional and Data Warehouse Workloads on System z**
  - http:// www.redbooks.ibm.com/abstracts/sg247726.html

- **DB2 for z/OS: Data Sharing in a Nutshell**
  - http:// www.redbooks.ibm.com/abstracts/sg247322.html

- **System Programmer's Guide To: Workload Manager**
  - http:// www.redbooks.ibm.com/abstracts/sg246472.html

- **Application Design for High Performance and Availability**
  - http:// www.redbooks.ibm.com/abstracts/sg247134.html

- **Workload Management for DB2 Data Warehouse, REDP-3927**
  - http:// www.redbooks.ibm.com/abstracts/redp3927.html

- **DB2 9 for z/OS Performance Topics**
  - http:// www.redbooks.ibm.com/abstracts/sg247473.html

Thank You

IBM Professional Services

Accelerated Success

- New "how-to" books deliver expertise
  - BI Strategy Book
  - BI on Cloud
  - BI Redbook

- Workshops to help shape strategy
  - Champion workshops
  - Business Analytics experience

- Services & Training
  - Proven practice workshops, learning assessment and user adoption services
  - Broader portfolio of self-paced training options

- Growth in communities
  - Innovation Center
  - DeveloperWorks, C^3 Blog
  - Twitter, facebook and linked-in

Miami-Dade County
Selects IBM System z platform to expand their IBM Cognos 8 BI enterprise infrastructure

"

*…We are now **able to expand the usage of our Business Intelligence reporting.** By the end of 2010, we will have **users from over 42 County departments with over 1500 users creating and consuming reports with stable environments on System z.***

—Jaci Newmark, Project Lead, Enterprise Business Intelligence Architecture,

Miami-Dade County

"

✓**11 days to go from distributed to System z deployment model**

✓**Consolidated multiple BI deployments onto a single platform**

✓**Consolidate multiple, disparate data sources onto a single platform**

✓**Ensured 99.999% availability & complete disaster recovery plan**

University of North Carolina Health Care
Deploys a hybrid data warehouse solution combining the strengths of InfoSphere and DB2 software on System z and System p platforms

" *With the deployment of the Carolina Data Warehouse for Health, we have been able to <u>increase the timeliness of the information</u> available to our researchers, staff and physicians," "Because the system can also support general queries that relate to the diagnosis and treatment of a wide array of patients, we are now able to <u>make more intelligent decisions</u> leading to improved patient care."* "

Donald Spencer, MD, MBA,
Associate Director of Medical Informatics, UNC Health Care

Numius' Client Success Story
Uses DB2 for z/OS and Cognos 8 BI for Linux on System z

✓Produced 400 reports in the same time as 1 report in the old environment

    ✓400 reports ran in 45 minutes as opposed to 1 report in 46 minutes

    ✓Easily supported 130 concurrent users, as opposed to 8 on the source environment

✓No reports timed out, not one user was rejected. Even when the system slowed down, it remained stable

✓The client would not need to redesign his application to achieve his objective of reaching out to a large community

✓This solution helps improve waste control in Belgium by informing producers, consumers and authorities faster and more thoroughly

*Jo Coutuer, Senior Business Intelligence Architect, Numius*

IBM Cognos BI Total Cost of Ownership Study

Explores the TCO of choosing an x86 based infrastructure vs. System z for a Cognos 8 BI deployment using proven IBM TCO measurement methodology

✓Average savings over 5 years of with a System z deployment: 36%

✓Reduction in high availability costs with System z: 50%

✓System administration savings alone pay for System z investment.

IBM
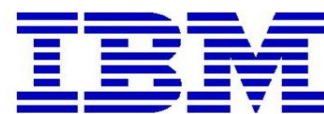
Blue Insight,
The IBM internal Private Analytics Cloud

> " *Our commitment to informed decision making led us to consider private cloud delivery of Cognos via System z, which is the enabling foundation that makes possible* **+$25M savings over 5 years***.* "

-IBM CIO Office

✓Consolidated 115 multi-product, departmental BI deployments to 1 Cognos 8 BI on System z

✓Support for our global workforce (2009: 72K, 2010: 130K, 2011: 200K)

✓Realizing value from +60 data sources across IBM

✓Projected $25M in savings (60% Consolidation, 35% Standardization, 5% Automation

Industry Analysts Agree

"In short, I believe that mainframes are the most modern platforms available in the commercial marketplace today."

Source: Clabby Analytics, Migrate From Mainframe? To What?, Joe Clabby, June 24, 2010

"Companies that buy into outdated hype about its complexity fail to realize the potential performance gains associated with mainframe use."

Source: Aberdeen Group, The Fable of Mainframe Complexity, Max Gladstone,

May 5, 2010

"Clients wishing to evolve their legacy application portfolios into more-modern technologies and architectures can do so on the IBM mainframe."

Source: Gartner, Mainframe Modernization: When the Platform Is the Solution,

Dale Vecchio, 8 January 2010

"The mainframe has long been recognized as a platform with an enviable reputation for reliability, security, and efficiency, Ovum considers that IBM by exploiting these characteristics has produced the next generation of data centre and cloud centre management platforms."

Source: Roy Illsley, Principal Analyst, Ovum

Typical Utilization for Servers

Windows: 5-10%   Unix: 10-20%   **System z: 85-100%**

System z can help **reduce** your floor space
up to **75%-85%** in the data center

For additional information please visit
www.ibm.com/software/data/businessintelligence/systemz/

System z can lower your total cost of ownership, requiring **as little as 30%**
of the power of a distributed server farm running equivalent workloads

The cost of storage is typically **three times more** in
distributed environments

IBM Product Portfolio
Business Analytics and Data Warehousing on System z

Business Intelligence

- Cognos 10 Business Intelligence

Predictive Analytics

- SPSS Statistics 19

- SPSS Modeler

- SPSS Collaboration and Deployment Services

Data Warehousing

- DB2 for z/OS VUE (Value Unit Edition)

- InfoSphere Warehouse

- Smart Analytics Optimizer

Solution Edition for Data Warehousing (pricing option)

Data Integration and Movement

- InfoSphere Information Server
- InfoSphere Change Data Capture
- InfoSphere Replication
- InfoSphere Federation
- Global Name Recognition

Master Data Management

- InfoSphere Master Data Management Server

InfoSphere Industry Models

- Banking, Insurance, Retail. Telco, Heath Payor, Heath Provider, Financial Markets

Flexible Deployment Options

- IBM Smart Analytics System 9600
- IBM Smart Analytics Cloud
- IBM Services