



Insight2014

The Conference for Big Data and Analytics

DB2 Cancun Enhancements

George Baklarz, DB2 Technical Sales



#ibminsight

Please Note

- IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.
- Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.
- The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.
- The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.
- Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Agenda

- An Introduction to DB2 BLU Acceleration
 - Columnar Technology
- Recent Enhancements
 - DB2 BLU Improvements
 - pureScale Enhancements
 - Shadow Tables
- Summary



An Introduction to DB2 BLU Acceleration

IBM

Insight2014

The Conference for Big Data and Analytics

**SEIZE THIS
MOMENT**
▶▶▶▶

#ibminsight

Introducing BLU Acceleration



A new generation of data management innovation

- 8-25x faster reporting and analytics¹
 - More than 1000x seen in lab test queries²
- 10x storage space savings³
 - Seen during beta tests
- Delivers superior query acceleration at lower cost, on less hardware, and greater ease of use

BLU Acceleration



Speed of Thought Analytics

Super Fast, Super Easy — *Create, Load and Go!*

No Indexes, No Aggregates, No Tuning, No SQL changes, No schema changes

¹ Based on internal IBM testing of sample analytic workloads comparing queries accessing row-based tables on DB2 10.1 vs. columnar tables on DB2 10.5. Performance improvement figures are cumulative of all queries in the workload. Individual results will vary depending on individual workloads, configurations and conditions.

² Based on internal IBM tests of pure analytic workloads comparing queries accessing row-based tables on DB2 10.1 vs. columnar tables on DB2 10.5. Results not typical. Individual results will vary depending on individual workloads, configurations and conditions, including size and content of the table, and number of elements being queried from a given table.

³ Client-reported testing results in DB2 10.5 early release program. Individual results will vary depending on individual workloads, configurations and conditions, including table size and content.

Seamless Integration into DB2

- **Built Seamlessly into DB2 – Integration and Coexistence**

- Column-organized tables can coexist with existing tables
 - Same schema, same storage, same memory
- Integrated tooling support
 - Optim Query Workload Tuner recommends BLU Acceleration deployments

- **Same SQL, Language Interfaces, Administration**

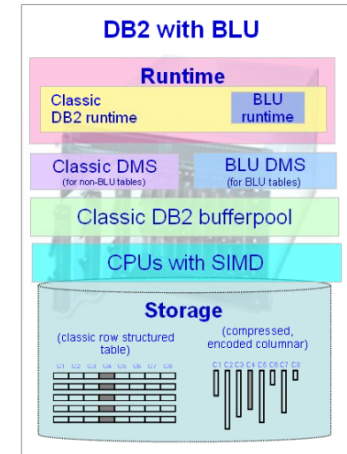
- Column-organized tables and row-organized tables can be accessed within the same SQL statement

- **Dramatic Simplification – Just “Load and Go”**

- Faster deployment
 - Fewer database objects required to achieve same outcome
- Requires less ongoing management due to its optimized query processing and fewer database objects required

- **Simple Migration from Row to Column-organized**

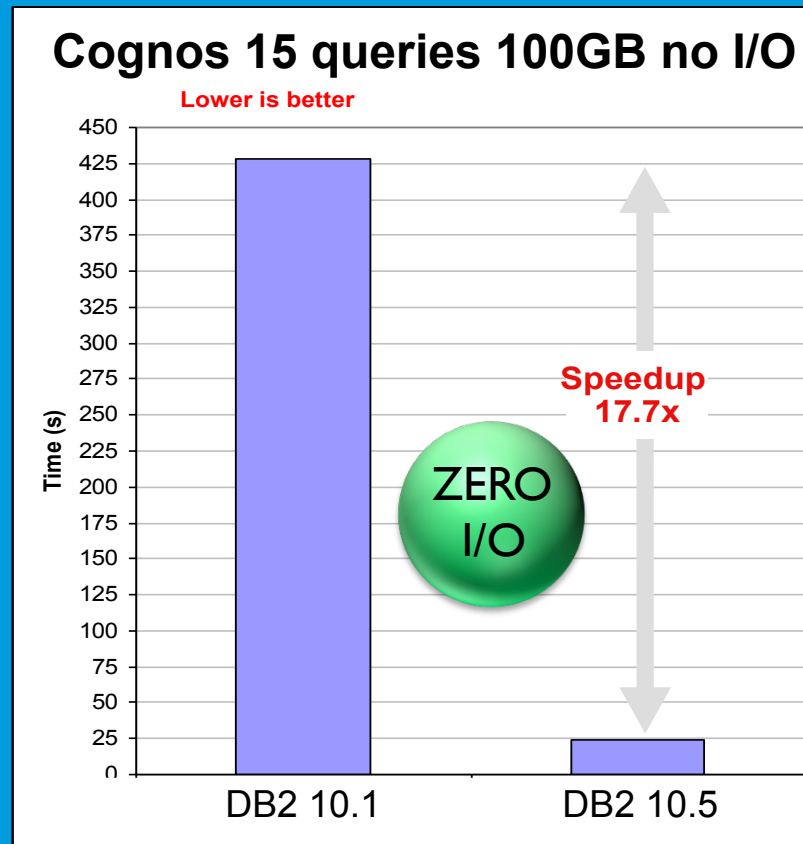
- Built-in db2convert utility



In-memory Isn't Everything

Great performance takes a lot more than just “in-memory”

Performance when all data is cached
Zero I/O in both DB2 10.1 and DB2 10.5
with BLU Acceleration¹



IBM Research & Development Lab Innovations

BLU Acceleration



- **Dynamic In-Memory**

In-memory columnar processing with dynamic movement of data from storage data



- **Actionable Compression**

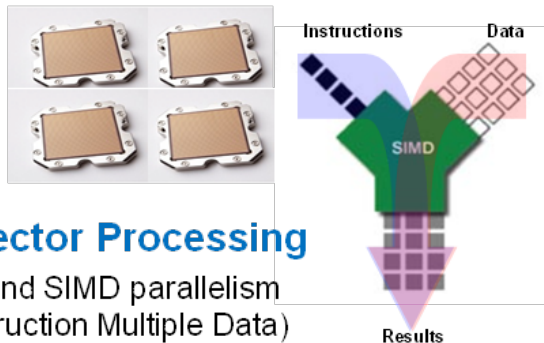
Patented compression technique that preserves order so that the data can be used without decompressing



Encoded

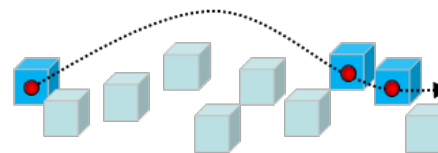
- **Parallel Vector Processing**

Multi-core and SIMD parallelism (Single Instruction Multiple Data)



- **Data Skipping**

Skips unnecessary processing of irrelevant data





DB2 BLU Enhancements

IBM

Insight2014

The Conference for Big Data and Analytics

**SEIZE THIS
MOMENT**
▶▶▶▶

#ibminsight

Recent DB2 Announcements

- **BLU Acceleration**
 - Performance improvements across the board
 - New Shadow tables provide BLU Acceleration for reporting while maintaining OLTP transactional performance
 - Oracle compatibility and extended SQL support
- **pureScale**
 - Online Table reorg, Incremental backups, improved administration and more flexible recovery
 - New virtualization options allow for a lower cost of entry by using Socket-based communication
- **SAP BW Support**
 - Additional BW Object support with significant performance improvements

BLU Acceleration Enhancements

- More Oracle ISV Enablement for BLU
 - Easier/Faster Migration of Oracle Data Marts to DB2
- Support for a Broader Range of Workloads
 - More ETL functionality
 - Significant I/U/D performance improvement
 - More complex query support
 - Broader support for complex joins, nested joins, etc.
 - Better High Availability and DR Support
- POWER8 Optimized
- Greater SAP Support
 - SAP BW DSO enablement
- Shadow Tables
 - 10x or more acceleration of OLAP style queries implemented on OLTP workload data servers and minimum impact on OLTP performance

Improved Oracle Compatibility

- Data types
 - DATE data type (Oracle Semantics)
 - NUMBER data type (Oracle Semantics)
 - VARCHAR2 data type (Oracle Semantics)
- Features and capabilities
 - DUAL
 - OUTER JOIN OPERATOR (+)
 - TRUNCATE TABLE
 - CHARACTER LITERALS
 - COLLECTION METHODS
 - PL/SQL compilation
 - Oracle data dictionaries
 - Oracle database links
 - INSENSITIVE cursors
 - INOUT parameters
 - SQL Data-Access-level enforcement



Improved Performance of Extract/Load/Transform

- Significant Insert/Update/Delete and Extract/Load/Transform Performance Improvements
 - Further support for primary key index exploitation for point queries and UPDATE/DELETE
 - Optimizations for batch UPDATE/DELETE statements
- Significant performance improvements for INSERT/UPDATE
 - Update with IN-list predicate is significantly faster
 - In FP1, UPDATE with IN-LIST was 14x slower than row store
 - In DB2 FP4, UPDATE with IN-LIST is up to 3x faster than row store
 - Fast INSERT = up to 1.8x faster than FP3
 - Fast UPDATE = up to 70x faster than FP3
 - Canadian University reported one single UPDATE statement was 84x faster than FP3
 - European client reported 40x faster UPDATE within INGEST compared to FP3
 - Fast DELETE = 26x faster than FP3

More Compression and Simplicity

- Automatic data sub-setting during LOAD ANALYZE phase
 - New sampling and data limits, with intelligent defaults
 - Sampling 1% of the data results in similar compression rates
 - Reduced memory consumption from the UTILITY HEAP
 - Significant CPU reduction
- Adaptive compression for INSERTs
 - LOAD processing generates exceptional compression via table and page-level dictionaries
 - INSERT/INGEST statements already create table-level dictionaries but now generate page-level dictionaries
- Enhanced compression for VARCHAR data
 - Detection of patterns throughout the data



More Performance

- Significant query performance enhancements
 - Complex, nested joins, and other joins now run in column engine
 - Common table expressions
 - Joins involving VARCHAR data
 - Data skipping for VARCHAR and CHAR predicates
 - CHAR and VARCHAR now available in synopsis table
- Support for MERGE statement (UPSERT only)
- Functional enhancements
 - ADD COLUMN support for columnar tables
 - Federation support for databases with columnar tables



Early Performance of DB2 Cancun Release

The logo for Cognos, featuring the word "COGNOS" in a bold, red, sans-serif font. The letter "O" is stylized with a white grid pattern.

30% faster

The logo for SAP Business Warehouse (BW), featuring the word "SAP" in large white letters on a blue background, with "BW" in smaller blue letters below it.

40% faster

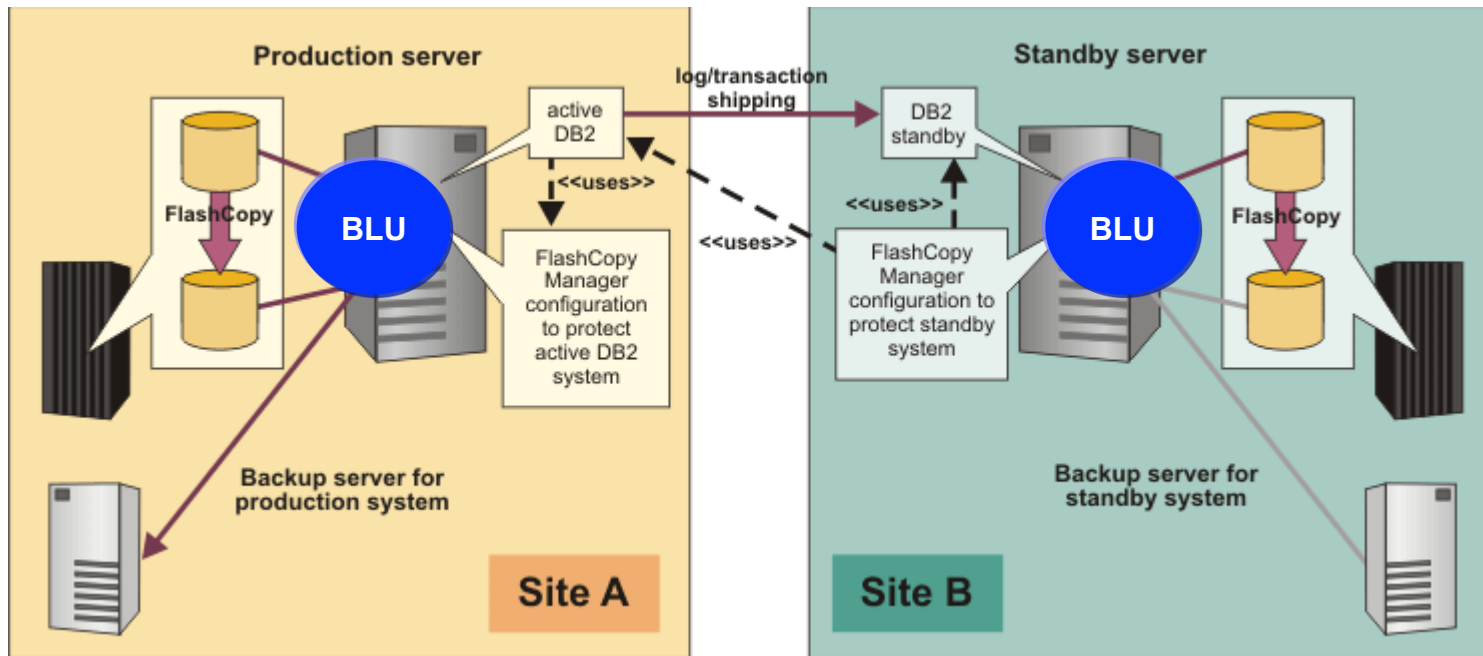
**Industry
Benchmark**

50% faster

Note that this is early beta testing on DB2 Cancun Release and we expect to have more improvements over these numbers. Your results may differ.

Simple HA and DR Solution for BLU

- BLU Acceleration with HADR availability
 - Highly available analytics
 - Use for both HA and DR
 - Includes most HADR capabilities
 - All synchronization modes, multiple standby, time delay, and log spooling
 - Read-on-Standby not supported in this release



Extreme Performance via Deep POWER8 Exploitation

- Performance

- More threading over Power7
- New 128-bit register instructions
- Vector processing of decimal data

- Faster range predicates for BLU tables

- POWER8 has new instructions that can be exploited by SIMD aware applications

- Reliability

- High-speed Data Page memory checking

- Cognitive compilation

- When compiling and optimizing DB2 runtime code, IBM uses special cognitive algorithms that watch DB2 processing BLU Acceleration workloads



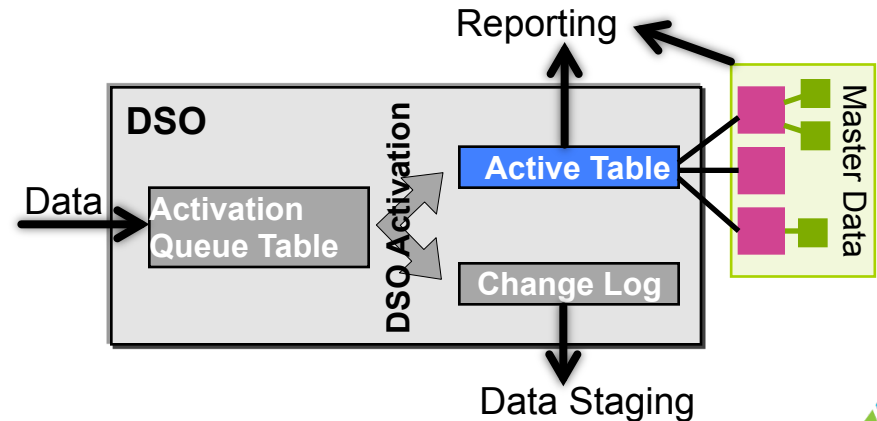
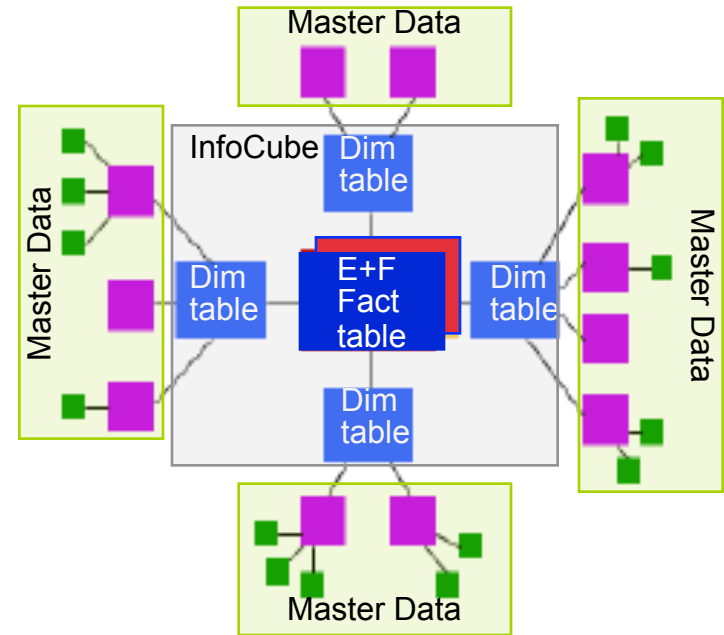
SAP BW Support in DB2

• DB2 10.5 FP1 Support

- Standard InfoCube
- Non-Cumulative InfoCube
- InfoCubes in Multi-Providers
- Semantical partitioned InfoProviders
- Near-Line Storage

• DB2 10.5 Cancun Support

- Standard DSO *
- Master Data *
- Flat InfoCube *
- InfoSet
- Write-Optimized DSO
- Persistent Staging Area (PSA)
- Transactional InfoCube
- Direct Update DSO





DB2 pureScale Enhancements

IBM

Insight2014

The Conference for Big Data and Analytics

**SEIZE THIS
MOMENT**
▶▶▶▶▶

#ibmin insight

pureScale with TCP/IP Interconnect (pureScale "Lite")

- TCP/IP (sockets) interconnect for faster cluster setup and lower cost deployments using commodity network hardware
 - Provides exactly the same level of high availability as RDMA-based pureScale environments
 - Appropriate for small clusters with moderate workloads where availability is the primary motivator for pureScale
- New DBM configuration parameter `CF_TRANSPORT_METHOD` gets set to either TCP or RDMA at instance creation time
 - Can be manually changed after the fact if the interconnect changes
- Prerequisites
 - 10 Gigabit Ethernet strongly recommended for production installations
 - Typical performance impact of 10GE TCP/IP is 30% relative to RDMA-based interconnect
 - Set `DB2_SD_ALLOW_SLOW_NETWORK=YES` for VMware or 1GE implementations
 - Hosts must be on the same subnet and must be able to ping each other

Virtualized Deployments of DB2 pureScale

- Previously, virtualized deployments of pureScale were limited to:
 - AIX LPARs, with dedicated RDMA network adapters per partition
 - KVM with RHEL, with dedicated 10 GE RoCE network adapters per partition
- Additional virtualized deployments now available in DB2 Cancun Release with TCP/IP (sockets) interconnect, including:
 - AIX LPARs
 - VMware (ESXi, vSphere) with RHEL or SLES
 - KVM with RHEL
- Virtualized environments are perfect for:
 - Development
 - QA and testing
 - Production environments with moderate workloads
 - Getting hands-on experience with pureScale



Virtualized Deployments: Supported Configurations

Operating System	Virtualization Technology	InfiniBand Supported?	10GE RoCE Supported?	TCP/IP Sockets Supported?
AIX, SLES, RHEL	No virtualization (bare metal)	Yes *	Yes *	Yes
AIX	LPAR	Yes *	Yes *	Yes
SLES	VMware	No	No	Yes
RHEL	VMware	No	No	Yes
	KVM	No	Yes *	Yes

* Dedicated interconnect adapter(s) per host/partition

- **VMware supported with**
 - Any x64 system that is supported by both the VM and DB2 pureScale
 - Any Linux distribution that is supported by both the VM and DB2 pureScale
- **KVM supported with**
 - Any x64 system that is supported by both RHEL 6.2 and DB2 pureScale
 - RHEL 6.2 and higher

Virtualized Deployments: VM Storage Configurations

Disk configuration	KVM hypervisor	VMware ESX/ESXi	Tiebreaker and I/O fencing
Virtual disks ¹	Yes	No ²	No ³
RDM disks in Physical Mode ⁴	No	Yes	Yes
SAN disks ⁵	Yes	No	Yes

Note:

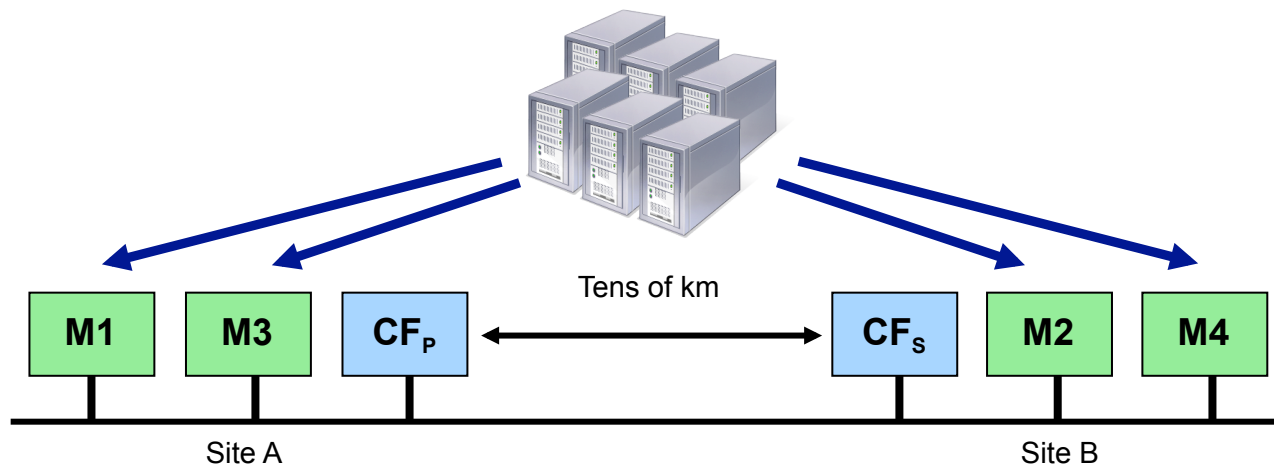
1. Virtual disks do not support SCSI-3 PR commands and cannot be used as tie-breaker disks. Virtual disks can be used to contain shared data.
2. Only supported in non-production environments.
3. I/O fencing requires SCSI-3 PR commands to be enabled, which are not supported on virtual disks.
4. Raw device mapping (RDM) disks are logical unit numbers (LUNs) that can be directly accessed from the VM guest operating system without going through a virtual machine file system (VMFS). RDM disk support is not available in KVM environments.

To support tie-breaker disk and SCSI-3 PR I/O fencing, each RDM disk must be assigned to only one virtual machine per physical server.

5. You can assign storage Fibre Channel (FC) adapters to the guest virtual machines by using the PCI device pass-through mode. After you assign storage adapters, you can directly access storage area network (SAN) disks from inside the guest VM. Tie-breaker disks and SCSI-3 PR I/O fencing are supported in this environment.

Additional GDPC Configurations

- A “stretch” or geographically-dispersed pureScale cluster (GDPC) spans two sites A and B at distances of tens of kilometers
 - Provides active/active access to one or more shared databases across the cluster
 - Enables a level of DR support suitable for many types of disasters (e.g. fire, localized power outage)



- GDPC now supports the following configurations
 - AIX with InfiniBand (IB) network
 - AIX with 10 Gigabit Ethernet (10GE) RoCE network **(new in DB2 Cancun Release)**
 - RHEL with 10 Gigabit Ethernet (10GE) RoCE network
 - SUSE with 10 Gigabit Ethernet (10GE) RoCE network **(new in DB2 Cancun Release)**

Support for IBM POWER8 Hardware

- Take advantage of the latest generation of POWER8 processors, with game-changing innovation that accelerates big data and analytics
- POWER8 support for pureScale implementations using
 - TCP/IP sockets
 - 10GE RoCE
 - Using 10GE RoCE PCIe gen 2 adapters that are supported today with POWER7
 - QDR InfiniBand (post- DB2 Cancun Release)
 - Using QDR IB PCIe gen 2 adapters that are supported today with POWER7
 - 40GE RoCE (post- DB2 Cancun Release)
 - Using 40GE PCIe gen 3 adapters



In-place (Online) Table Reorganization

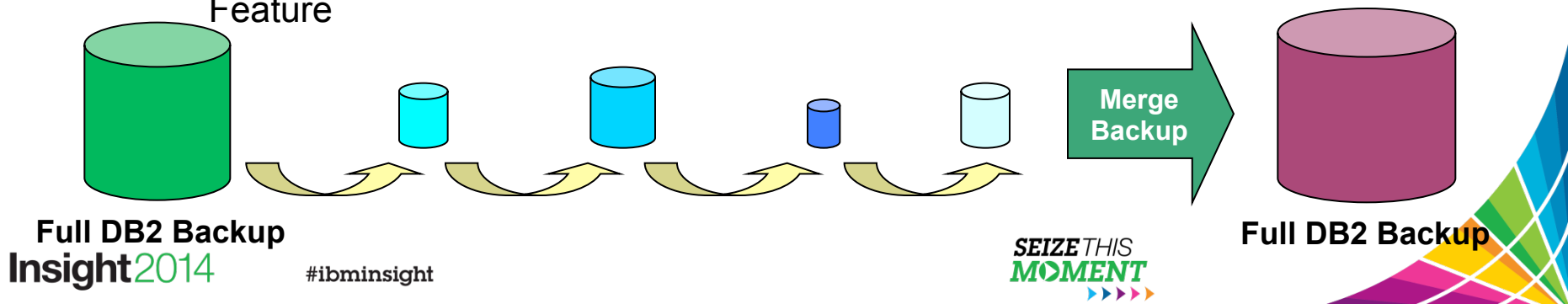
- Online table reorganization now fully supported in pureScale
 - Reclaim free space
 - Eliminate overflows
 - Re-establish clustering
- Example:

```
REORG TABLE <tableName> INPLACE ALLOW READ ACCESS
```



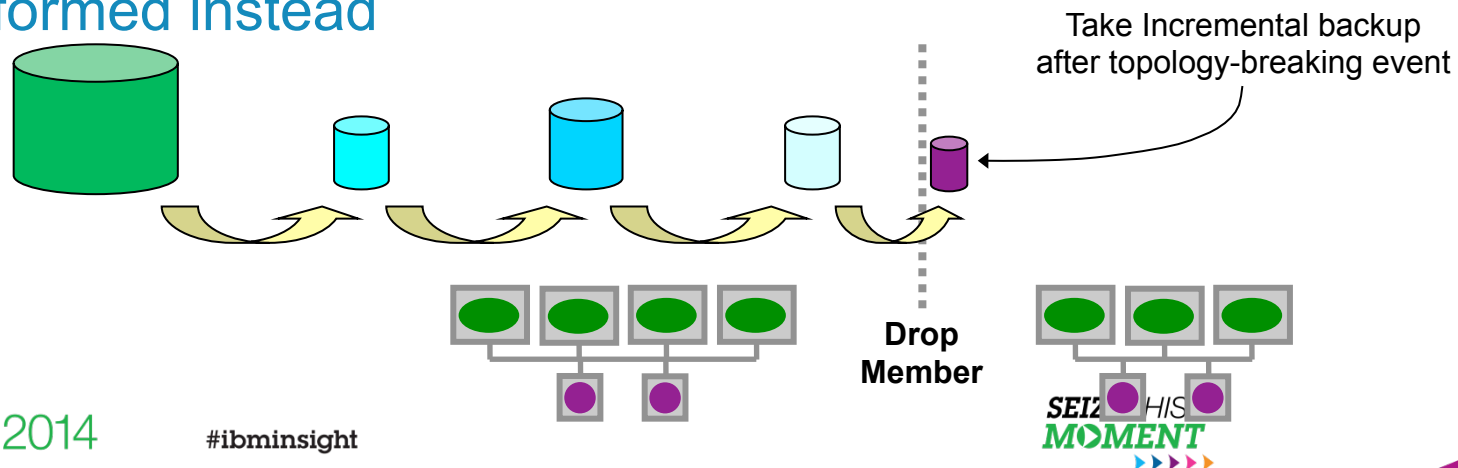
Incremental Backup/Restore and DB2 Merge Backup

- **Incremental backups now supported for pureScale**
 - Allows for smaller backup images, as unchanged data not backed up
 - Applicable to database-level or table space-level backups
 - Enabled via `TRACKMOD` database configuration parameter
- **Two types of backups**
 - Incremental: Copy of all data that has changed since the most recent, successful, full backup operation (also known as cumulative backup)
 - Delta: Copy of all data that has changed since the last successful backup of any type (full, incremental, or delta) (also known as a differential backup)
- **Support for pureScale in DB2 Merge Backup v2.1 FP1 (shipped in parallel with DB2 Cancun Release)**
 - The Merge Backup utility combines an older full backup with subsequent incremental and delta backups to create a new full backup image
 - Tool available separately but also included in the IBM DB2 Advanced Recovery Feature



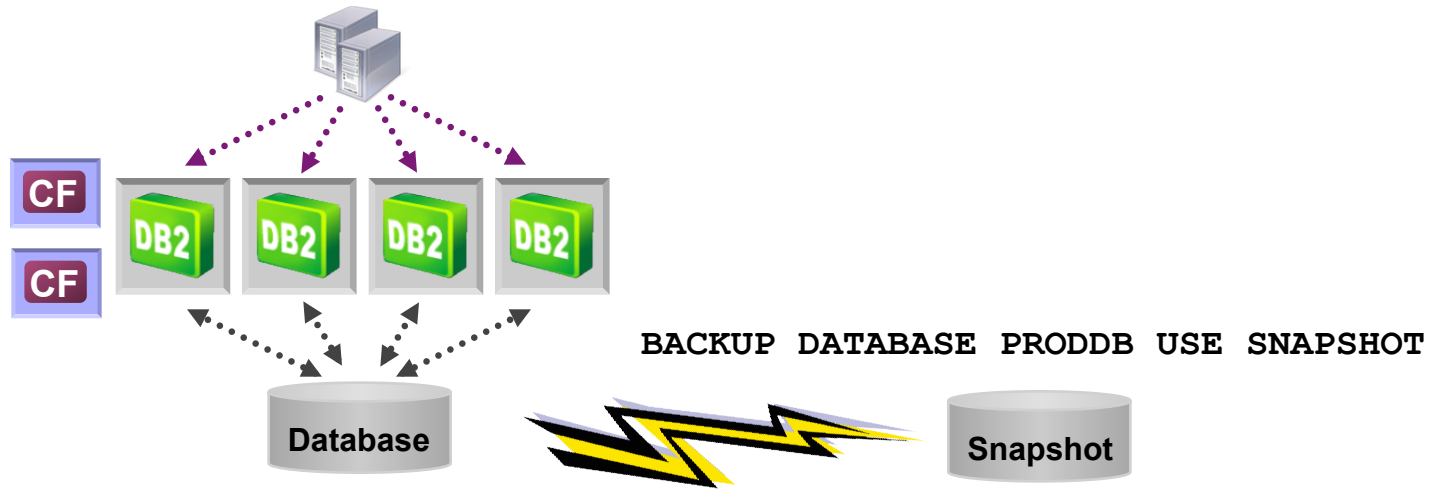
Database Topology Changes with Incremental Backup

- Certain operations are considered "topology-breaking" for a database
 - Drop member from cluster
 - Restore database backup to a cluster with a subset of the members
 - Restore non-pureScale database backup into pureScale instance
 - Restore pureScale database backup into non-pureScale instance
- Previously, a full offline database backup was required following these events to provide a new recovery starting point for the database
- Now, an incremental offline database backup can be performed instead



Integrated Snapshot Backups

- Backup large pureScale databases very fast!
 - In DB2 Cancun Release and DB2 10.1 FP5



- Alternative methods previously available:
 - Manual snapshot process
 - SET WRITE SUSPEND
 - {storage-level snapshot commands}
 - SET WRITE RESUME)
 - Snapshot backup scripts (added in DB2 10.5)
 - E.g. `BACKUP DATABASE PRODDb USE SNAPSHOT SCRIPT '/scripts/snapshot.sh'`

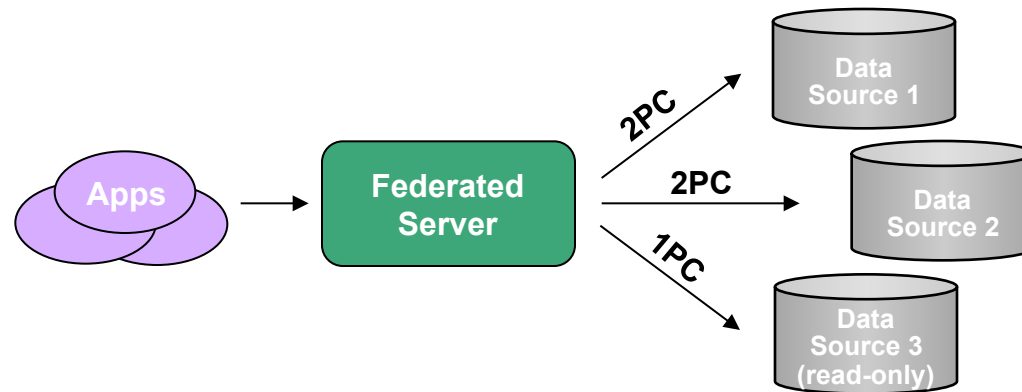
OPM Enhancements for pureScale

- Wait time metrics in SQL Statements Dashboard and Overview Dashboard expanded to include
 - Cluster Caching Facility (CF) wait time
 - Amount of time spent communicating with the CF (not including time spent by CF performing processing)
 - Page reclaim wait time
 - Amount of time spent waiting on page locks, where the lock request caused a page to be reclaimed
- New alerts related to pureScale environments with HADR
 - HADR Primary Cluster Member Disconnected
 - HADR Primary Cluster Disconnected
 - HADR Standby Cluster Not Ready for Role Switch
 - HADR Primary Cluster Member Logger Blocked
 - HADR Primary Cluster Member Logger Slow
 - HADR Standby Cluster Log Receiver Falling Behind
 - HADR Standby Cluster Log Replay Falling Behind
 - HADR Standby Cluster Member Log Receiving Buffer Utilization



Federated Two Phase Commit (2PC) Support

- Federated two phase commit (F2PC) allows for insert/update/delete against different remote data sources within a single transaction



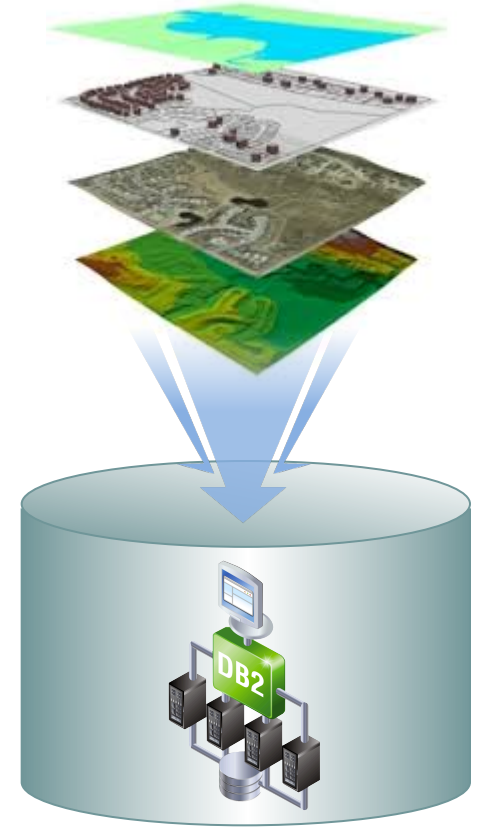
- Previously, F2PC not supported with DB2 pureScale database server acting as a federated server
 - Data sources could not be defined with the two phase commit option

```
create server oracserv type oracle version 10.1 wrapper oracwrap
... options(node 'ora10node1', password 'Y', pushdown 'Y',
           db2_two_phase_commit 'Y')
alter server udb1 options (add db2_two_phase_commit 'Y')
```

Both fail with: SQL1881N "DB2_TWO_PHASE_COMMIT" is not a valid "SERVER" option

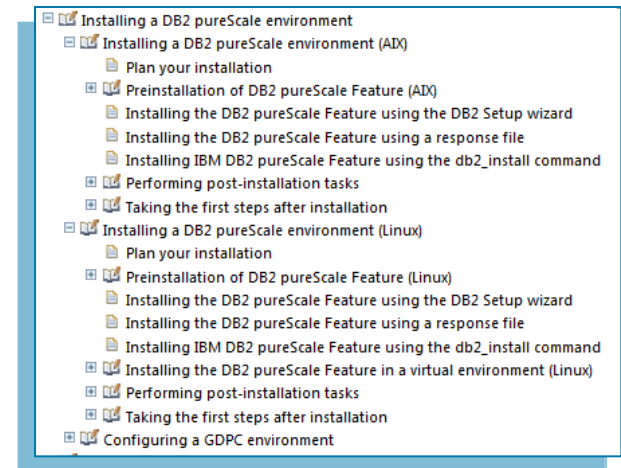
DB2 Spatial Extender

- You can now store, manage, and analyze spatial data in a DB2 pureScale environment
- No difference between pureScale and non-pureScale in terms of setting up Spatial Extender and creating a project that uses spatial data



Improved Serviceability and Usability

- Restructured install topics in Information Center,
 - Some sections completely rewritten for improved clarity and flow
 - Topics have been rearranged to provide an order to the install process
 - Topics now grouped together by OS
 - New "planning" topics added for installation
- Enhanced prerequisite checking for pureScale
 - Improved checks for network adapters
 - New checks test actual functioning of RDMA and Ethernet capabilities, in addition to checking configuration settings



Improved Serviceability and Usability (cont.)

- Improved online (rolling) fixpack update usability
 - Through robust cluster status validation and enhanced error reporting that will avoid unplanned outages
 - Don't allow primary CF to be brought down if secondary CF still in catchup state
 - Don't allow only active member to be brought down for an online update
 - Block online add member during a rolling update
 - Block cluster removal while update is in process
 - Improved error messaging around invalid parameters and space validation
- Option to force `installFixPack` command to overwrite manually installed or updated GPFS, TSA, or RSCT components with new version included in DB2 fix pack
 - New "`-f TSAMP | RSCT | GPFS`" option

Parallelized DB2 Instance Upgrade of Members and CFs

- Reduced downtime for version level upgrade of DB2 pureScale instance by performing upgrade of members and CFs in parallel
- Steps involved:
 1. Basic validations for instance upgrade to be done on install-initiated host (IIH)
 2. Mandatory upgrade steps that must be completed on IIH before moving onto upgrade of other hosts
 3. Parallelized instance upgrade on the rest of the hosts in the cluster
 4. Instance upgrade results from all hosts are summarized at completion time
- New global option `-g` added to `db2iupgrade` command to enable this behavior
 - Default for pureScale instances
 - Not default for non-pureScale instances (local host upgrade is the default)

DB2 BLU Shadow Tables

IBM

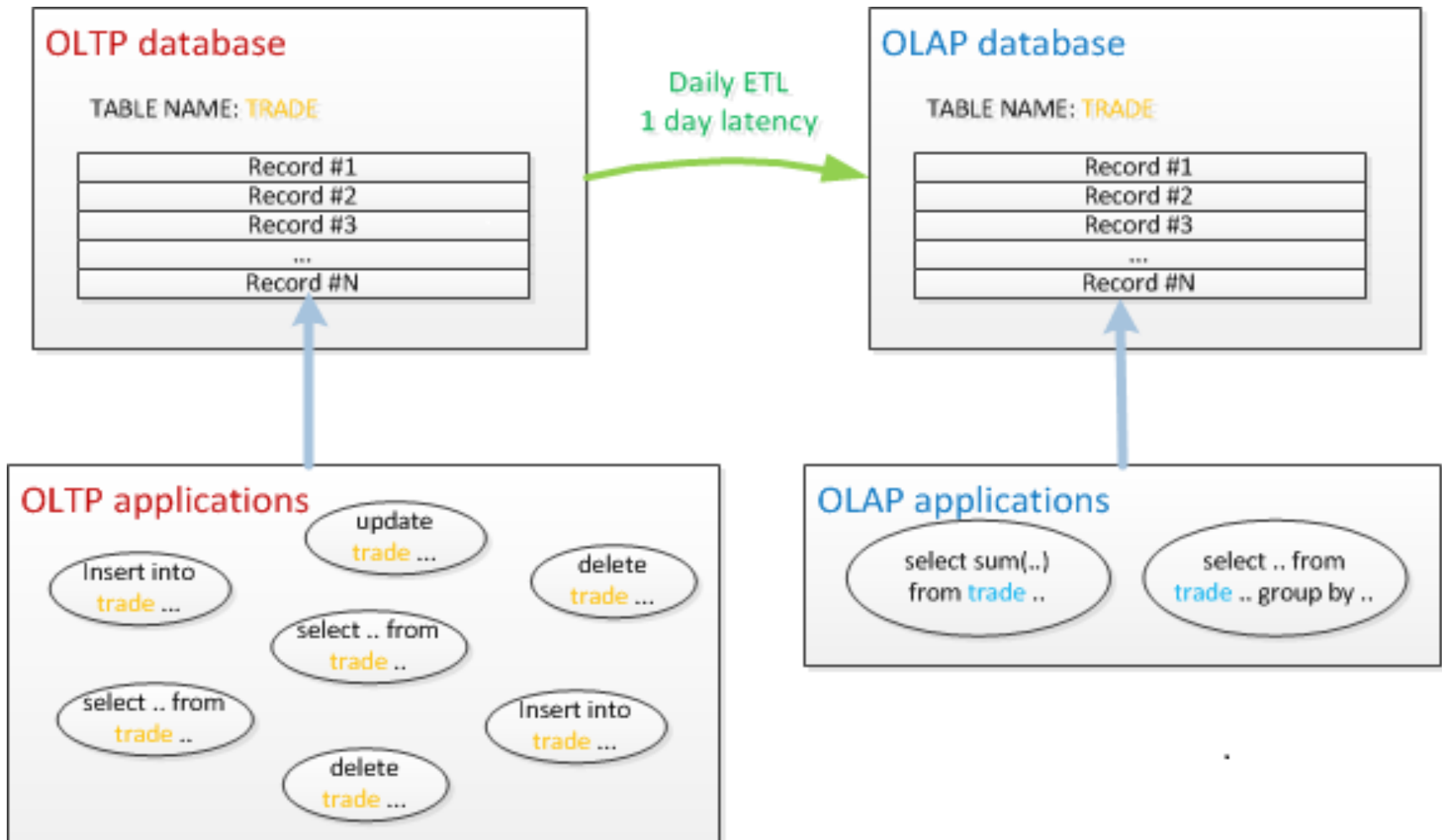
Insight2014

The Conference for Big Data and Analytics

**SEIZE THIS
MOMENT**
▶▶▶▶

#ibminsight

High Level Architecture: Traditional OLAP



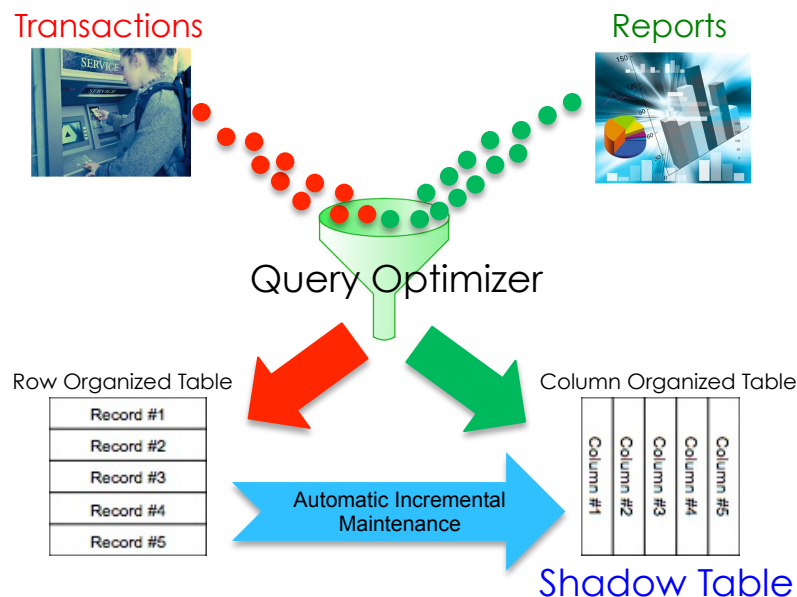
DB2 Shadow Tables

- Transparent DB2 BLU “Shadow Table”

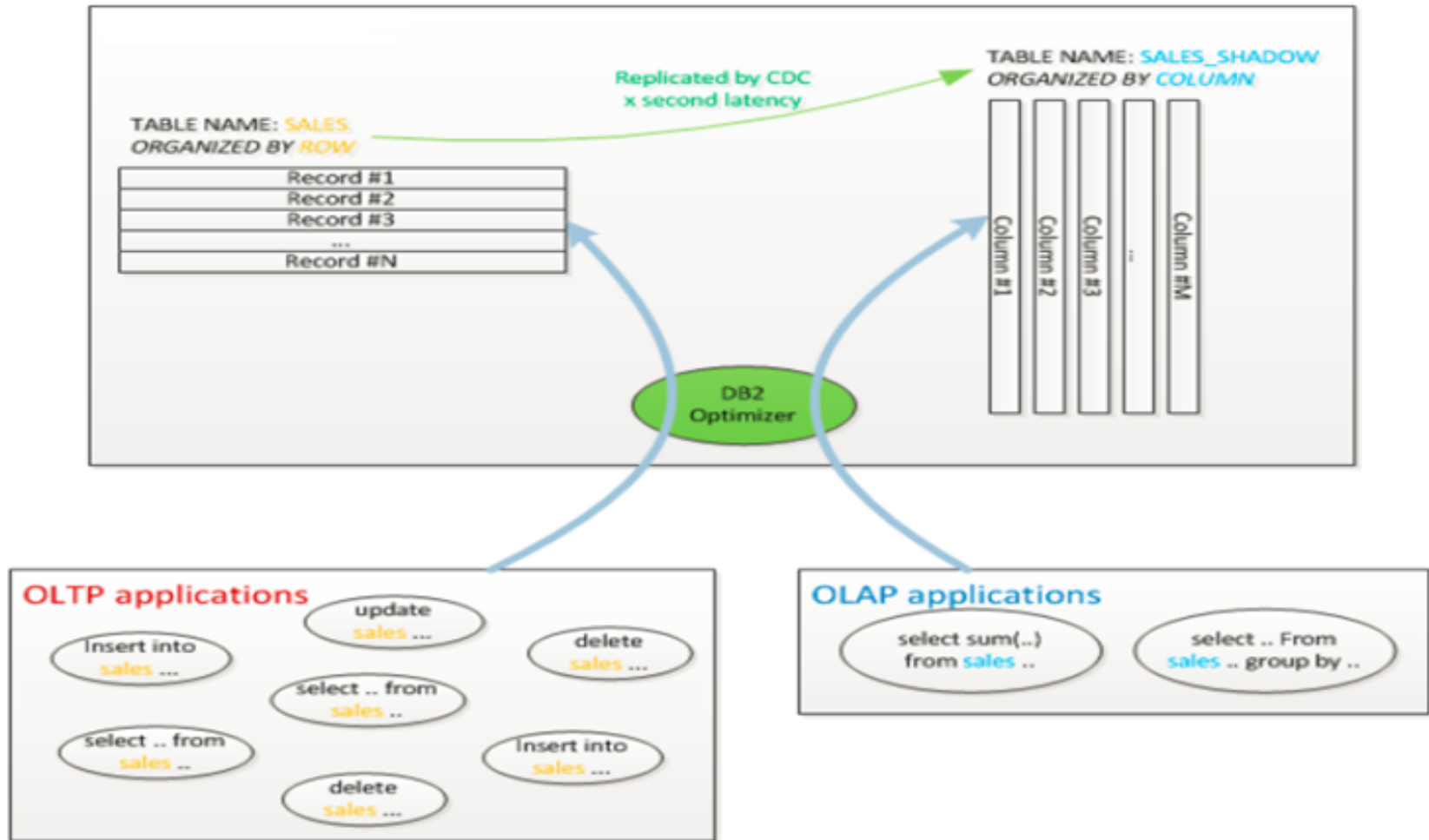
- A new way to optimize queries using a columnar data store
- Shadow Tables are a projection or a full copy of a base row organized table stored in column organized format

- Powered by DB2 BLU Acceleration

- Analytical queries issued against the normal row based table are diverted to shadow tables to take advantage of BLU Acceleration
- DB2 optimizer decides when to route queries to row or column organized table
- Data in the shadow table is kept up to date with the base table using CDC

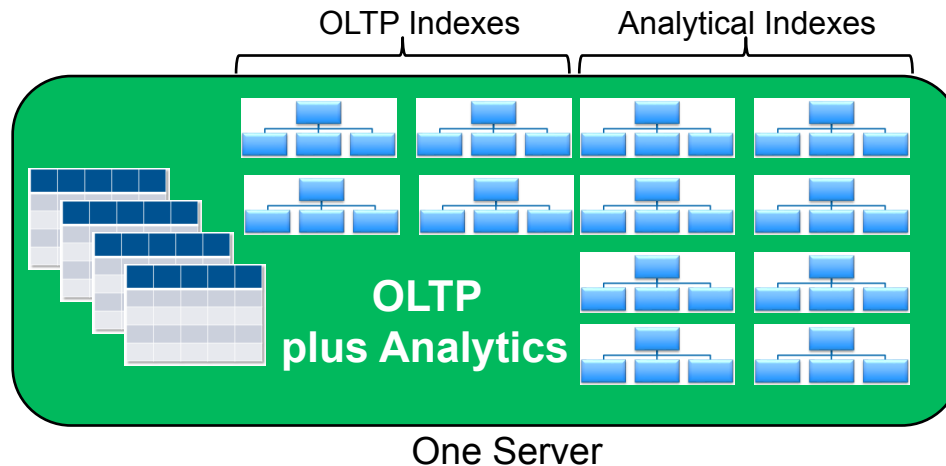


High Level Architecture: DB2 Shadow Tables



Shadow Tables – Simplified Administration

- OLTP system with OLTP indexes and several extra indexes to support reporting queries

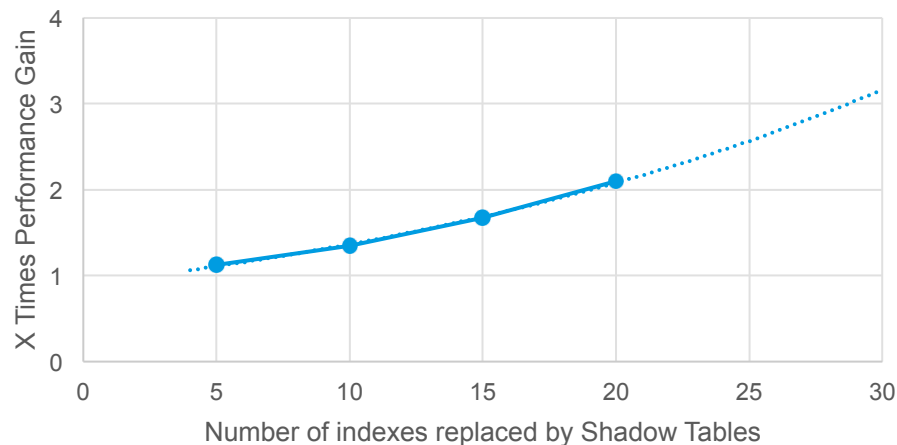


- Shadow Tables simplify administration and boost performance
 - Greatly reduces indexes needed for performance – no analytical indexes

Shadow Tables – Improved Performance

- With only 4 OLAP indexes replaced by BLU Shadow Tables, performance of OLTP queries is the same
 - No additional impact to OLTP when using Shadow Tables
 - 10x less memory consumed for indexes with Shadow Tables
 - Other vendors are talking about replacing 10-20 indexes (not typical with OLTP Systems)
 - Reporting queries >10x faster with BLU vs. running reports on base OLTP tables¹

OLTP Performance Gain with Shadow Tables



1) Lab results and may not be representative of the performance you may achieve.

Shadow Table - Creation

- A Shadow Table is a column-organized projection (potentially with subset of columns) of a row-organized table that is user-defined
 - ```
CREATE TABLE TRANS_FACT_shadow
 (select c1, c3 from TRANS_FACT)
 data initially deferred refresh deferred
 enable query optimization maintained by
 replication organize by column;
```
- After creation, Shadow Tables will be put in SET INTEGRITY PENDING state
  - Need to issue SET INTEGRITY .... IMMEDIATE UNCHECKED to bring the Shadow Table out of pending state

# Shadow Table – Creation (Cont)

- Just like any other user-maintained MQTs, SET INTEGRITY .... IMMEDIATE CHECKED and REFRESH TABLE statements are not allowed on a Shadow Tables
- The primary key on the base table must be included in the select list of the Shadow Table
  - The primary key on the Shadow Table is required to provide a one-to-one mapping for each row in the base table to the corresponding row in the Shadow Table.
  - The primary key also facilitates maintenance of the Shadow Table

# Restrictions

- Only one Shadow Table per row-organized base table is allowed
- A Shadow Table must be
  - Column-organized (must be specified on creation only) and MAINTAINED BY REPLICATION must be specific on creation
    - If an MQT is maintained by REPLICATION (Shadow Table), it must be column-organized
      - The keywords ORGANIZE BY COLUMN must be specified
      - DFT\_TABLE\_ORG is ignored
  - Must be a REFRESH DEFERRED MQT. REFRESH IMMEDIATE is not supported
- For Shadow Tables to be considered in query routing, all row-organized tables in a query must have Shadow Tables defined

# Restrictions

- In the query definition of the Shadow Table, the following will be enforced
  - Only one base table can be referenced (i.e., no join is allowed)
  - Base table must be a row-organized table
  - The projection list can only reference base table columns that are valid in a column-organized table (no LOBs, XML, etc.)
    - No expression is allowed in the projection list
    - The columns cannot be renamed through the column list or AS clause
  - The projection list must include at least one set of enforced unique constraint or primary key columns from the base table
  - The query definition of the Shadow Table cannot include
    - References to a nickname, a typed table, or a view
    - SELECT DISTINCT
  - Only a sub-select consisting of a select-clause and a from-clause is allowed

# Recognizing a Shadow Table

- A new flag (position 23) in the PROPERTY column of SYSCAT.TABLES is used to indicate that the table is a Shadow Table

| Column Name | Data Type    | Nullable | Description                                                                                                                                                                                                                                                                                                      |
|-------------|--------------|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PROPERTY    | VARCHAR (32) | -        | <p>Properties for a table.<br/>A single blank indicates that the table has no properties. The following is position within string, value, and meaning</p> <p>1, Y = User maintained materialized query table<br/>....<br/>20, Y = Column-organized table<br/>21, Y = Synopsis table<br/>23, Y = Shadow Table</p> |

- `select varchar(tabname, 20) as tabname, substr(property, 23,1) as Shadow,tableorg from syscat.tables where tabname like 'T1%';`

```
TABNAME SHADOW TABLEORG

T1 R
T1_SHADOW Y C
```

# INSERT/UPDATE/DELETE – Shadow Tables

- To allow replication tools to apply IUD transactions to the MQT, a Shadow Table MQT is also defined as a user-maintained MQT
  - Similar to regular user maintained MQTs, IUD statements on the Shadow Table by users are allowed but not recommended
  - To avoid inadvertent changes to Shadow Tables, which are maintained by an external replication product (CDC), you are advised to revoke INSERT, UPDATE, and DELETE privileges on these tables.
  - Write permissions on these tables should only be allowed on the DB2 userid used to configure the external replication product



# INSERT/UPDATE/DELETE – JDBC Batching

- As a replication tool, Change Data Capture (CDC) will use JDBC batching to apply groups of single IUD statements to the target Shadow Table
  - This will trigger the fast direct column-organized IUD and array IUD optimization that were also added in DB2 Cancun Release

# INSERT/UPDATE/DELETES – CDC Latency

- The current design will buffer INSERTs and flush when the buffer is full or at a transaction boundary. We also create the SYNOPSIS table at a transaction boundary
  - It is, therefore, beneficial, for replication to avoid small APPLY transactions to the column-organized MQT table
  - To address this, CDC provides a new CDC system parameter, `acceptable_latency_in_seconds_for_column_organized_tables` to allow buffering of data on the replication end.
  - It is recommended that `acceptable_latency_in_seconds_for_column_organized_tables` be set to a value smaller than CURRENT REFRESH AGE setting



# INSERT/UPDATE/DELETE – Primary Key

- For CDC APPLY logic to perform the UPDATE and DELETE operations efficiently, enforced primary key is supported on Shadow Tables
  - The main purpose of the primary key on the Shadow Table (that matches the primary key or unique key constraint of the source table) is for efficient INSERT/UPDATE/DELETE during CDC APPLY
- The primary key column set on the Shadow Table must correspond to the primary key or unique constraint column set on the base table
  - A dependency between the Shadow Table primary key index and the base table index for the associated primary key or unique constraint will be created
  - Dropping the base table primary key or unique constraint that has an associated primary key on a Shadow Table is not allowed

# Query Routing

- Shadow Tables will be considered only for dynamic queries
- Analytical queries against the OLTP (row-organized) tables will be automatically routed to Shadow Tables to take advantage of BLU Acceleration if the replication latency is within an acceptable limit specified by users and is still a cost-based decision
- All tables that are referenced within a query must have associated Shadow Tables; otherwise, none of the Shadow Tables will be considered for use
  - Latency-based routing is limited to Shadow Tables only
- Once we have determined that a query can make use of a Shadow Table, the final decision on whether to route the query to the Shadow Table or not, is still cost based



# New SYSTOOLS Schema Table - REPL\_MQT\_LATENCY

- REPL\_MQT\_LATENCY has the following format. It is populated by the replication software

| Column Name  | Data Type | Description                                                                                                                                                       |
|--------------|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| COMMIT_POINT | BIGINT    | The last commit time after applying change to target Shadow Table. It is the timestamp (in seconds since epoch in UTC) when the DELAY_OFFSET value was generated. |
| DELAY_OFFSET | BIGINT    | The number of seconds between the time when the source table data is read and the last commit time of applying changes to target Shadow Table                     |

# New SYSTOOLS Schema Table - REPL\_MQT\_LATENCY

- Using these two values, DB2 will compute the refresh timestamp during its query processing
- $\text{REFRESH\_TIMESTAMP} = \text{COMMIT\_POINT} - \text{DELAY\_OFFSET}$   
After the CDC UPDATE is committed, the transactions on the source table of the Shadow Table that had been committed before the REFRESH TIMESTAMP are guaranteed to have been applied and committed in the Shadow Table
- During query processing, DB2 compares REFRESH TIMESTAMP with the system timestamp and CURRENT REFRESH AGE special register to decide if the query should be routed the Shadow Table
  - In the case of CDC, when the source tables have no INSERT/UPDATE/DELETE activities CDC periodically updates the timestamp to indicate that all tables are up to date

# Query Routing – Non Latency-Based Routing

- You can choose to have Shadow Tables be considered for optimization of queries that is not based on a latency period
- This might be the case if the degree of synchronization between a Shadow Table and its underlying base table is not especially important within the context of your business needs
- A Shadow Table defined with `ENABLE QUERY OPTIMIZATION` can be used to optimize the processing of queries if all of the following conditions are true
  - `CURRENT REFRESH AGE` special register is set to `ANY`
  - `CURRENT MAINTAINED TABLE TYPES FOR OPTIMIZATION` special register is set to `ALL` or `REPLICATION`
  - `CURRENT QUERY OPTIMIZATION` special register is set to `2` or a value greater than or equal to `5`

# Special Registers/DBM Configuration Parameters

- **CURRENT REFRESH AGE special register (CRA)**
  - The routing for Shadow Tables can be configured to depend on the latency between the data in the base table and in the Shadow Table. We will allow routing if the latency is within the interval specified in CURRENT REFRESH AGE special register
  - This special register can now take on values between 0 and ANY (999999999999999) if maintained by replication
    - Note you must first set the CURRENT MAINTAINED TABLE TYPES FOR OPTIMIZATION registry variable to REPLICATION (next page)
- **dft\_refresh\_age DB CFG parameter**
  - Also applies to the dft\_refresh\_age DB CFG parameter, which supplies the default value used for REFRESH AGE if CURRENT REFRESH AGE special register is not set



# Special Registers/DBM Configuration Parameters

- **CURRENT MAINTAINED TABLE TYPES FOR OPTIMIZATION special register (CMTTFO)**
  - For a Shadow Table to be considered for query routing, one prerequisite is the CURRENT MAINTAINED TABLE TYPES FOR OPTIMIZATION special register must have one of REPLICATION or ALL in the list it defines
- **dft\_mttb\_types DB CFG parameter**
  - It is currently possible to limit which MQT types can be used to optimize queries using the CURRENT MAINTAINED TABLE TYPES FOR OPTIMIZATION special register, which gets its default value from the dft\_mttb\_types database configuration parameter. Currently the special register can have the value ALL, NONE, or a comma-separated list of table types from SYSTEM, USER, and FEDERATED\_TOOL and a new type, REPLICATION.

# Valid Combinations of Special Registers

- Latency based routing will be limited to Shadow Tables
- For a Shadow Table to be used (not necessarily latency based routing), CMTTFO must be either REPLICATION or ALL
- If CMTTFO = REPLICATION, CURRENT REFRESH AGE can be any of 0, <intermediate value>, or ANY
  - When CRA is set to some intermediate value, latency-based routing is considered; Shadow Tables will be considered if their content is not too stale
  - When CRA is set to 0, deferred MQTs (including Shadow Tables) will not be considered; no latency based routing
  - When CRA is set to ANY, deferred MQTs (including Shadow Tables) will be considered (regardless of staleness of the Shadow Table data); no latency based routing. Routing to these deferred MQTs, including Shadow Tables, does not depend on staleness of MQT content
- If CMTTFO = ALL, CURRENT REFRESH AGE cannot be set to an intermediate value. CRA can only be 0 or ANY
  - Same implication as above. That is, no latency based routing

# EXPLAIN/ERROR Diagnostics for Shadow Tables

The `EXPLAIN` utility will generate extended diagnostic messages if Shadow Tables are not used in a query for the following reasons

- `EXP0053W` – MQT isolation level is not `UR` or `CS`
- `EXP0054W` – `CRA` is equal to zero
- `EXP0076W` – All row-organized tables in the query must have an associated Shadow Table
- `EXP0079W` – The following MQT was not used in the final access plan, because the plan cost with this MQT was more expensive or a better candidate was available
- `EXP0087W` - Replication latency is not within the limit defined
- `EXP0088W` - `CMTTFO` was not set to `REPLICATION` or `ALL`
- `EXP0089W` - Cannot route to Shadow Table if query is run on the read-on-standby system

# How Do Applications Use Shadow Tables

- It is important to separate applications into OLTP and OLAP-based on factors such as query complexity and acceptable staleness of data
- A Connect Procedure provides a general mechanism to enable applications to implicitly execute a specific procedure upon connection. It is a stored procedure that is called every time an application connects to a DB2 database
- Connect Procedures can perform any of the usual procedural tasks, such as issuing SQL statements
- You can use this approach to write custom logic to identify which connections are to enable the use of Shadow Tables for query optimization for your OLAP queries and applications
- You create a single connect procedure and specify the applications and environment you want to have setup.



# Shadow Tables - Customer Quotes

## DataProxy LLC

*“Our initial results using **DB2 with BLU Acceleration** were impressive; massive analytic query speed-up, up to **273x improvement**. With the **Shadow tables** that are part of the **DB2 Cancun Release**, we can now leverage that **great analytic performance** in our transactional ODS and let **DB2 do all the work for us. One database, one system, no extra effort on our part!**”* – Ruel Gonzalez - Information Services



*“In the telecom business, reporting on data as it is happening is critical to our clients success. We need to be able to **run complex reports directly on our transactional data** to provide a competitive advantage for our business. **Shadow Tables**, in the **DB2 Cancun Release**, **gives us that ability**. They are simple to create and very easy to manage and our end users are now **reporting directly on our OLTP data without having to add in a separate BI system.**”* – Paul Peters, Lead Database Administrator, VSN Systemen BV



# Thank You

**IBM.**

**Insight2014**

The Conference for Big Data and Analytics

**SEIZE THIS  
MOMENT**  
▶▶▶▶▶

**#ibminsight**