

**IBM Information**  
>>> On Demand **2007**



## Best Practices for Dynamic Warehousing on System z

*Jaime F. Anaya, DB2 for z/OS solutions Architect, IBM,  
janaya@us.ibm.com*

*Burton Corwin, Program Manager for Strategic DB2  
z/OS initiatives, IBM, bdcorwi@us.ibm.com*

*TDW – 1022A*



**Act.Right.Now.**

*IBM INFORMATION ON DEMAND 2007  
October 14 - 19, 2007  
Mandalay Bay  
Las Vegas, Nevada*

# Disclaimer

The information contained in this presentation has not been submitted to any formal IBM review and is distributed on an "As Is" basis without any warranty either expressed or implied. The use of this information is a customer responsibility.

The materials in this presentation are also subject to

- enhancements at some future date,
- a new release of DB2, or
- a Programming Temporary Fix (PTF)

*IBM MAY HAVE PATENTS OR PENDING PATENT APPLICATIONS COVERING SUBJECT MATTER IN THIS DOCUMENT. THE FURNISHING OF THIS DOCUMENT DOES NOT IMPLY GIVING LICENSE TO THESE PATENTS.*

TRADEMARKS: THE FOLLOWING TERMS ARE TRADEMARKS OR ® REGISTERED TRADEMARKS OF THE IBM CORPORATION IN THE UNITED STATES AND/OR OTHER COUNTRIES: AIX, AS/400, DATABASE 2, DB2, e-business logo, Enterprise Storage Server, ESCON, FICON, OS/390, OS/400, ES/9000, MVS/ESA, Netfinity, RISC, RISC SYSTEM/6000, iSeries, pSeries, xSeries, SYSTEM/390, IBM, Lotus, NOTES, WebSphere, z/Architecture, z/OS, zSeries, System z.

*THE FOLLOWING TERMS ARE TRADEMARKS OR REGISTERED TRADEMARKS OF THE MICROSOFT CORPORATION IN THE UNITED STATES AND/OR OTHER COUNTRIES: MICROSOFT, WINDOWS, WINDOWS NT, ODBC and WINDOWS 95.*

***For additional information visit the URL***

***<http://www.ibm.com/legal/copytrade.phtml> for "Copyright and trademark information"***

1



IBM INFORMATION ON DEMAND 2007

**Act Right. Now.**

## Agenda

- IBM Dynamic Warehousing
- Why Warehousing on System z and DB2 z/OS
- What's Available on System z to Support DW
- z-centric BI solution architecture
- Some Best Practice Recommendations
- Summary
- Q & A

2



IBM INFORMATION ON DEMAND 2007

**Act Right. Now.**

# IBM Data Servers

## **Reduce cost of deployment and management of data**

- *Innovation to reduce the cost of infrastructure*
- *Innovation to manage the lifecycle of data - from modeling and design through change management and sunsetting*

## **Enable rapid use of data throughout the enterprise**

- *Innovation that accelerates SOA and XML initiatives*
- *Innovation that leverages Web 2.0 and situational applications*

3



IBM INFORMATION ON DEMAND 2007

**Act Right. Now.**


## Dynamic Warehousing

### A New Approach to Leveraging Information

Information On Demand  
to Optimize Real-Time  
Processes

OLAP & Data Mining  
to Understand Why and  
Recommend Future Action


Query & Reporting  
to Understand  
What Happened



Action




Analysis



Reporting

**Dynamic  
Warehousing**

**Traditional Data  
Warehousing**



IBM INFORMATION ON DEMAND 2007

**Act Right. Now.**

Traditional warehousing focused on query and reporting to understand what happened, and evolved to enable OLAP and data mining to understand the why those things happened and recommend future action.

[click for transition]

Dynamic warehousing is a new approach to address the primary business challenges organizations face today, which requires the ability to deliver the right information to the right people at the right time to more effectively leverage information and enable more effective business decisions. It's about information on demand to optimize real-time processes.

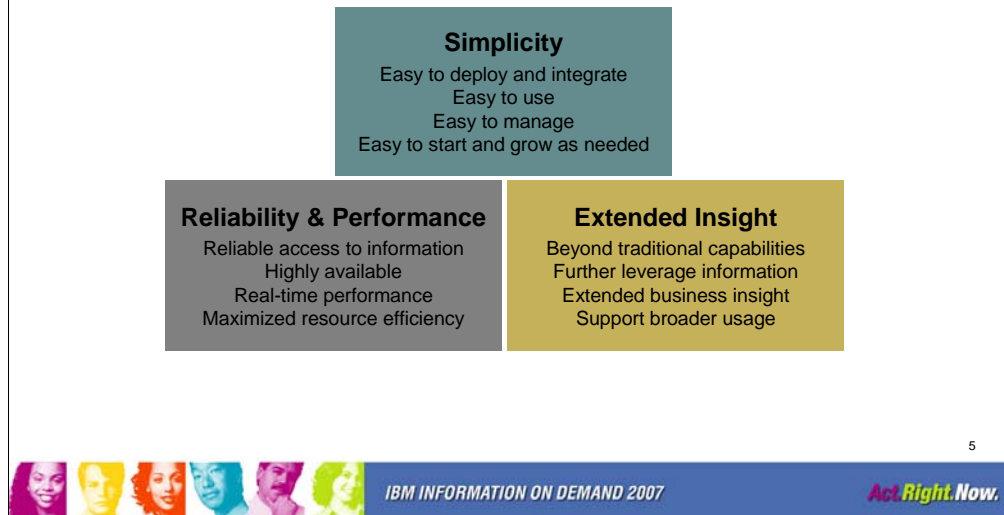
[click for transition]

And Dynamic Warehousing requires four key things:

1. Support for real-time access to aggregated, cleansed information, which can be delivered in the context of the activities and processes being performed;
2. Analytics that can be leveraged as part of a business process;
3. The ability to incorporate knowledge from unstructured information; and
4. A complete set of integrated capabilities that extend beyond the warehouse to enable Information on Demand

## Warehousing strategic pillars

### *Guiding principles for innovation*



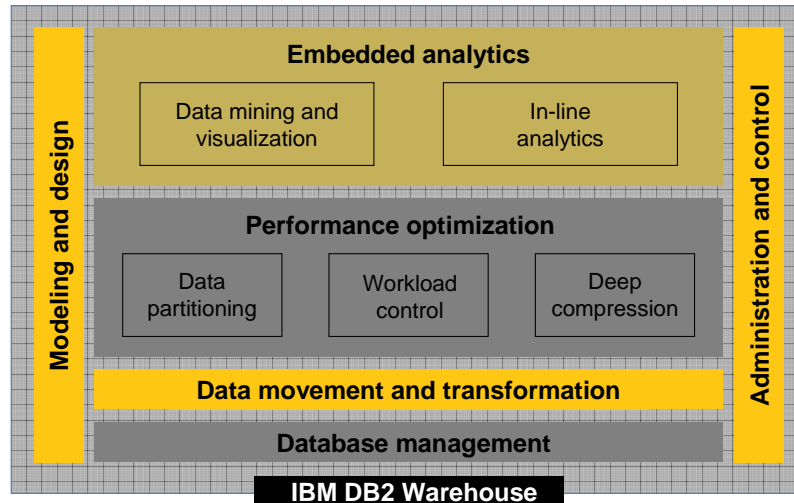
So, before we jump into the different components of IBM DB2 Warehouse, it's important to understand the guiding principles that drive what goes into our warehousing solutions. IBM has a set of core values, such as "Dedication to every client's success," which drive our actions as an organization. Likewise, we have identified a set of strategic pillars that we use as guiding principles when deciding what features to add to our warehousing offerings and how to innovate. These revolve around:

- *Simplicity*—to make our solutions as easy to use as possible;
- *Reliability and performance*—to provide efficient, reliable, highly available access to information for historical analysis OR operational purposes, with real time performance requirements, while maximizing resource utilization; and
- *Extended insight*—going beyond traditional warehousing capabilities to enable our customers to get more value out of their information by providing built in capabilities for generating greater business insights and supporting broader usage of the warehouse across the organization.

Every feature we add and each new offering we create is driven by one or more of these guiding principles.

## IBM DB2 Warehouse software

*A complete, integrated platform*



6



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

Our goal with the DB2 Warehouse platform is to provide a complete set of tightly integrated services that meet the objectives of our guiding principles. Customers can leverage all of these services together or selectively implement desired options for ultimate flexibility.

It starts out with DB2 9, a highly scalable, enterprise class database that can address the performance characteristics of any type of application. We also provide out-of-the-box data movement and transformation capabilities to reduce the complexity and lower the costs typically associated with loading data into the warehouse and preparing that data so that it can be leveraged more effectively. We then add a set of performance optimization features that enable the warehouse to address broader enterprise requirements. This includes Database Partitioning for dedicated warehousing that can scale linearly, Workload Control features for prioritizing queries to ensure that the most critical applications are serviced accordingly, and Deep Compression to increase efficiency of the warehouse and reduce storage costs.

We then extend the value of the warehouse by including analytics capabilities that can be delivered inline within applications or made available as a service, along with data mining and visualization to provide more dynamic business insight. These services are embedded in the warehouse to provide better performance, increased efficiencies and reduced costs. Finally, we provide a common set of integrated tools for data modeling and design, administration and control of the warehouse and all related services, making them easier to use and manage.

## Operational BI Definition & Examples

- Major BI Growth Area is Embedded Analytics
  - Deliver BI to customer facing humans and applications
  - Integrate BI Components with Operational Systems & Information Portals
  - Looks like a “sweet spot” for DW on DB2 z/OS
- Customer Service/ Self Service Examples
  - VZ Customer Service
  - Credit Card Self Service Reporting



IBM INFORMATION ON DEMAND 2007

Act Right.Now.

7

### Major BI Growth Area is Operational Business Intelligence.

The market place is moving in the direction of Operational BI, moving Information Based Decision Making down to the operational level (Sales, Customer Service, etc). BI has traditionally been for Analysts to Executives. This is a completely new growth area. This area has been targeted by Microsoft, Information Builders, Business Objects, Cognos, Microstrategy and Oracle.

### Sweet spot for System z

Operational Reporting requires massive transaction scalability. Customer facing staff number in the thousands. Kraft foods has 7,000 sales people, Verizon has 10,000 customer service representatives. The information Operational BI systems deliver require aggregating hundreds to thousands of records as opposed to millions or billions of records for strategic BI systems.

### Embedded Analytics

Customer Service level applications require efficiency.

Improving efficiency leads to a return on an investment.

The best way to improve efficiency is to reduce the number of screens and clicks a customer service representative requires to access information.

The best way to accomplish the above is to “embed” the Business Intelligence information within an operational application.

Alphablox is the best tool to accomplish that goal.

This has been proven by my customers



## Delivering Operational BI on System z

- Many core business Operational Systems & Information Portals store data on System z
  - AlphaBlox uniquely provides the technology to deliver embedded analytics with these applications
  - Embedded analytics depends on historic and semi-aggregated data stored in a data warehouse
  - Advantages of co-locating operational & warehouse data can drive DW on z
  - Qualities of Service required by the Operational Systems & Information Portals can also be provided by a companion DW on z
- Sweet Spot for AlphaBlox & DW on DB2 z/OS together
  - Requires massive transaction scalability (z strength)
  - Requires semi-aggregated data, i.e. data aggregated at a low level (z strength)
  - System z provides the OLTP Systems & DW; AlphaBlox provides the BI Components

8



IBM INFORMATION ON DEMAND 2007

Act Right.Now.

Operational BI, and specifically Embedded Analytics, can improve Customer Sales and Service efficiency.

Revenue generation and efficiency improvements (cutting head count) are the only selling arguments to banks today. (The same is likely true to other large organization in the financial services and utilities industries.)

Implementing these solutions will also drive Master Data Management solutions like Websphere Customer Center.

Implementing any of these solutions on System z will drive further MSU usage.

It all depends upon driving efficiency!

Embedded Analytics (Alphablox), not just operational BI, is the best way to improve efficiency, in this respect it becomes easy to sell.

We would be selling to the same people who are manage the operational systems and are concerned about customer service efficiency.

Embedded Analytics can range from simple semi-aggregated historic data to predictive mining to support cross sell and up-sell opportunities.



## **Why Warehousing on System z and DB2 z/OS**

**Act.Right.Now.**



Next slide please.

Now that we know about dynamic warehousing, let move on to a related by slightly different topic – how to benefit from a consolidated and/or modernization of your warehouses and data marts to achieve near real-time requirements, single enterprise wide “version of the truth”, best of industry security and regulatory compliance

## Data Warehousing on z/OS Drivers

- **System z customers are asking for increasing support and capabilities for DWH on z**
  - Protects their investment in System z
  - Significant amount of enterprise data is on System z
- **Many z customers have a DWH on DB2 z/OS**
- **Demand for embedded analytics in System z based operational applications is growing**
- **The BI operational needs for reliability, availability and security are growing.**
- **Significant DW capabilities added to DB2 V8 with**

10



**even more DB2 9**

IBM INFORMATION ON DEMAND 2007

**Act Right. Now.**

## Value proposition of DW on System z

- **Qualities of Service**
  - Superior Quality
  - Continuous Availability
  - Security and Regulatory Compliance
  - Scalability
  - Backup and recovery
- **Operational data and the ODS together means**
  - Reduced complexity
  - Reduced cost
  - Shared processes, tools, procedures
  - Streamlined compliance and security
- **Positioned for the future**
  - Web-based applications
  - XML support
  - Service Oriented Architecture (SOA)
- **zIIP and other specialty engines improve Total Cost of Ownership**
- **Better leverage System z skills and investment**

11



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

In summary, on slide 40,. I want to the emphasize key DB2 z/OS and System z differentiators which make them ideally suited for Dynamic Warehousing.

- First is the qualities of service: System z and DB2 for z/OS provide the highest levels of availability, resiliency, security, and recovery at very competitive cost of ownership. With BI and warehousing workloads moving from the back office to the front office, those System z strengths make it a very attractive platform.
- Remember also that by keeping your operational data and warehouse together in System z, you can streamline compliance and security, reduce complexity and reduce the overall cost of your BI solution.
- Even further savings can be achieved by leveraging your existing System z skills and investment. If you are expanding your warehouse or creating a new one, consider that it may be handled with existing people and systems or with minimal increases as opposed to having to develop new skills and deploy new systems.
- Finally, your investment in DB2 z/OS is protected as it is very well positioned for the future with support for web-based applications, XML, and Service Oriented Architecture.



**What** is available on System z to support and run your Data Warehouse

*Act.Right.Now.*



# DB2 Warehouse Capabilities that Exist Today for System z

- **Database Management**
  - **DB2 z/OS V8 & V9 EDW Base**
    - Functional and performance enhancements
    - DB2 family compatibility – including MQTs
- **Data Movement and Transformation**
  - WS DataStage EE, WS II Classic Federation, WS Classic Event Publishers, Distributed DBMS Event Publishers, BatchPipes for OS/390, DB2 Unload/Load Utilities
  - Information Server (WS Information Analyzer, WS QualityStage)
- **Performance Optimization**
  - DataPartitioning: Range Partitioning, Data Sharing
  - Workload Control: WLM, IBM Director
  - Deep Compression: Hardware/Software compression, Data Compression, Index Compression (V9)
- **Analyze/Report**
  - Alphablox, QMF, DataQuant
  - Partner offerings from Cognos, Hyperion, Business Objects, MicroStrategy, SAS, IBI
- **Modeling and Design**
  - Rational Data Architect
- **Administration and Control**
  - IBM Tivoli Omegamon XE for DB2 Performance Expert on z/OS, DB2 Query Monitor, Optimization Service Center (V9)

13



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

## Additional DW/BI Capabilities for System z

- **Performance Management**
  - IBM Tivoli Omegamon XE for DB2 Performance Expert on z/OS, DB2 Query Monitor, Optimization Service Center (V9)
- **Security and Regulatory Compliance**
  - DB2 Data Archive Expert, DB2 Test Database Generator, DB2 Audit Manager Expert, IBM Encryption for DB2 and IMS Databases
- **Application Management**
  - DB2 Table Editor, DB2 Web Query Tool, DB2 Automation Tool, DB2 Query Monitor
- **System z Specialty Engines**
  - zIIP, zAAP, IFL, ICF

14



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

## DB2 V8 enhancements benefiting DW

- Usability, Availability & Scalability
  - Online Partitioning Changes, More Partitions, Rotate
  - Schema Evolution:
  - VSCR with System z 64-bit Architecture
- Data Warehousing Support
  - Star Join Improvements
  - Materialized Query Tables
- Overall System & Query Performance
  - Locking Improvements
  - Multi-row INSERT & FETCH
  - More index Access for VARCHAR, data type mismatch, ...
  - DDF Performance Improvements
  - zIIP for improved cost of ownership

15



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

On slide 22, I list the enhancements to DB2 V8 that benefit dynamic warehousing.

Queries are improved a lot in V8. Optimization improvements provide a performance boost and make the job simpler. Improved optimization techniques like ability to use indexes more, star join and scale improvements allow reduced work for computers and for people. Enhanced data helps get the best access path. Visual Explain improves the ability to analyze and resolve any problems.

The many improvements for indexes, materialized query tables and partitioning can save space and add new options for improved performance and availability, even while simplifying the process. Not padded, clustering, longer and backward scans help indexes. Being able to add, rotate and rebalance partitions improve partitioning options. BI tools like QMF and DataQuant can build upon these strengths and add new function to reporting, dash boards, visualization. SQL enhancements on this page and the next improve portability of the SQL, improve the ability to express queries, and help with performance.



## DB2 9 enhancements benefiting DW

- Usability, Availability & Scalability
  - Universal & Partition by Growth Table Space
  - Clone Tables (Online Load Replace)
  - XML
- Data Warehousing Support
  - Star Join Improvements (Dynamic Index ANDing)
  - SQL Enhancements (RANK, DENSE\_RANK, Text, XML, INTERSECT, ...)
- Overall System & Query Performance
  - TRUNCATE Statement for fast delete, MERGE for insert or update
  - Index Compression
  - Significant CPU Reduction for DB2 Utilities & many queries
  - Insert and update scaling

16



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

On slide 23, I list the enhancements to DB2 V9 that benefit dynamic warehousing.

Today's complex applications include both transactions and reporting, so performing both well is imperative. The key improvements for reporting are optimization enhancements to improve query and reporting performance and ease of use. Improved data is provided for the optimizer, with improved algorithms and a rewritten approach to handling performance exceptions.

More queries can be expressed in SQL with new SQL enhancements. The set operators INTERSECT and EXCEPT clauses make SQL easier to write. OLAP extensions for RANK, DENSE\_RANK and ROW\_NUMBER add new capabilities. Other SQL statements improve consistency with the DBMS industry. V9 continues the progress in SQL, with many new functions, statements and clauses. The biggest changes are in XML. New SQL data manipulation statements are MERGE and TRUNCATE. New data types with DECIMAL FLOAT, BIGINT, BINARY and VARBINARY. Improvements in LOBs provide new function, more consistent handling and improved performance. Security is improved with ROLES and network trusted context. Data definition consistency and usability are improved. V9 is another big step in DB2 family consistency and in the ability to port applications to DB2 for z/OS.



# Hardware Specialty Engines

*Act.Right.Now.*



## DB2 & IBM zIIP can add value to database work

Portions of the following DB2 for z/OS V8 work may benefit from zIIP\*

- 1 - ERP, CRM, Business Intelligence or other enterprise applications
  - Via DRDA over a TCP/IP connection (enclave SRBs, not stored procedures or UDFs except call, commit & result set processing)



- 2 - Data warehousing applications\*
  - Requests that use parallel queries
- 3 - DB2 Utilities LOAD, REORG & REBUILD\*
  - DB2 utility functions used to maintain index structures
- 4 - DB2 9 for z/OS remote native SQL procedures

\* zIIP allows a program working with z/OS to have all or a portion of its enclave Service Request Block (SRB) work directed to zIIP. Above types of DB2 work are those running in enclave SRBs, of which portions can be sent to zIIPs



IBM INFORMATION ON DEMAND 2007

Act Right Now.

Allow me on slide 31, to discuss DB2 exploitation of zIIP specialty processor which brings excitement to warehousing on System z.

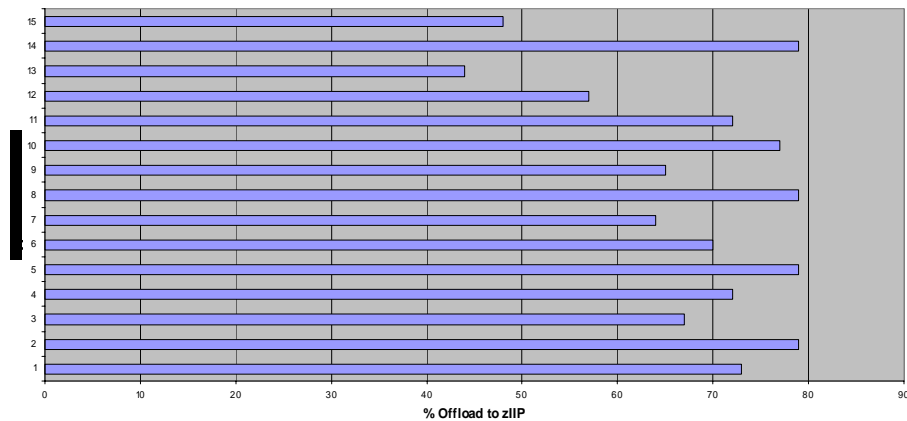
The zIIP is designed so that a program can work with z/OS to have all or a portion of its enclave Service Request Block (SRB) work directed to the zIIP. Not all of this work will be run on zIIP. z/OS will direct the work between the general processor and the zIIP.

IBM DB2 for z/OS version 8 is the first IBM software able to take advantage of the zIIP. Initially, the following workloads can benefit:

- SQL processing of DRDA network-connected applications over TCP/IP: These DRDA applications include ERP (e.g. SAP), CRM (Siebel), or business intelligence and are expected to provide the primary benefit to customers.
- Stored procedures and UDFs run under TCBs, so they are not generally eligible, except for the call, commit and result set processing.
- In V9, remote native SQL Procedure Language is eligible for zIIP processing.
- BI application query processing utilizing DB2 parallel query capabilities; and
- functions of specified DB2 utilities that perform index maintenance.

For more, see <http://www.ibm.com/systems/z/ziip/>

## Typical zIIP offload for a DW workload



19



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

Next on slide 32, you can see a graph which clearly shows that there is a wide range of query work that can be redirected to a zIIP. The queries shown are selected TPC-H queries executed on a 1TB warehouse.

The primary work for a data warehouse is the parallel processing.

- For large queries that are processed in parallel, as much as 80% of the work can be redirected to a zIIP. This should be most if not all of the costly BI queries.
- Some cases will be smaller, since part of the work will not be run in parallel.
- Small queries and the initial part of large queries (under 100 ms of cpu time) do not use the zIIP.
- If the work comes in remotely over TCP/IP and DRDA (not stored procedures), then the DRDA redirect can be used as well for the initial work.



# IBM BI Tools for System z

*Act.Right.Now.*



## Available since March 2007: IBM DataQuant

- provides a comprehensive query, reporting and data visualization platform for both web and workstation-based environments
- Together with over 100 built-in analytical functions, IBM DataQuant allows organizations to derive maximum value from their data and rapidly build and distribute comprehensive data visualization solutions across the enterprise.
- Visual Dashboards, Enhanced Graphical Reporting, Security and Personalization, SOA Layer, Enhanced Analytics



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

Going on to slide 26 ... I'll mention recently announced DataQuant.

**DataQuant for z/OS, V1.1 (5697-N64)** delivers a comprehensive query, reporting, and data visualization platform for both Web and workstation-based environments. While remaining compatible with the IBM Query Management Facility (QMF™) product line, DataQuant introduces a variety of powerful business intelligence capabilities, including:

- Powerful graphical reporting environment that allows rapid development and deployment of executive dashboards, information portals, and interactive data visualization solutions.
- Drag-and-drop development of OLAP analytics, SQL queries, tabular reports, graphical reports, pivot tables, and data analysis views.
- Dozens of charts, controls, and graphical primitives to visual reports, along with the ability to embed subqueries that provide supporting details.
- Full compatibility with QMF infrastructure and objects, including the ability to create, open, edit, delete, and save QMF queries, procedures, forms, and visual reports.
- Over 100 built-in mathematical and analytical functions accessible in both dashboards and visual solutions as well as in printable reports.

It also provides a rich security infrastructure, providing personalization of both the query and reporting environment as well as business intelligence content distributed within it. This infrastructure includes support for single sign-on and optionally interfaces with existing LDAP directories or internally defined DataQuant user directory.

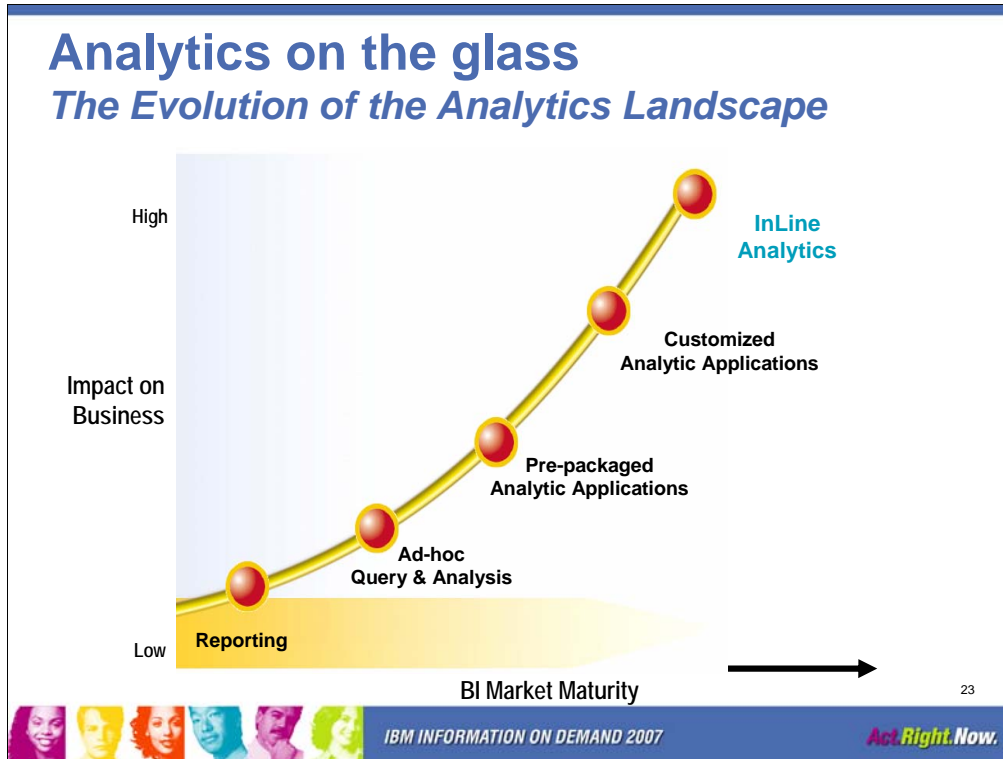
# What is Alphablox?

- Platform for Customized Analytic Applications and Inline Analytics
- Pre-built components (Blox) for analytic functionality
- Allows you to create **customized** analytic components that are **embedded** into existing business processes and web applications



IBM INFORMATION ON DEMAND 2007

Act Right. Now.



An off-to-the-side tools approach to BI has had limited success

In contrast, analytics integrated into business processes are demonstrating real ROI

This has been our belief since 1996

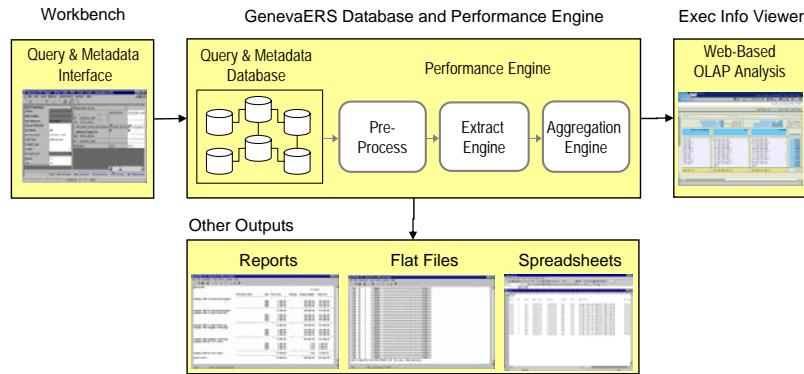
The challenge we saw was how to effectively and efficiently imbed analytics into the applications that drive critical business processes

We and our customers believe our infrastructure approach to analytics uniquely accomplishes this



# GenevaERS: Harnessing Economies of Scale for Enterprise Reporting

GenevaERS is a high-performance data scanning tool for z/OS that enables analysis procedures not practical with standard tools. GenevaERS is a combination business intelligence application and database query engine. It efficiently uses available memory, I/O devices and compute capacity to resolve multiple business problems simultaneously in a single pass through a database or data repository. For certain problems, GenevaERS can offer performance up to 100 times greater than other products.



24

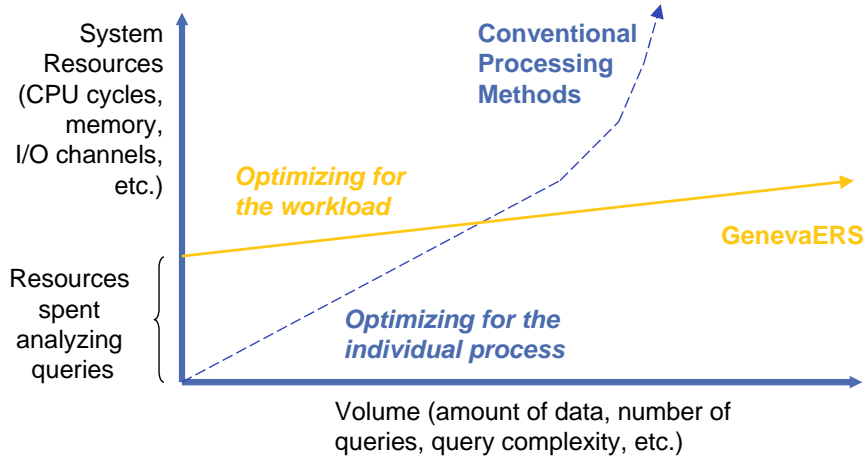


IBM INFORMATION ON DEMAND 2007

Act Right. Now.

# Solution for the Scaling Problem

GenevaERS's aggregation of queries allows it to harness economies of scale to yield efficiencies unmatched by conventional methods.



25



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

## Dealing with the Shrinking Batch Window

*When performance problems are caused by:*

- High query volumes - GenevaERS responds with:
  - Single pass architecture – processing multiple queries in a single scan, avoiding the use of indexes
- High table lookup volumes - GenevaERS responds by:
  - Exploiting the 64-bit addressing inherent in z/Architecture and sharing in-memory tables among multiple queries, thus leveraging the SMP architecture of the mainframe
- High data volumes – GenevaERS responds with:
  - Generated machine code, tailored to the specific problem at hand
  - Code highly optimized for System z, with short instruction path lengths
  - Efficient I/O for both disk and tape
  - Parallel processing
  - In-memory summarization of data
  - Data piped from one transform to the next, minimizing I/O

26



IBM INFORMATION ON DEMAND 2007

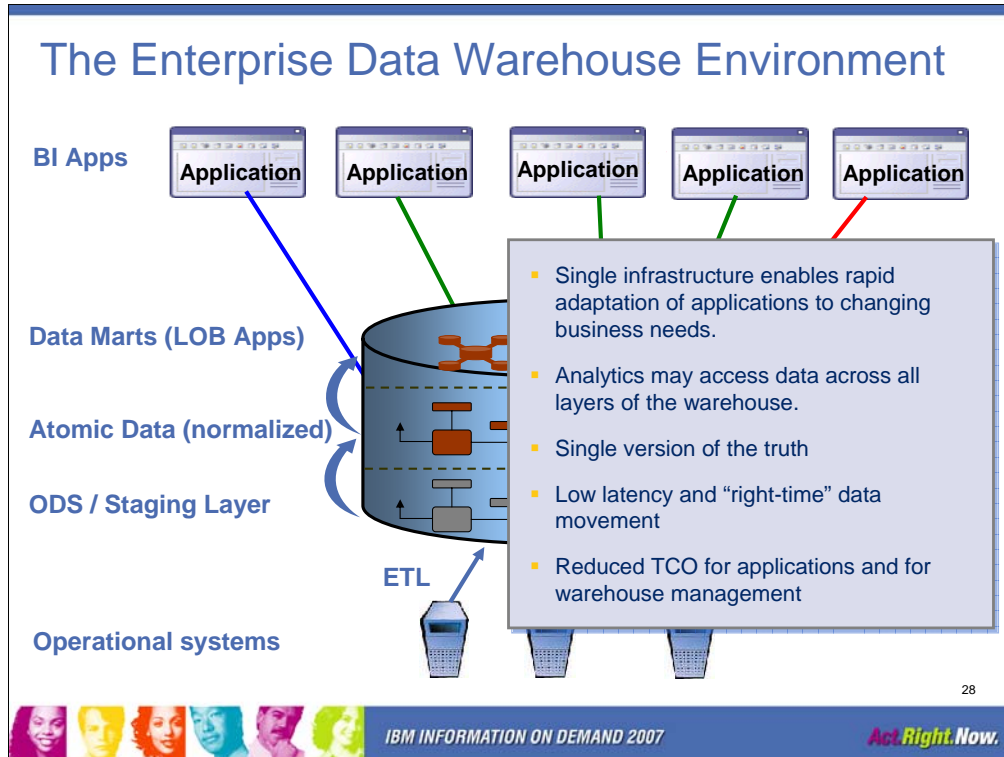
**Act Right. Now.**



# BI Solution Architecture

*Act.Right.Now.*





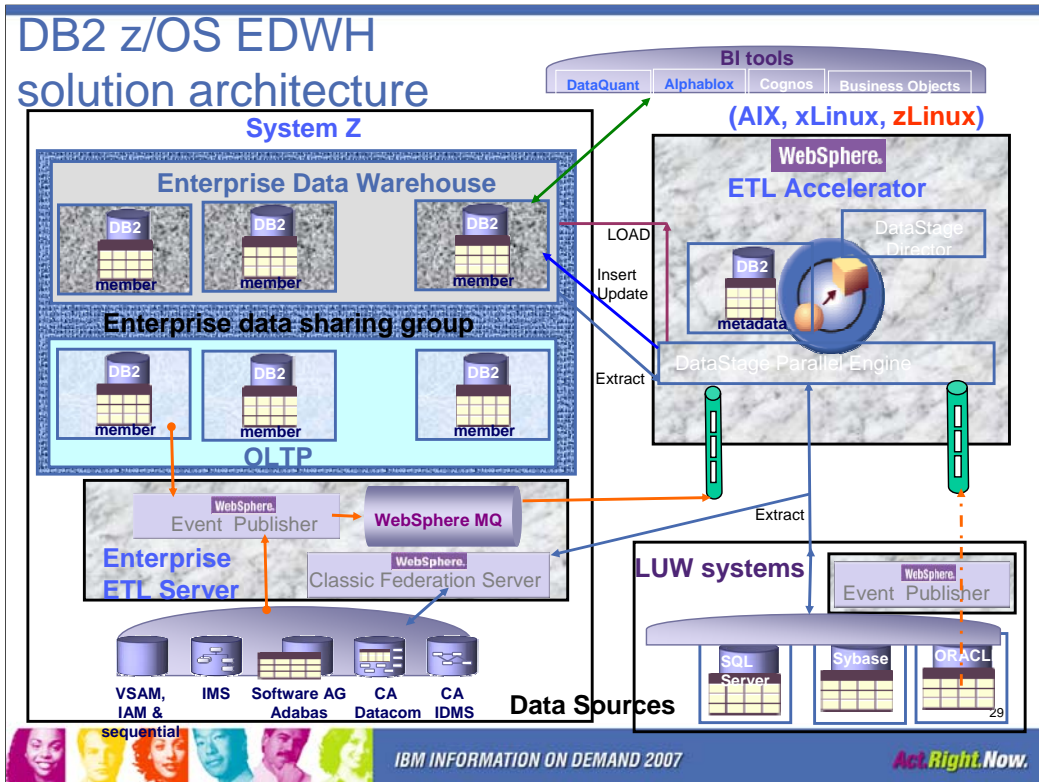
The enterprise data warehouse simplifies the architecture by consolidating multiple layers within the data warehouse itself. Operational data stores, the normalized data warehouse, and the logical marts are all architected to be part of the consolidated data warehouse. This simplifies the interchange between these layers and enables the possibility of right data loading of data. It simplifies access for application, as they now have one single place to go for information.

Because the information is consolidated in the warehouse under a single owner, you have better coordination and visibility across and within the layers of the warehouse. The result: a single version of the truth and reduced TCO for applications and overall management.

Common Questions:

How do you provide a single normalized schema that will support multiple business applications?

Answer: you don't have to. One of the advantages of consolidating the warehouse is that it makes it much easier to create and maintain offshoots or variants. They can be separate schema instances, or views on the current instance. Because it's all within the same logical warehouse instance, maintenance becomes easier.



On to slide 34 then. Here you see all the different components coming together.

## EDW

Highest Qualities of Server: highly available, best security,  
Single DB system view with OLTP

It exploits the flexible yet high performance layered DWH architecture

Flexible logical data marts (views, partitions, online DDL) with performance objects and the flexibility to convert to physical data marts for even more performance once flexibility is not as important

Performance objects (MQT, user aggregates, parallelism)

## ETL Accelerator

Single integration metadata store

Single event manager

Flexible transformations

Parallel engine

Near real time data streaming and/or high performance bulk transfers

Extensive data sources on z/OS and distributed systems

Modern BI reporting and Analytics

## DB2 Isolation of OLTP and DWH

Proposed DB2 ZPARM Configuration differs between an OLTP and a DWH. This makes it necessary to isolate the workloads from each other.

- **Proposed solution, a DSG environment:**  
If the customer already has a Sysplex/Data Sharing Environment, the DWH is moved to new members. Recommendation for availability is of course to have at least two members in different CECs.
- **Solution in different LPAR – allows different OS levels:**  
The customer doesn't have and/or doesn't want a data sharing environment. In that case the DWH can be in a different LPAR with its own OS version and DB2 subsystem. This allows resource assignments/control on LPAR level and mixed OS versions.
- **Solution within the same LPAR:**  
Two different DB2 subsystems for OLTP and DWH workload within the same LPAR. This allows workload specific ZPARM configurations.

→ versus argument: One system for all different Workloads...



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

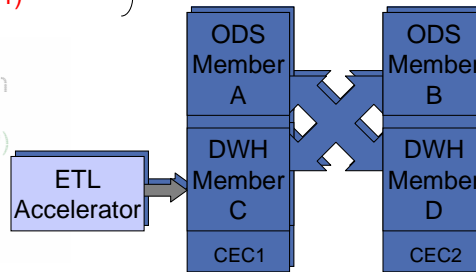
# Proposed solution, a DSG environment

- Having the DWH members within the same DSG as the operational database:
  - + Single System Image allows easy adding of further members
  - + Single view of System (Backup, DR)
  - Possible Catalog contention (for analytics every user should have an own database to reduce contention on DSNDB01)
  - More complex DB administration

## Recommended solution

- Must isolate buffer pools
- Different DB for
  - OLTP
  - DWH
  - Each analytical user

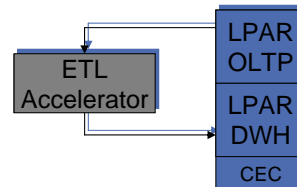
- Having the DWH members within different DSG than the ODS
  - + Better isolation (no common GBP)
  - More effort
  - Loosing single view of System





## Solution in different LPAR – allows different OS levels

- + Managing resources at LPAR level is easier
- + Min/Max assign per LPAR (IRD)
- + Assign more zIIP to DWH and more CP to OLTP
- + Good isolation (low risk for OLTP)
- + Allows different OS levels
- Loose single view of DB system
  - Loose single backup
  - Loose single DR
- No HA for OLTP or DWH

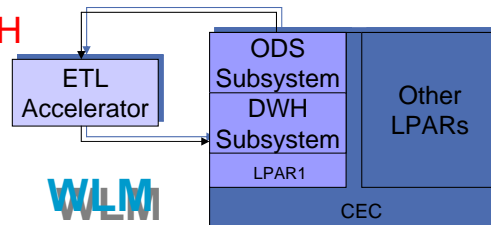


32



## Solution within the same LPAR

- + Still different ZPARM and BP configuration
- + OLTP can be optimized for its predicted workload
- + Better for the “One platform for all kinds of workloads” argument
- More complex WLM configuration
- No HA for ODS or DWH



33



IBM INFORMATION ON DEMAND 2007

Act Right. Now.



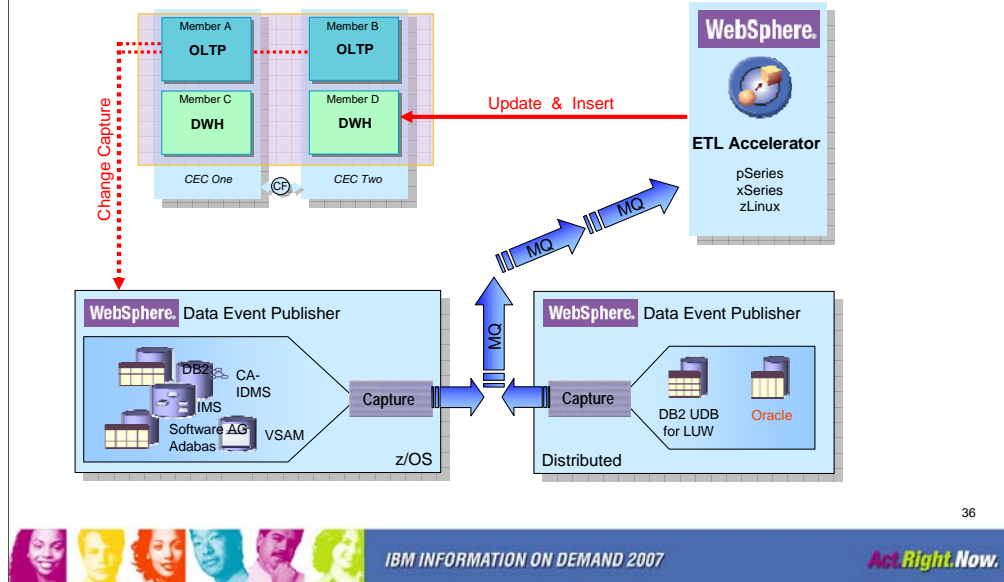
# Data Movement and Transformation best practices

*Act.Right.Now.*





## At runtime, the Data Warehouse is updated incrementally



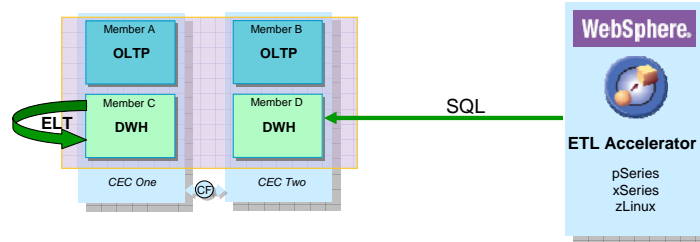
36

Go on now to slide 37:

Once the data warehouse is loaded, it needs to be updated in incremental mode. On the host side, WebSphere Classic Event Publisher is used to capture any changes in the data sources like IMS, VSAM or even the OLTP part of DB2. These changes are then pushed through MQ series pipes directly into specialized DataStage input stages. The changes are then fed into the data warehouse immediately or within defined windows.

Distributed data sources like DB2 LUW or, in the immediate future, Oracle can also be used for change capture and feed their updates into the accelerator.

## In Database ELT is triggered by DataStage



```
-- Aggregate by salary by department into AGGRSALARY
```

```
INSERT INTO AGGRSALARY ( DEPTCODE, AVGBAND, AVGSALARY )  
  SELECT DEPTCODE, AVG( BAND ) AS AVGBAND, AVG( SALARY ) AS AVGSALARY  
  FROM STAFF GROUP BY DEPTCODE
```

Whenever possible, "In Database" transformations (ELT) are used to spare the transport of the data to the accelerator. But the used SQL is still sent from the ETL Accelerator to the database to have one place of documentation for all ETL steps.

This can also be used to shift the data up the hierarchy within the Layered Data Architecture.

37



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

Next on slide 38:

Most efficient transformations happen within the database without the data ever leaving the DB2 system. This can be done whenever SQL delivers all necessary functionality to perform the transformation. Aggregations from detailed data into day, week or month summaries can be one example for this kind of in database transformations, here described as ELT (Extract Load Transform) vs. ETL.



# DB2 z/OS Configuration best practices

*Act.Right.Now.*



## DB2 Recommendations for the DWH

- Use sliding secondary quantities on table/index space creations
- V9: Use universal tablespace and just limit the dataset size with DSSIZE parameter. Let DB2 add new partitions automatically as needed. This also allows partitioning w/o partitioning key.
- Limit the parallelism with the PARAMDEG ZPARM.  
DB2 Optimizer tends to spawn unhealthy amount of parallel tasks if not limited. Recommendation is to limit it to 1..2 times of existing processors to avoid latency problems in storage.
- Switch STARJOIN ZPARM to DISABLE unless you're sure that it is necessary with the environment (mainly for SAP customers)
- Place workfiles into own buffer pools
- Use largest page size for tables that are potentially scanned most of the time
- Within DSG: Contention on spacemap pages might be problem in high ingest scenarios. Use member clustering in these situations.
- Use page level locking

39



IBM INFORMATION ON DEMAND 2007

Act Right. Now.



## DB2 Recommendations for the DWH (continued)

- Plan RUNSTATS on heavy used tables (with high update ratios).
- Compression: ESA data compression helps a lot in I/O or memory bound scenarios because it reduces channel activity and keeps more data in buffer pools. The CPU overhead can be moved in parts to zIIP. Of course doesn't make sense in CPU bound scenarios.
- In V9 you can switch of logging for specific tables. This makes sense if high ingest workload is causing log problems within a staging area. Logging should be switched off for temp. tables always.
- Limit Indexes on high ingest tables. Switch ingest tables on periodic base to standard tables by adding necessary indexes. Create a UNION view over all related standard tables and current ingest table.
- Perform MultiRow Inserts wherever possible.

40



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

## DB2 ZPARAMs for DWH

OPTIXIO = YES	Yes means to switch off optimizer extensions which were introduced with V8. Otherwise problems in access path evaluation might arise on complex queries.
OPTIORC = YES	Should be YES in all customer environments anyway.
PARAMDEG=1..2 times engines	Used to limit the parallel degree to a healthy limit
STARJOIN=DISABLE	Unless physical star joins are used
SRTPOOL=8MEG	Sort area per thread.
TABLES_JOINED_THRES HOLD=10	Amount of tables which are checked by optimizer to find the optimal access pass. CPU cost for this operation increases exponentially with the amount of checked tables.

41





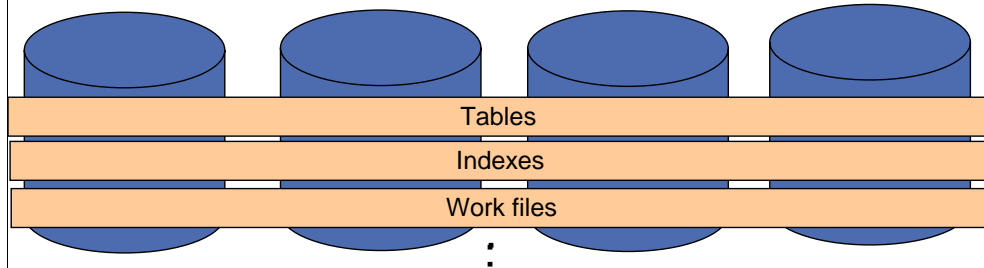
# Storage best practices

*Act.Right.Now.*



## Data Set Management

- Avoidance of logical volume (UCB) contention
- Random allocation but even distribution of data
- Suitable for single or multiple DS8000 subsystem configurations



- Use Dynamic PAV (Parallel Access Volume) and multiple allegiance
- Use DS8000 MIDAW facility
- Use VSAM data striping

43



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

# Storage Recommendations

- Rule of thumb storage requirement = 3 to 5 times raw data.
  - If 5 times, 10TB of warehouse data would lead to 50TB storage requirements
- **Storage objects to account for**
  - Indexes: The size of index spaces is 2 to 3 times of the size of the tablespace
  - Tablespace: an average of 10 to 15 indexes per table
  - Workfiles: Sizing for active/archive logs and workfiles is necessary
  - DB2 Directory and Catalog: Distribute the logs (2-5 volumes)
  - Temporary tables: Bigger amount of workfiles, spread over several volumes, should be constantly allocated. This improves performance on merge operations. A minimum of 5 workfiles should be allocated. Override SMS for workfile placement.
  - MQTS

**Space  
1:3-5**

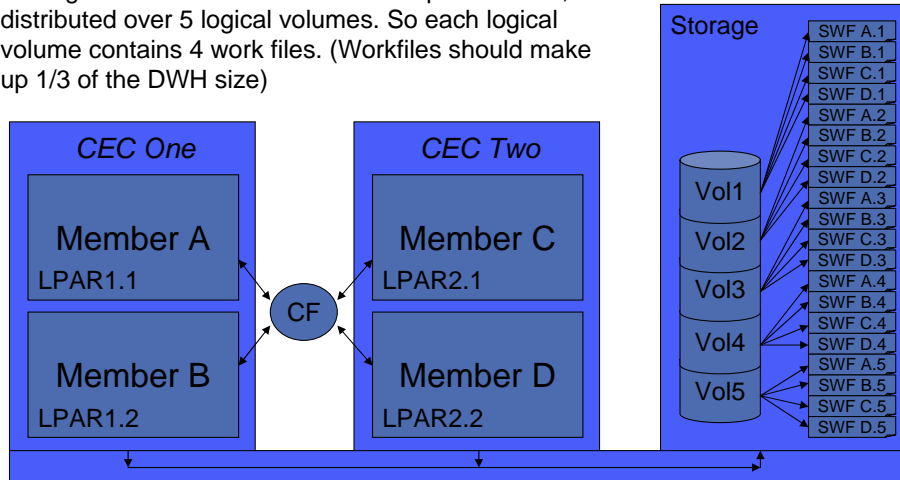
**Distribute  
workfiles  
over different  
volumes**

44



## Example of Sort Workfile Distribution with 4 DSG Members

Having a minimum of 5 sort work files per member, distributed over 5 logical volumes. So each logical volume contains 4 work files. (Workfiles should make up 1/3 of the DWH size)



45



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

My email discussion with Ute was about DB2 sort workfiles and not DFSORT sort datasets. Dynamic PAV has diluted and largely eliminated IOSQ time as a serious contributor to IO response time in most cases i.e., IOSQ is virtually zero. The only exception where I see significant IOSQ is on logical volumes containing sort workfiles. However, it is starting to re-emerge on densely packed highly accessed large logical volumes. For OLTP based systems, customers are using very large and customised (e.g., VPSEQT=100, VDWQT=50, DWQT=60) dedicated sort workfile bufferpool to try to get concurrent small sorts completed in memory without spilling to sort work files and read back from sort workfiles i.e., configured for optimum performance. So in this case the isolation and separation of DB2 sort workfiles is a moot point for discussion. This is not true for batch and not true for DWH/BI, where the sort size is highly variable and can be XXXL size, and the degree of concurrent sort is highly variable. In DWH/BI need allow for something like 30% of total DB2 user table data size for DB2 sort workfiles. Here the sort workfile bufferpool should be configured for robust defensive performance (e.g., VPSEQT=100, VDWQT=10, DWQT=50), and the sort workfiles separated and isolated to avoid excessive IOSQ. Assume minimum data sharing configuration to be 2 boxes, 4 LPARs (2 LPARs per box), one DB2 member per LPAR (4-way data sharing) then I would dedicate an absolute minimum of 5 volumes for sort workfiles, have 5 workfiles per member (one per logical volume) which means 4 workfiles per logical volume (one for each DB2 member).

John J. Campbell

Distinguished Engineer

DB2 UDB for z/OS Development

IBM Software Group - Information Management

eMail: [CampbelJ@uk.ibm.com](mailto:CampbelJ@uk.ibm.com)

## SMS Implementation

- Use DB2 Storage groups (to define and manage data sets) in conjunction with SMS Storage Classes / Groups (for data set allocation)
- Simple ACS Routines using HLQs to direct datasets to appropriate Storage Class / Storage Group
- Storage Groups requiring maximum concurrent bandwidth should consist of addresses spread across the I/O configuration
- For Storage Classes, set IART parameter to non-zero so SMS will use internal randomization algorithm for final volume selection

46



IBM INFORMATION ON DEMAND 2007

Act Right. Now.



# Warehousing in System z - Summary

*Act.Right.Now.*





## Where you put your Data Matters....

### Confidence in System z, z/OS and DB2 for z/OS

- **Integrity**

- z/OS® System Integrity Programming Standard – in writing
- IBM System z™ integrity features that help protect data

- **High availability**

- Designed with a 'Never go down' philosophy as opposed to a 'rapid reboot' philosophy
- Capability of providing concurrent HW maintenance and upgrades and rolling changes to DB2® for z/OS (in a Parallel Sysplex® cluster) can mean fewer database outages

#### DB2 for z/OS in:

- 25 of the top 25 WW banks\*
- 23 of the top 25 US retailers\*\*
- 9 of the top 10 global life/ health insurance providers\*\*\*

- **Security**

- Encryption, encryption, encryption – comprehensive solution
- MLS – merge data into single server and helps preserve data isolation.
- Helping address regulatory compliance with ability to establish centralized policies and procedures for privacy, security and audit

- **Total Cost of Ownership (TCO)**

- **Systems and database management**

48



IBM INFORMATION ON DEMAND 2007

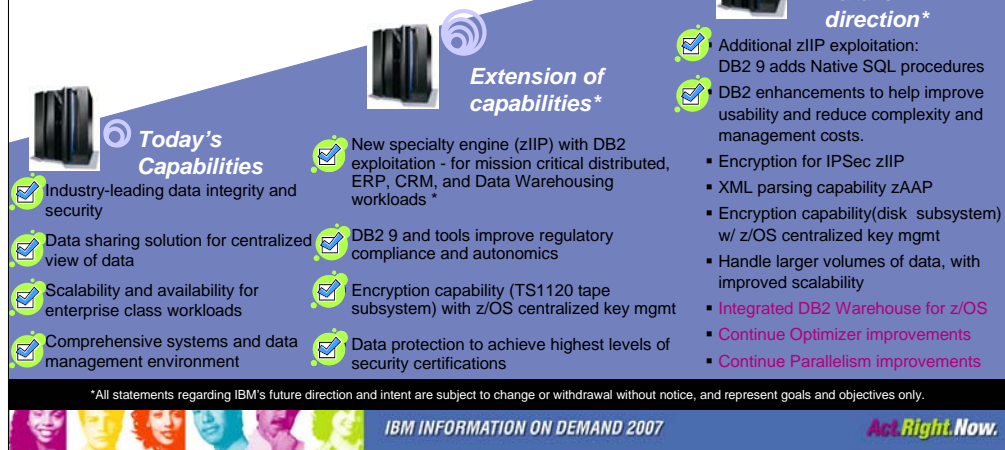
Act Right. Now.

In summary, on slide 31, I want to re-iterate that System z and DB2 for z/OS provide the highest levels of availability, resiliency, security, and recovery at very competitive cost of ownership. With BI and warehousing workloads moving from the back office to the front office, those System z strengths make it a very attractive platform.

Remember also that by keeping your operational data and warehouse together in System z, you can streamline compliance and security, reduce complexity and reduce the overall cost of your BI solution.

## A vision for System z advanced data serving System z Enterprise Hub for Mission Critical Data

- ❖ With a strong foundation for transaction processing, built on 40+ years of technology innovation, System z servers with z/OS and DB2 can provide a premier platform for data serving, today and into the future\*
- ❖ IBM plans to continue to invest in new solutions to address customers' strategic information on demand goals\*



On slide 41, I show the roadmap that positions System z and DB2 z/OS as a hub for mission critical data. It re-emphasizes the investment IBM has made in warehousing and shows that we will continue to make enhancements in this space. The main point to take home is that IBM understands that warehousing on System z and DB2 for z/OS is important and we are committed to continue making enhancements that improve the value of BI solutions on DB2 z/OS and System z.



On a final note, I would like to thank everyone for attending and to mention that I've included backup slides with additional warehousing and DB2 z/OS information that you may find useful.

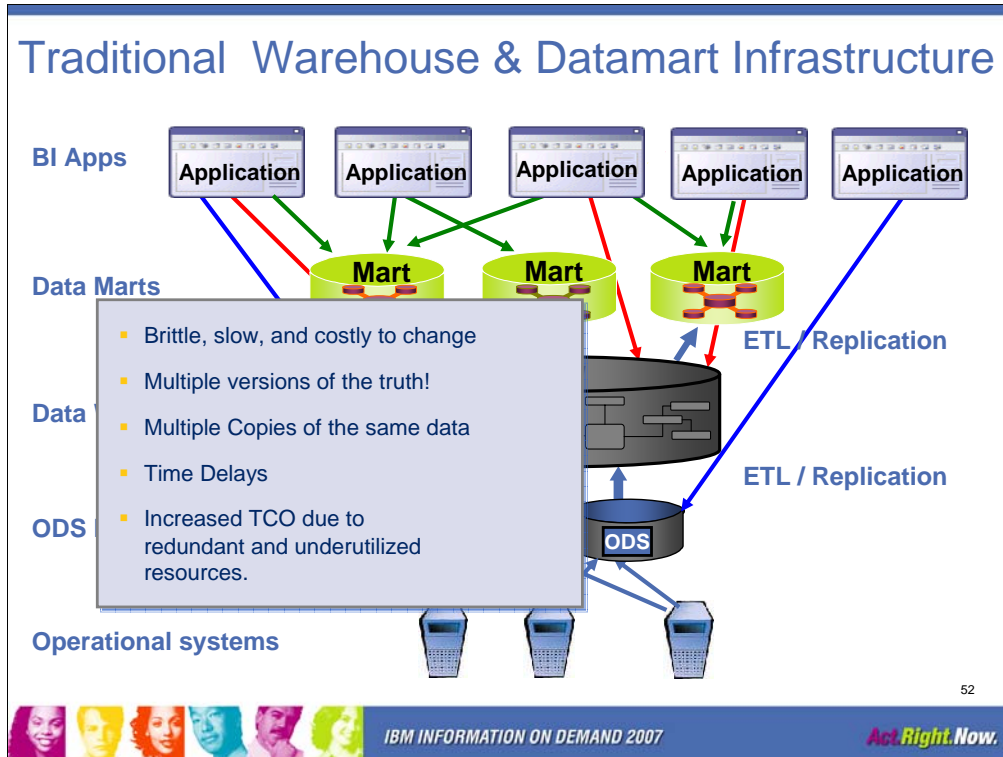
With that, I'd like to now open it up for any questions you may have.



# Backup Slides

*Act.Right.Now.*





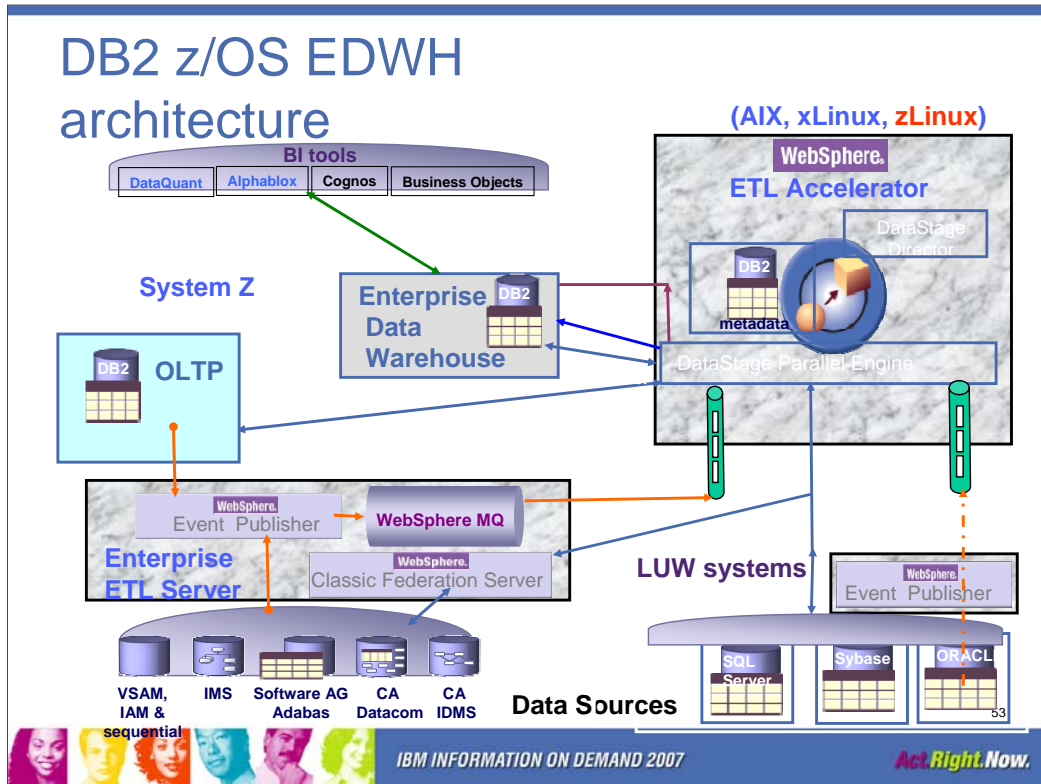
This is likely where the customer is today. A layered architecture with complicated interactions between the layers. Often the layers are owned by different parts of the business (for example, the CIO typically stops with the warehouse. The LOB user owns the marts). Individual marts being created and maintained, often at the departmental level, to meet isolated business needs. Applications being developed in isolation, being forced in interface with multiple layers in the stack.

Many of our customers are finding that this is a difficult and expensive infrastructure to manage and grow. Application or data requirement changes ripple through multiple layers (and potentially across organizations). The risk of disrupting service is higher – the cost of growing the environment is higher. You have multiple copies of the same data resulting in problems of synchronization.

It is also difficult to implement more dynamic right time data loading, which as we've discussed is becoming a more significant requirement for our customers.

Examples of how this breaks down (there are no customers attached to these scenarios, but we see these types of scenarios often in the field).

**Proliferation of marts :** A department loses faith in the existing BI infrastructure and chooses to go off and implement their own mart, their own view of the world. Because the marts are decentralized, there is no one who



## EDW

Highest Qualities of Server: highly available, best security,

Single DB system view with OLTP

Flexible yet high performance layered DWH

Flexible logical data marts (views, partitions, online DDL) with performance objects and the flexibility to convert to physical data marts for even more performance once flexibility is not as important

Performance objects (MQT, user aggregates, parallelism)

## ETL Accelerator

Single integration metadata store

Single event manager

Flexible transformations

Parallel engine

Near real time data streaming and/or high performance bulk transfers

Extensive data sources on z/OS and distributed systems

## Modern BI reporting and Analytics

GenevaERS for static report consolidation (satisfy many reports with one access to the DWH)

DataQuant and/or Alphablox for reporting and analytics

## Slides and North America Handouts

- **DB2 presentations:**

<ftp://ftp.software.ibm.com/software/data/db2zos/>

- **1 DB2 and SOA today**
- **2 Leveraging zIIP for DB2 workloads**
- **3 DB2 for z/OS evolution in an SOA world**
- **4 XML in DB2 for z/OS**
- **5 A brave new world:**

**Accessing DB2 for z/OS data in SOA**

- **6 DB2 tools update: The key to optimized DB2 environments and reduced TCO**
- **Handouts for North America road show sessions:**

<http://www.ibm.com/software/os/zseries/roadshows/handouts/>



54



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

### 1 DB2 and SOA today

DB2 for z/OS V8 supports a number of SOA capabilities, and this session will offer an update on how IBM customers are benefiting as they upgrade to V8 and fully exploit these rich functions.

### 2 Leveraging zIIP for DB2 workloads

IBM announced its new zIIP specialty engine offering in 2006. The zIIP is designed to help process certain DB2 related tasks at a reduced software license and processor cost. The workloads which benefit are distributed SQL access, parallel queries, and some utility processing. SAP customers are a good example. In this session, we will explain how you can start making use of the zIIP and which workloads qualify for the zIIP. We will also provide some best practices on how to maximize the usage of the zIIP.

### 3 DB2 for z/OS evolution in an SOA world

This presentation will provide an overview of trends and directions of DB2 on z/OS. Special focus will be on SOA enablement, with the primary focus on DB2 9.

### 4 XML in DB2 for z/OS

DB2 for z/OS provided an XML extender in DB2 V7, then added XML publishing functions in DB2 V8. DB2 9 provides extensive changes for pureXML. In this session we explain how to use XML with DB2 for z/OS and provide different usage scenarios including how to leverage XML data in an SOA.

### 5 A brave new world: Accessing DB2 for z/OS data in an SOA

DB2 data and stored procedures can be accessed via Web services, or via Java programs using JDBC or SQLJ. Come learn about the benefits of each of these technologies, along with tips on how to select the right method for your requirements, and best practices in the deployment of each.

### 6 DB2 tools update: The key to optimized DB2 environments and reduced TCO

DB2 tools can help you streamline database administration, reduce complexity by providing autonomic features that add capability and minimize error potential, and preserve investment in z/OS applications - all factors that help reduce total cost of ownership. Learn how DB2 tools can help optimize your current DB2 environment, assist in migrating to DB2 Version 8, and prepare for DB2 V9. We'll also look at how tools can help you meet auditing requirements, address regulatory compliance issues, improve recovery and performance, and enhance your change management process.

ftp://ftp.software.ibm.com/software/data/db2zos/

Name	Size	Type	Modified
DB2utilityBestPracticesBSmith.pdf	883 KB	Adobe Acrobat 7.0 ...	6/8/2007 5:02 PM
DM07DynamicDataWarehousingz2007June01.pdf	2.30 MB	Adobe Acrobat 7.0 ...	6/8/2007 5:02 PM
AccessDBzOSDataSOA2007April18.pdf	1.00 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
Application_perform_tuningV9.pdf	9.37 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
C05BryanPaulsenDB29experencer.pdf	429 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
C06TengDB2_9_Administrators.pdf	119 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
DB29zOSextract2007June01.pdf	575 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IDUGC03Attack_DB2_CloneTables.pdf	461 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IDUGC05BryanPaulsenDB29experencer.pdf	429 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IDUGC06TengDB2_9_Administrators.pdf	119 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IDUGDBfor tomorrow_2007dist.pdf	4.89 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IMTechzOS17hottopics.pdf	1.29 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IMTechzOS18protectcomply2007May28.pdf	1.52 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IMTechzOS19bestpracticesecurity2007June08.pdf	1.08 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IMTechzOS19bestpracticesecurity2007May28.pdf	883 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IMTechzOS20planmigration.pdf	2.09 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IMTechzOS21zIP.pdf	1.42 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
IMTechzOS22DB2_9.pdf	887 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
PortingDB2toDB2_2007June01.pdf	439 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
RTFW2007June05.pdf	1.48 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
security_enhancements_V9.pdf	316 KB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
WAS_DB2_Integration_Update.pdf	1.34 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
XMLDBzOS2007April18.pdf	1.42 MB	Adobe Acrobat 7.0 ...	6/8/2007 3:03 PM
FLhosp_ConvertingFetch.pdf	2.14 MB	Adobe Acrobat 7.0 ...	5/16/2007 3:02 PM
FLhosp_MultrowFetch.doc	37.0 KB	Microsoft Word Doc...	5/16/2007 3:02 PM
FLhosp_Mult-RowFetchExample.doc	37.0 KB	Microsoft Word Doc...	5/16/2007 2:02 PM
EnterpriseCOBOLupdate_smiller.pdf	44.0 KB	Adobe Acrobat 7.0 ...	5/9/2007 1:02 PM
EnterpriseCOBOLupdate_smiller.pdf	27.1 KB	Adobe Acrobat 7.0 ...	5/9/2007 1:02 PM
IDUGNA200707migration2007May06_smiller.pdf	1.96 MB	Adobe Acrobat 7.0 ...	5/9/2007 1:02 PM
WDzV7COBOLupdate_smiller.pdf	103 KB	Adobe Acrobat 7.0 ...	5/9/2007 1:02 PM
SHAREdb2zPerformanceShibamiya.pdf	243 KB	Adobe Acrobat 7.0 ...	3/7/2007 4:02 PM
SHAREdb2z9MigrationMiller.pdf	2.00 MB	Adobe Acrobat 7.0 ...	3/7/2007 4:02 PM

55

IBM INFORMATION ON DEMAND 2007 Act Right. Now.

This is the ftp web page as of June 12, 2007. See the 23 new in the past week and 7 more in the prior month.

ftp://ftp.software.ibm.com/software/data/db2zos/ is the web page. I use Internet Explorer for this page, and sort by the modified date.



## Data Warehousing for DB2 on System z – more info

- **Whitepaper** : Why Data Warehousing on System z available on the web  
<http://www.ibm.com/software/data/db2bi/systemz.html>
- DW on system z – **Demo** available in the Technical Marketing Competence Center, Böblingen, Germany, [TMCC@de.ibm.com](mailto:TMCC@de.ibm.com)
- DW on system z – **Customer Briefings** possible in the Technical Marketing Competence Center, Böblingen, Germany or at Silicon Valley Lab [TMCC@de.ibm.com](mailto:TMCC@de.ibm.com)
- DW on system z – '**Best Practices**' paper planned to be available 2Q2007
- ...



56



IBM INFORMATION ON DEMAND 2007

Act Right. Now.

Here are some resources on the web about System z and data warehousing.

United States [ change ]

Search

Home Business solutions · IT services · Products · Support & downloads · My IBM ·

Welcome Mr. Roger Miller [Not you?] [ IBM Sign in ]

Software > Information Management

**Dynamic Warehousing from IBM**

A new approach to delivering enhanced business insight.

**MOVE FROM WAREHOUSING TO DYNAMIC WAREHOUSING**

Leverage embedded analytics in your business processes.  
 Incorporate knowledge from unstructured information.  
 Support real-time access to aggregated, cleansed information.  
 Use integrated capabilities for Information On Demand.

**Why IBM**

IBM is the first to fully understand and define this next generation of data warehousing. Only IBM can deliver all of the Dynamic Warehousing capabilities with a product portfolio that is best in class.

**New white paper**

From conflicting unintegrated historical data to actionable insight.

IBM INFORMATION ON DEMAND 2007 **Act Right. Now.**

See these web pages for more on Dynamic Warehouse.  
<http://www.ibm.com/software/data/db2bi/data-warehousing/>

Software > Information Management >

## Business Intelligence on System z

**Business Intelligence on System z**

**Success stories**

**News**

**Events**

**Related software**

- More Business Intelligence
- DB2 for z/OS

**Related hardware**

- IBM System z9 Integrated Information Processor (zIIP)
- Warranties, licenses and maintenance

**Description**

The distinction between data warehousing applications and online transaction processing (OLTP) applications is blurring. Data warehousing and analytic applications are accessing operational data or near-real-time data.

IBM provides a sophisticated data warehousing solution including - integrating/building warehouses, analyzing and reporting on that data, streamlining security and regulatory compliance, performance tuning, and application management capabilities. It combines the traditional strengths and new capabilities of [DB2 Universal Database](#) and the

**Product offerings**

- **WebSphere DataStage for z/OS** takes ETL performance to a new level with powerful data integration and transformation.
- **WebSphere Replication Server for z/OS** replicates data between DB2 for z/OS and other databases for high availability
- **DB2 Alphablox** imbeds analytics directly into web based applications.
- **Rational Data Architect** helps data architects understand information assets, design federated databases, and streamline database projects.

**Information Management software**

**IBM Information On Demand 2007**

**Oct 14-19, 2007**

**Register Now!**

Early, Early Bird Rate \$1595

The Premier Information Management global conference

**Highlights**


- Announcement Letter 206-181

**White papers**

[Why Data Serving on a Mainframe](#)

**Events**

- Replay: "Why BI on z9/System z makes "cents"
- Sept 7 Webcast on "Data Warehousing on ..."



IBM INFORMATION ON DEMAND 2007

**Act Right. Now.**

Business intelligence on System z:

<http://www.ibm.com/software/data/db2bi/systemz.html>