Why DB2 on z and What’s New in DB2 for z/OS V8

Dr. Jim Teng
IBM Distinguished Engineer
Information in this presentation about IBM's future plans reflect current thinking and is subject to change at IBM's business discretion. You should not rely on such information to make business plans. Any discussion of OEM products is based upon information which has been publicly available and is subject to change. The opinions expressed are those of the presenter at the time, not necessarily the current opinion and certainly not that of the company.

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*, Enterprise Storage Server, ESCON*, IBM, iSeries, Lotus, NOTES, OS/400, pSeries, RISC, WebSphere, xSeries, z/Architecture, z/OS, zSeries, System p, System i, System z

The following terms are trademarks or registered trademarks of the Microsoft Corporation in the United States and/or other countries: MICROSOFT, WINDOWS, ODBC

For more copyright & trademark information see ibm.com/legal/copytrade.phtml
Topics

- Introduction
- Why DB2 z?
- DB2 z/OS V8 Functional Overview and Value Proposition
- Q&A
Introduction and Overview of DB2 z/OS
Brief DB2 History

1983

Sequential Prefetch

Referential Integrity

1992

DB2 for z/OS

Parallel IO

1993 1995 1997 1999 2001 2004

Hiperspaces

Data Sharing

T2 Indexes

CPU Parallelism

System

V6 V7 V8

Universal

Database

Online Reorg

TCPI

Sysplex

Parallellism

SQL Enhancements

Statistics History

Utility Wildcards

Triggers

UDF, UDT

LOBs

Data Spaces

GUI Tools

Stored Proc Language

Rexx

Breakthroughs

VLDB

New Workloads

Unicode

Long Names

Memory

Availability

Performance

SQL, SOA
DB2 for z/OS Strategic Direction

- Continue to Evolve as “The” Secure Enterprise Data Hub Server
  - Centralized Processing of Massive Quantities of Data
  - DB2 Family Compatibility
  - OBDC/CLI, JDBC, SQLJ, Unicode, SOA Stored Procedure, Java, ...

- Premier Use of Parallel Technology and z/OS - zSeries Integration
  - Data Sharing
  - Continuous Availability
  - Incremental Growth
  - Work Load Balancing

- Improved Application Enablement Through Extensive SQL Functionality
  - Native XML support
  - Object-Oriented Extensions - User defined functions and types
  - Rules Driven Integrity - referential integrity and table constraints
  - Vendor Enabling and Portability

- Continual Performance Improvement
  - Optimizer Technology
  - Transaction, Query, Batch, Utility Concurrency with Application Processing
  - Utilities and Tools Focus
Why DB2 z/OS?
Why DB2 z for Enterprise Database Solutions

- Availability
- Scalability
- Security
- TCO
DB2 and zSeries: Designed for the Highest Availability

- Continuous Availability
  - ‘Do Not Go Down’ philosophy
  - Maintenance / Upgrades Without System Downtime
- Integration with OS and HW means
  - Less complexity
  - More robustness
  - Higher scalability
  - Industry leading disaster recovery
- DB2 for z/OS V8 Continues to Improve
  - Schema evolution
    - Modify Tables while database is running
    - Dynamically add partitions
  - Change system parameters on the fly
    - As conditions change, Data remains available
  - Online Reorg and Load
    - Avoid downtime
    - Improve performance
  - System Level Point in Time Recovery
    - Easier, more flexible, less disruptive, faster recovery
  - Continually enhanced tools that support newest features
Availability: Disaster Recovery

- DB2 Tracker
  - Continuously applies logs
  - Limits recovery time
- Snap Without Quiesce
- Replication – Log Based Capture
- Remote Disk Mirroring
  - Low cost recovery method option
  - XRC – Asynchronous mirroring over long distances
  - Data consistency on the host, DB2 and non DB2 data.
- GDPS
  - Disk mirroring with automated scripts,
  - Continuously and safely maintains a time consistency copy of critical data – DB2 and non DB2 data
  - Switches workload to a backup site when primary site shuts down
  - Automatically accesses copy of data and rapidly restarts critical applications
  - Enables database restart (in minutes) with no manual intervention
Architecture: Parallel Sysplex / Oracle RAC

- Single system image, one large db
- Dynamic workload management
- Sophisticated systems management
- Integrated, reliable HW/SW/OS solution
- No single points of failure

- High-speed inter-system links
- No zSeries Coupling Facility means:
  - More inter-system messaging
  - Performance overhead
  - DB partitioning, application may need changes
Scalability: Average Capacity Utilization
Oracle RAC vs. DB2 Parallel Sysplex

Graph showing capacity utilization for Oracle RAC and DB2 Parallel Sysplex at various numbers of nodes.
Scalability: Workload Balancing

Same workload

Mixed Workloads on zSeries

z/OS WLM balances resources for you
- Based on business goals
- No other platform comes close

You do more work!
- Workloads not integrated
- Partitions managed like separate servers
- Servers under utilized

Mixed Workloads on UNIX
4 UNIX Instances, 4 separate SMPs or Partitions

Processor Utilization Percentage

Web Serving
Business Intelligence & Data Mining
Batch
SAP
DB2 for z/OS: Scales Like No Other

Resource Optimization
- Run multiple applications on same machine – higher HW utilization
- Efficient storage optimization with high performance
- Mature DB2 optimizer

Special Hardware Assists
- Sort, Compression
- Unicode Conversion
- Encryption
- Workload Management
- Coupling Facility

Add Members
- Non disruptive, no repartitioning required
- Little to no performance impact
- Maintain Application Transparency
Security: Multilevel Security on zSeries

Integrated Solution: Exclusive to zSeries

Secure
- MLS with row-level granularity
- Highly secure access to a single DB2 database, with single security manager
- RACF, Digital Certificates, PKI, Kerberos, Intrusion Detection
- Isolated logical partitions
  - The only platform to have highest international certification level (EAL5) and the U.S. Government FIPS4 certification
- zSeries Cryptography
- DB2 V8 Built-in Encryption

Autonomic & Resilient
- High Availability (with Parallel Sysplex® Clustering)
- End-to-End prioritization from browser to storage
- Workload Management
- Intelligent Resource Director

Other Solutions:
- Applications must be aware & protect
- Passwords are passed across lines, lose control
- Table level security
“Unbreakable” Oracle Breaks Again

So much for Oracle never breaking
A few weeks ago, I wrote this piece on absurd marketing claims in the latest Oracle TV ad that claims "never breaks". After seeing these ads almost every other night since then, we get news that Oracle just released their latest mega patch that addressed 49 security issues. I could be wrong but I can't remember the last time any software company that released such a large set of patches to address so many vulnerabilities.

George Ou
ZDNet.com blog, 7/14/2005
http://blogs.zdnet.com/Ou/?m=20050714
DB2 for z/OS: Provides Better TCO

- Less Complexity Keeps Costs Down
  - Simplified management minimizes people cost
  - Grow MIPS up to 10x with existing resources
- Costs Align With Consumed Resources
- Applications not modified to be “cluster aware”
- No “hidden” costs
  - SW, maintenance, switches, underutilized disk, license fees for backup
- Monitoring and problem diagnosis much easier:
  - Synchronized clock across cluster nodes (event sequencing)
  - z/OS System Logger using CF for merged message logs
  - Tools especially designed for performance monitoring and analysis
What Folks Are Saying about RAC “cost”

- Is RAC TCO higher than the marketing hype? People have a lot to say
- Again, look at http://www.freelists.org/webpage/oracle-l
- Try these search strings:
  - Excellent Way to Increase Consultant Cost
  - RAC Performance Everything Gets a lot More Expensive
  - Cost So High Bumps Up Against Actual Hardware Costs
DB2 for z/OS: The REAL Standard, Without the Hype

- Deep HW, SW, OS integration provides differentiated value for mission critical applications
  - No other solution comes close, and marketing hype can’t change that
- No other solution can match DB2’s industry leading
  - **Availability**
    - Failures are truly isolated
    - Manage planned maintenance without downtime
    - Use of coupling facility greatly reduces complexity, which greatly reduces outages
  - **Scalability**
    - Dynamically react to changing workload conditions
    - Consolidate workloads
    - Use of coupling facility provides for industry leading performance and scaling
  - **Reliability**
    - Industry leading RACF Security
    - Industry proven on ‘bet your business’ applications
  - **TCO**
    - Simplified management – do more with existing resources
    - Align costs with resources actually consumed
“I make fun of a lot of other databases - all other databases, in fact, except the mainframe version of DB2. It's a first-rate piece of technology.”

Larry Ellison; eWeek October 31, 2003
In Larry's Own Words By Matthew Symonds
DB2 z/OS V8 Functional Overview and Value Proposition
DB2 V8 General Availability
March 26, 2004
V6 end of service June 30, 2005
DB2 for z/OS Version 8 is

- SQL OLTP Leadership: Multi-row Fetch & Insert, Select within Insert, Dynamic Scrollable Cursors, ...
- Breaking through limitations: storage, name lengths, SQL statements, partitions, logging
- Performance Enhancements: MQTs, index use
- Database changes without an outage: add partition, rotate partitions, alter cluster
- WebSphere software & Java Integration
- Building the e-infrastructure
DB2 for z/OS Version 8

- Scalability Enhancements including 64 Bit Storage
- SQL Optimizations
- Improved Security
- Performance Enhancements
- Availability
- Data Sharing Improvements
- Utility Optimizations
64 bit virtual storage

- Line $2^{24}$ Bar $2^{31}$ $2^{64}$
- Above the bar
  - Buffer pools, EDM pool, Sort Pool
  - Rid Pool, Compression dictionaries
  - IRLM locks, PC=YES
- Real storage requirements probably increase
- Improve scalability & availability
- Much reduced management for virtual storage

![64-bit address space memory map](image)
SQL and DB2 Family

- Multi-row INSERT, FETCH & UPDATE
- GET DIAGNOSTICS
- INSERT within SELECT
- IDENTITY Column enhancements
- SEQUENCES
- CURRENT PACKAGE PATH
- SQL Procedure Language
- Dynamic Scrollable Cursors
- Common Table Expressions
- Scalar Fullselect
- Materialized Query Tables
- UNICODE SQL, Multiple CCSIDs
- XML Publishing
- ...
### SQL Optimizations

**Breaking through limitations - SQL**

- **Table name sizes**: 18 to 128
- **VIEW & ALIAS names**: 18 to 128
- **Column name sizes**: 18 to 30
- **SQL statement length**: 32K to 2MB
- **Tables in a join**: 15 to 225
- **Character Literals**: 255 to 32704
- **Hex literal digits**: 255 to 32704
- **Predicates**: 255 to 32704
- **Index key**: 255 to 2000
SQL Optimizations

Breaking through limitations SQL ...

- Current optimization 8 to 128
- CURRENT PACKAGESET 18 to 128
- CURRENT PATH 254 to 2048
- SCHEMA 8 to 128
- Authorization ids still 8 bytes

Image of Earth from Moon, Source: NASA (Public Domain)
Improved Security

- Multilevel security with row level
- Session variables
- Special registers for identification
- RACF access control & DB2 operator commands
- Encryption and decryption built-in functions
- Encryption for DRDA on the wire
Multilevel Security by Row

- Table column defined AS SECURITY LABEL
- Check for each new seclabel value accessed
- Mandatory access control: run time user to data
Row Granularity Multilevel Security

Table has column defined AS SECURITY LABEL
Each row value has a specific security label
Get security labels from RACF
Save in rows for INSERT, UPDATE, LOAD, ...
Check for each new seclabel value accessed
  If access is allowed, then normal access
  If access is not allowed, data not returned
Runtime user to data checking
Seclabel values are cached to minimize cpu
Requires z/OS V1R5 and Security Server (RACF)
Session Variables

Variables set by DB2, connection or signon exit

Built in function to retrieve value for a variable

Use function in views, triggers, stored procedures & constraints to enforce security policy

Can have more general, flexible access checks

Multiple columns, AND/OR logic, ...

PLAN, PACKAGE, SECLABEL, VERSION, ...

Complements other security mechanisms

CREATE VIEW V1 AS SELECT * FROM T1 WHERE COL5 = GETVARIABLE('SYSIBM.SECLABEL');
New Special Registers

Client information for this connection
Provided by sqleseti, Java methods, RRS
SIGNON & SET_CLIENT_ID

• CLIENT_ACCTNG      accounting string
• CLIENT_APPLNAME     value of application name
• CLIENT_USERID       client user ID
• CLIENT_WRKSTNNAME   workstation name
New Built-in Functions

- ENCRYPT_TDES: encrypt column in a table with a user-provided encryption password
- ENCRYPTION PASSWORD special register
- DECRYPT_BIT, DECRYPT_CHAR, DECRYPT_DB
- GET_HINT: obtain hint to help remember ENCRYPTION PASSWORD
- GENERATE_UNIQUE creates CHAR(13) FOR BIT DATA value that is unique across Sysplex
Performance: DB2 UDB for z/OS

Performance Enhancements

– Ability to use indexes more often
  – More matching in predicates
  – Index backward scan
  – Index varying length
  – Distribution statistics on non-index columns
– Materialized Query Tables
– Multi-row fetch & insert
– Long Term Page Fix Bufferpool option
– Data Sharing improvements
Without Materialized Query Tables: Each Query Re-Computes!

Q11, Q12, ...  

Join  

Aggregation  

Compute Many Times  

WarehouseDB  

Q21, Q22, ...
With Materialized Query Tables: Avoid Redundant Computation

Q11, Q12, ...   Reuse Many Times   Q21, Q22, ...

MQT

Precompute Once

WarehouseDB
Materialized Query Tables

- Sometimes called Summary Tables, Automatic Summary Tables, Automatic Materialized Query Tables, Materialized Views, ...
- Optimizer can rewrite queries to access MQT instead of base table or view
- Pre-computed information, very significant performance improvement
- Managed by user or system (SQL REFRESH)
- Automatic rewrite or manual
- Informational Referential Integrity (not enforced)
Star Join Sparse Indexes

- Optimizer improvement that addresses the same requirement as Hash Join
  - ACCESS_TYPE='T' in PLAN table
  - Uses sparse index to process the contents of work file
  - Improves upon APAR PQ61458 on V7
Parallel Sort

- Optimizer is enhanced to use parallel sort for multiple tables when cost tradeoff is advantageous
Trigger Performance

- TRIGGER work files are eliminated when old/new transition variables are returned for small number of rows
- Very significant performance enhancement when few or no triggers fired
Table Function Improvements

- CARDINALITY clause
  
  SELECT * FROM TABLE(TABUDF(C2) CARDINALITY 30) AS ABC

- Performance improvement to move data between table functions and DB2 using block data movement
RUNSTATS Distribution Statistics Enhanced

- Non-uniform distribution statistics on non-indexed columns can improve performance 10X, 100X
- RUNSTATS improvement allows optimizer to consider non-uniform distribution statistics on columns that aren't part of an index
- Current technique is separate DSTATS program
- Significant performance improvement
- RUNSTATS with UPDATE NONE REPORT NO
Multirow performance

- FETCH:
  - Up to 50% cpu savings
  - DSNTIAUL change made
  - DSNTEP4 usage

- INSERT
  - Up to 30% cpu savings
Long Term Page Fix

- **ALTER BUFFERPOOL**
  - New option PGFIX(YES)
  - By buffer pool
  - Use where IO rate is high
  - Must have real storage
  - Up to 10% cpu savings
  - Allowed in all modes, e.g. compatibility mode
Availability: DB2 UDB for z/OS

Continuous Availability

- Schema Evolution: database changes ALTER instead of DROP / CREATE
- Data Partitioned Secondary Indexes
- System-Level Point in Time Recovery
- Improved LPL Recovery
- Additional online zparms
New ALTERs

- Add a partition to the end of a table
- Rotate partitions
- Extend CHAR(n) column lengths
- Change type within character data types (CHAR, VARCHAR)
- Change type within numeric data types (small integer, integer, float, real, float8, double, decimal).
- Change type graphic data types (GRAPHIC, VARGRAPHIC)
- Includes column data type changes for columns that are referenced within a view
- Includes column changes for indexed columns
New ALTERs ...

• Add a column to an index
• Drop partitioning index (or create table without one)
• Alter to clustering index or to not cluster
• Alter an index to have true varying length character columns within a key (not padded) or to padded
• Alter identity column attributes
• Alter sequence attributes
• Alter VOLATILE attribute
• Add column AS SECURITY LABEL
• Alter function & procedure failure options
• Alter view regenerate
Online Schema Evolution 1

- Changing Partition Definitions
  - Add Partition
  - Rotate Partition
    - Move first part to last
    - E.g. keep rolling 36 months

- ALTER INDEX : ADD COLUMN, PADDED or NOT, CLUSTER or NOT

- CREATE INDEX while online
Online Schema Evolution 2

- Unbundling Partitioned Table Attributes
  - Partition without an index
    - Table controlled partitioning
    - May be able to have one less index
  - Data Partitioned Secondary Index
  - Cluster on any index
    - May be able to have more efficient clustering
    - E.g. partition by date, cluster by account
  - ALTER CLUSTER attribute
Online Schema Evolution 3

- Table or column Changes
  - Increase column within numeric data types
    - smallint, integer, decimal, float
    - No loss of precision allowed
  - Change to expand character data type
  - Change varchar to / from char
Data Partitioned Secondary Indexes (DPSIs)

- Benefits include partition independence:
  - More efficient utility processing
  - Higher availability
  - Streamline partition level operations
  - Potential for lower data sharing overhead

- Potential impact to query performance
  - Partition key is not specified
  - Many partitions to search
  - Not allowed for unique index
Data Partitioned Secondary Indexes (DPSIs)

- DPSI = physically partitioned secondary index
  - Number of parts in DPSI = Number of parts in table
  - Keys in part 'n' of DPSI refer only to rows in part 'n' of table

- 3 kinds of indexes now:
  - Partitioning Index (PI).
    - As today, except optional in V8 and may or may not be partitioned
  - New Data Partitioned Secondary Index (DPSI).
  - Non Partitioned Secondary Index (NPSI) As today's NPI
Sample New Style Table - Table-based partitioning

- Partition data by month (PI is optional!)
- Clustering by id or name (DPSI clustering)
- Ideal for Online Reorg with fast switch, no BUILD2
Index: DB2 UDB for z/OS

Index Improvements

- Variable length index keys
- Index-only access for varchar data
- Maximum index key 2000 bytes
- Predicates indexable for unlike types
- Backward Index Scan
- Partitioning separate from clustering
- Data-partitioned secondary indexes (DPSI)
- Create index online during select, insert
Indexable Predicates: DB2 for z/OS

- Predicates indexable for unlike types
  - Column is decimal; Host variable is float
  - Column char(3); Literal or host variable char(4)
  - Can be used with transitive closure
  - Some restrictions still for stage 1, indexable
Stage 1 Indexable Unlike-types

- DB2 enhanced to allow index access when host variable & target column are not the same data type
- Deals with programming languages that don't support the full range of SQL data types
  - C/C++ has no DECIMAL data type
  - Java has no fixed length CHAR data type
  - etc.
- Significant performance improvement for many applications
- Simplifies application programmer & DBA tasks
4096 Partitions

- Maximum number of parts raised from 254 to 4096
  - Table spaces and indexes
  - Table space use DSSIZE to go beyond 254 parts
- ALTER TS ADD PART adds partitions to the end
- Maximum table size remains 16 TB for 4 KB pages
- Data set naming convention
  - 'Axxx' - partitions 1-999
  - 'Bxxx' - partitions 1000-1999
  - 'Cxxx' - partitions 2000-2999
  - 'Dxxx' - partitions 3000-3999
  - 'Exxx' - partitions 4000-4096
- Maximum number of parts allowed depends upon page size & DSSIZE
  - 4K page size, DSSIZE=1GB => 4096 parts, 4 TB max table
  - 4K page size, DSSIZE=64GB => 256 parts, 16 TB max table
System Level Point In Time Recovery

- Easier, more flexible, less disruptive, faster recovery
- Handle large numbers of table spaces & indexes
- Two new utilities introduced:
  - BACKUP SYSTEM: Fast volume-level backups
    - DB2 databases and logs
    - Data sharing group scope
    - z/OS V1R5, DFSMSHsm, DFSMSdss, & FlashCopy required (FlashCopy Version 2 recommended)
  - RESTORE SYSTEM
    - To an arbitrary point-in-time
    - Handles creates, drops, LOG NO events
Provide an easier and less disruptive way for fast volume-level backup and recovery

- Use FlashCopy to backup DB2 data and logs
- No longer need to suspend logs
- Backups are managed by DB2 and DFSMSHsm to support system level Point-In-Time recovery
Data Sharing Enhancements

- Batching of GBP writes and castouts
  - Write/castout multiple pages in a single CF operation
  - Improved data sharing performance, especially for batch updates
  - Requires z/OS V1R4, CFLEVEL=12

- Reduced global contention for table space L-locks
  - Reduced XES-level contention across members
  - Improved data sharing performance, especially for OLTP
  - RELEASE(DEALLOCATE) may not be needed
Data Sharing Enhancements …

- Changed pages written to GBP at Phase1 instead of Phase2
  - Transactions invoking other transactions at syncpoint for same data
  - Unusual "record not found" from another member
  - Easier to manage
  - Equivalent performance

- More efficient index split processing for data sharing
Automatic Storage Management

- No need to specify space parameters
- Start with small disk space
  - Start with small secondary extents
  - Larger secondary extents as table grows
- Improves DBA productivity
- Avoids application outage
- Reduces need to reorganize
VLDB: DB2 for z/OS

Very Large Database Improvements

- Separate partitioning from clustering
- Data-partitioned secondary indexes
- 4096 Partitions
- Add partitions
- Rotate partitions
- Extend columns
- Optimization improvements
- Partition without an index
Utility Improvements

- Easier restart (V7 PQ72337)
- Schema Evolution
  - REBALANCE partitions
- On-line REORG Enhancements
  - DISCARD
  - Avoid BUILD2 with DPSI
  - REORG DB2 catalog SHRLEVEL REFERENCE
- LOAD & UNLOAD delimited input & output
- SCOPE PENDING
- RUNSTATS non-uniform statistics on non-index columns
- System-level point-in-time backup and recovery
- Improved defaults for performance