
DB2 10.5 for LUW and BLU Acceleration Cancun Release (10.5.0.4)



Chris Eaton
Worldwide Information Management Technical Specialist
IBM Toronto Lab
ceaton@ca.ibm.com

The Next Generation Data Platform

Transact

- Provide high levels of service without the high price
- Seamlessly expand or contract as needed, paying only for what you use when you use it
- Enhance customer experience by delivering data when and where it's needed

Innovate

- Free your applications from database complexity for faster delivery
- Transform your ability to make business decisions with 35x to 73x faster analytics, with some queries running more than 1400x faster^{1,2}
- Optimize IT resources and utilization with built-in simplicity and autonomies

Analyze

- Leverage next generation in-memory technology
- Get instant insight into operational and warehouse data without compromising performance of either
- Deploy analytics solutions faster with “load and go” and operational simplicity

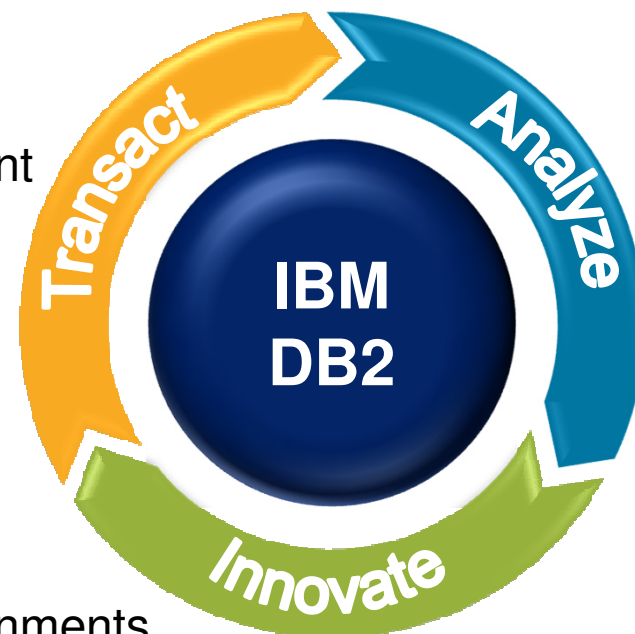


Introducing DB2 “Cancun Release”

New

Get the answers you need in the moment

- **Get real-time answers to ALL questions with instant insight into historical and “as it happens” data**
 - Change the economics of continuous availability with broad infrastructure choices at every price point
 - Improve performance of transactions and analytics while reducing complexity and overhead
 - Get fast time-to-value using skills you already have for Oracle database
 - Superior performance at lower cost for SAP environments



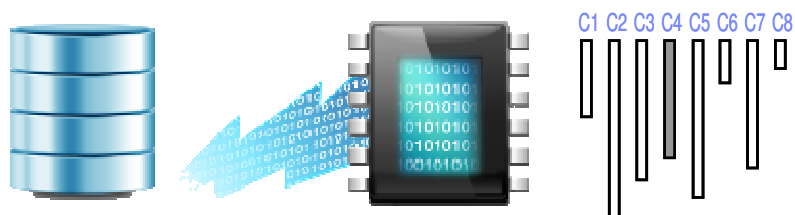
What Makes BLU Acceleration Different?

Unmatched innovations from IBM Research & Development laboratories

Now on Linux
on System z

Next Generation In-Memory

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



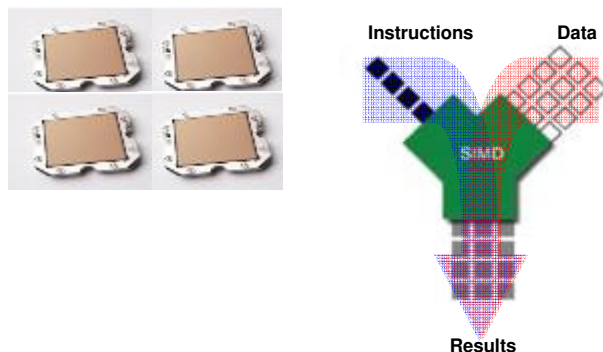
Analyze Compressed Data

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



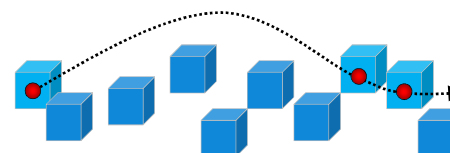
CPU Acceleration

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



Data Skipping

Skips unnecessary processing of irrelevant data

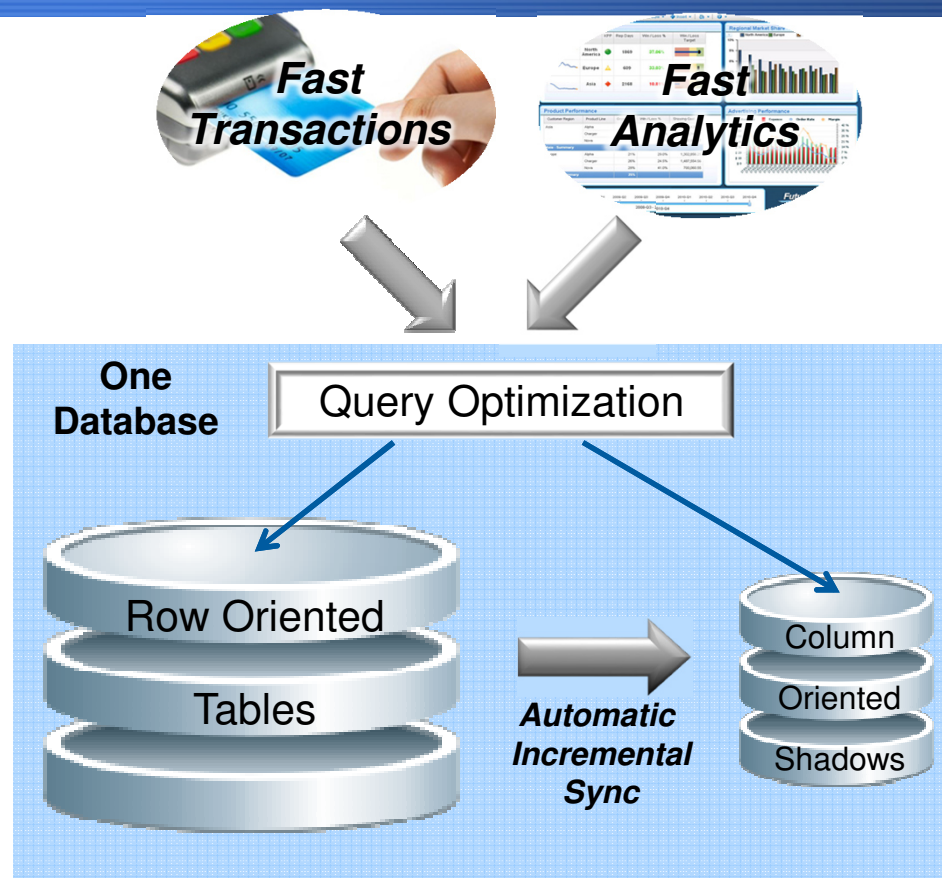


BLU Shadow Tables New

Now on Linux
on System z

Fast Answers. Simply Delivered.

- Instant insight into operational data without compromising transaction performance
- DB2 creates column-based 'Shadow Table' versions of row-based operational data
- Analytic queries **are seamlessly routed to Shadow Tables** to take advantage of BLU Acceleration analytics performance **in the transaction processing environment**
- With BLU Shadow Tables, the performance of analytical queries improves by 10x or more, with equal or greater transactional performance



Reporting and Transactions in the same continuously available system

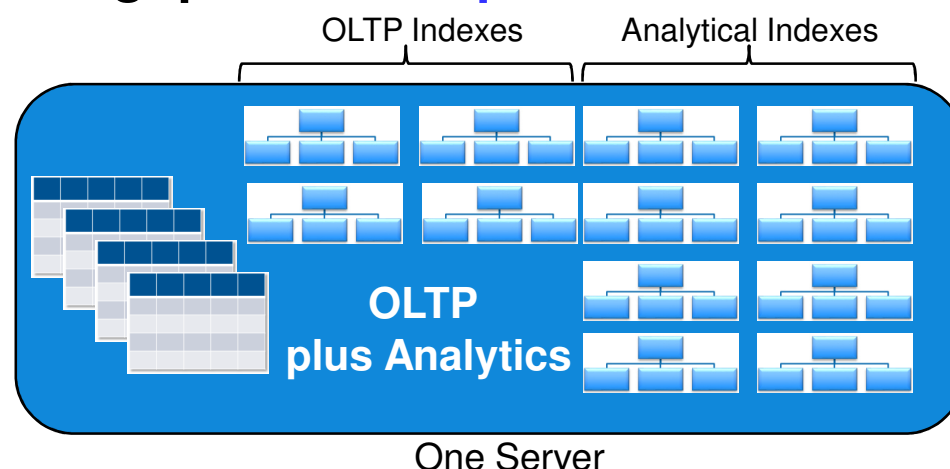
Shadow Tables: Key Facts

- OLTP transactions access the row-organized tables
- Analytic queries access the copies of the same tables, but column-organized. These are called Shadow Tables
- Both types reference the same table names. **No change is required to transactions or queries. Application transparent.**
- **Analytics are much faster** when accessing BLU, column-organized tables than row-organized tables
- **Shadow Tables**, based on BLU technology, **do not require any index other than primary key index**
- **Shadow Tables simplify administration and boost performance for both the OLTP workloads and the analytic workloads.**

Shadow Tables

Great Performance – Simplified Administration

- **OLTP system with OLTP indexes and several extra indexes to support reporting queries – Replace Indexes with Shadow Tables**

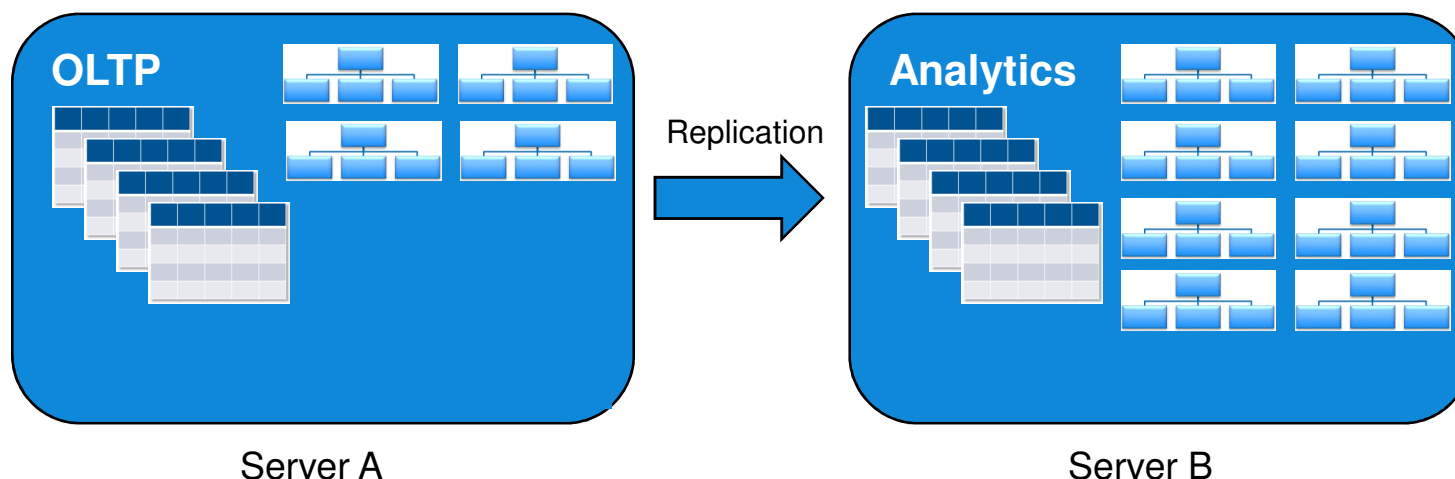


- **Shadow Tables simplify administration and boost performance**
 - Greatly reduces indexes needed for performance – no analytical indexes
 - Greater than 10x faster reporting
 - No impact to OLTP throughput by replacing just 4 analytic indexes with Shadow Tables
 - In fact, when dropping 20 indexes OLTP performance improves by more than 2x

Shadow Tables

Great Performance – Simplified Administration

- **OLTP system with replication to second reporting system**
– **Replace reporting system with Shadow Tables**

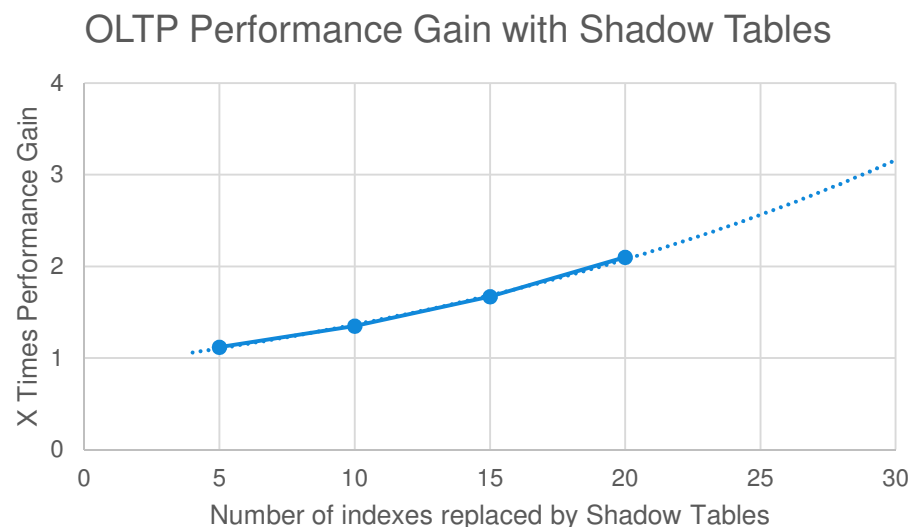


- **Shadow Tables simplify administration and boost performance**
 - Greatly reduces indexes needed for performance – no analytical indexes
 - Greater than 10x faster reporting
 - No impact to OLTP throughput by replacing just 4 analytic indexes with Shadow Tables
 - In fact, when dropping 20 indexes OLTP performance improves by more than 2x

Shadow Tables

Performance Compared to a Single System

- **No negative impact to OLTP performance with BLU Shadow Tables**
 - Equal OLTP performance by replacing just 4 indexes with shadow
 - Replacing more OLTP indexes with a Shadow Table will improve the performance of OLTP
 - By replacing 20 analytic indexes with a Shadow Table, performance of OLTP improves by 2x



- **Reporting queries >10x faster with BLU vs. running reports on row organized OLTP tables**

DB2 with BLU Acceleration 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



*“When we first started using BLU Acceleration in DB2 10.5 we were very pleased with the **10x compression** and the **10x performance improvement**. We are now using the **DB2 Cancun Release**, specifically **Shadow Tables**, to allow our clients to **run reports directly on top of transactional tables**. **The results are delighting my end users** and we **don't see any impact to our transactional performance.**”*

– Paul Peters, Lead Database Administrator, VSN Systemen BV

Oracle Compatibility – Full Use of Compatibility Vector

- **BLU will support the following Oracle capabilities**

- Datatypes

- 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

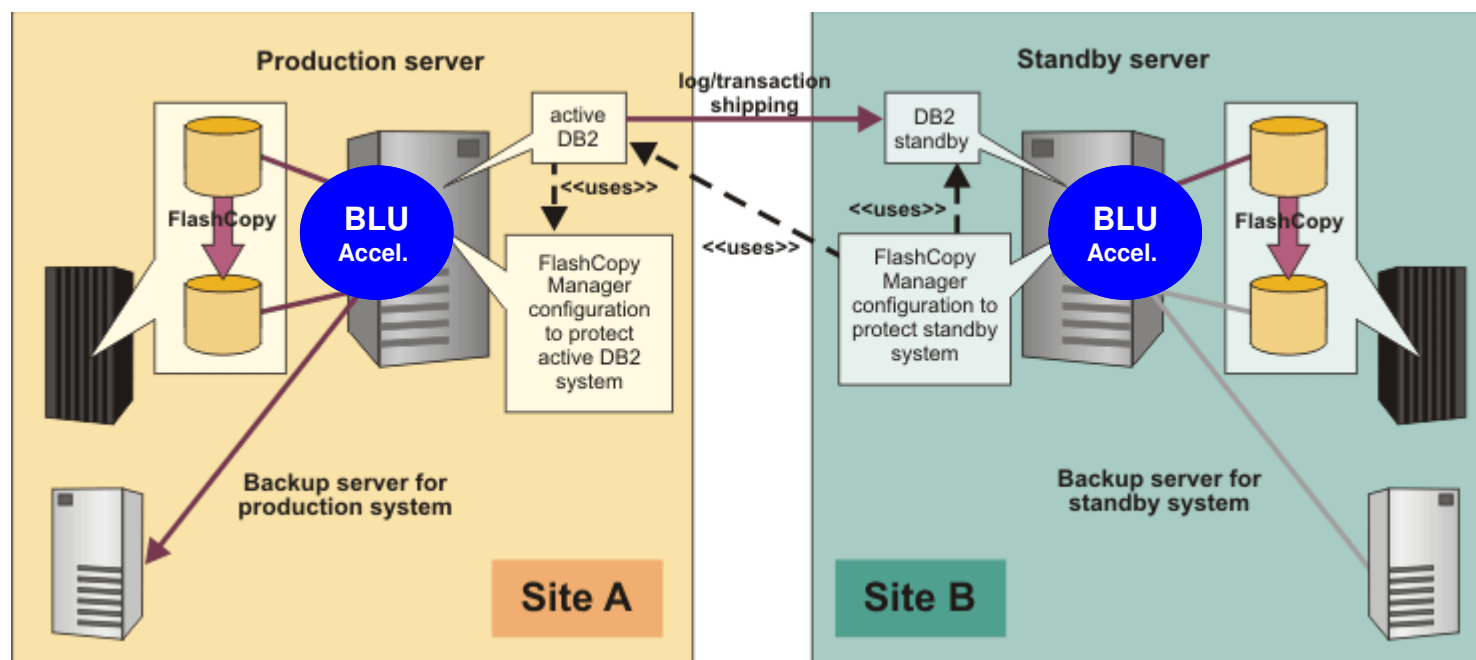
- Supported on BLU tables but not pushed down

- ROWNUM
- Hierarchical Queries(CONNECT BY)
- LIMIT OFFSET clause

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 sync modes, multiple standby, time delay, log spooling



BLU Acceleration – More Performance

- **Performance gains for all analytic workloads**
- **Significant query performance enhancements**
 - complex, nested joins and other joins that are currently run in row engine
 - common table expressions
 - joins involving VARCHAR data
 - data skipping for VARCHAR and CHAR predicates
 - CHAR and VARCHAR now available in synopsis table
- **Significant I/U/D and ELT performance improvements**
 - Further support for primary key index exploitation for point queries and UPDATE/DELETE
 - Optimizations for batch UPDATE/DELETE statements
 - Further improvements for batch INSERT and array INSERT processing
- **New Power 8 exploitations**
- **DB2 BLU with Cognos workload : up to 60% faster**
- **DB2 BLU with SAP BW workload : up to 40% faster**
- **Industry Standard Benchmark : 50% - 350% faster**

Improve BLU Performance of Complex ELT in “Cancun”

▪ Warehousing with More Complex ELT and Operational Analytics

- More general primary key or unique constraint index exploitation for
 - point queries
 - point Update and Delete statements

- Significant performance improvements for Insert and Update
 - Using index access that returns 1 row is significantly faster
 - Update with IN-list predicate is significantly faster
 - In FP1 update with in-list was 14x slower than row store
 - In FP4 update with in-list is up to 3x faster than row store
 - **Fast Insert = 1.8x faster than FP3**
 - **Fast update = 70x faster than FP3**
 - UofT reported back 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**

- Oracle compatibility allows for PL/SQL jobs for ELT jobs

- Support for MERGE statement

Extreme Performance via Deep Power8 Exploitation

- **Cognitive compilation**

- When compiling and optimizing DB2 runtime code, IBM uses special cognitive algorithms that watch DB2 processing BLU Acceleration workloads
- This learning is then used to reorder instructions within the product for even faster runtime performance

- **Faster range predicates for BLU tables**

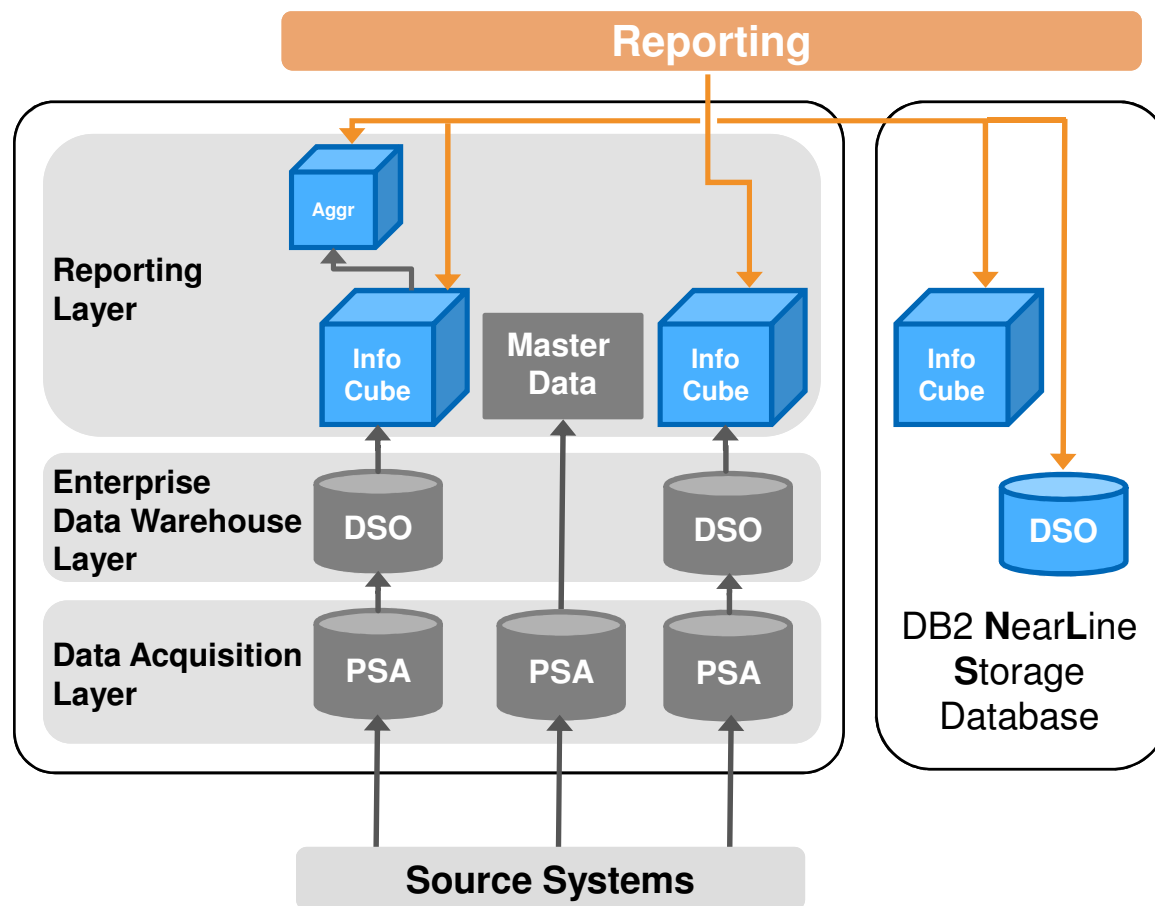
- Power8 has new instructions that can be exploited by SIMD aware applications
- DB2 will leverage these new instructions for range predicates to evaluate many more column values simultaneously compared to Power7 or Intel
- Resulting in even greater performance and faster analytics

BLU Acceleration – More Compression and Simplicity

- **Automatic data sub-setting for LOAD's ANALYZE phase**
 - Sampling and data limits, with intelligent defaults
 - Table sizes shown to be within 1% of the size of table with full analyze on load
- **Adaptive Compression for Inserts**
 - Inserts/INGEST support for creation of page-level dictionaries
- **New compression for VARCHAR data**
- **Add Column support for columnar tables**
- **Federation support for databases with columnar tables**

First DB2 BLU roll-out Phase for SAP BW

The **highlighted** BW objects are supported with **BLU** in initial roll-out phase (\leq DB2 10.5 **FP3**)



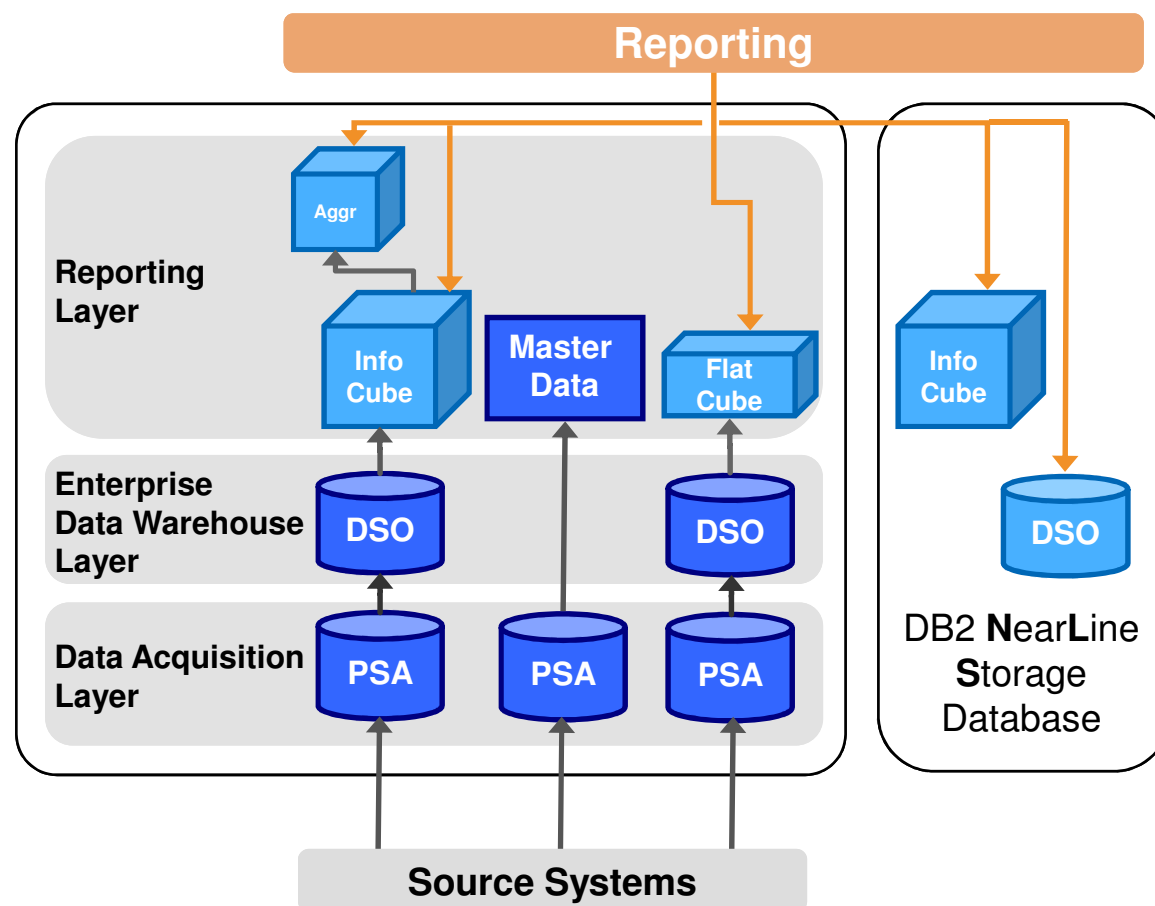
- Support for SAP BW 7.0 or higher
- Supported BW objects:
 - InfoCubes and Aggregates
 - BW temporary tables
 - NLS InfoCubes und NLS DSOs

Advantages:

- ✓ **Faster BW Reports**
- ✓ **No index tuning required**
- ✓ **Significant storage savings**
- ✓ **Less Aggregates required**

Enhancements with DB2 10.5 Cancun (FP4)

BLU support for all SAP BW objects



- Supported BW objects:
 - InfoCubes / Aggregates
 - Flat InfoCubes (with SAP BW 7.40)
 - DSO and PSA
 - InfoObjects
 - BW temporary tables

Advantages:

- ✓ **Faster BW Reports:**
 - Queries only access BLU tables (columnar)
 - Flat InfoCubes require less table joins
- ✓ **BLU support for DSO objects can help to reduce InfoCubes**
- ✓ **Potentially less ETL processing**

Sizing for BLU Acceleration



Determine Total Size of Uncompressed Table Data

- For a new database, use the filesystem reported size of delimited ASCII input files
- For an existing, uncompressed database, calculate the total size of table data only
- For an existing, compressed database, estimate the uncompressed table data size
 - Ensure all table statistics are up-to-date, and then
 - Use a SQL query on the system catalog tables to estimate the total size. Note that this query depends on table statistics being up to date

```
select sum(a.fpages * (1.0/(1.0 - (cast(a.pctpagessaved as
decimal(5,2))/100))) * c.pagesize/1024/1024/1024) as
uncompressed_table_data_GB from syscat.tables a,
syscat.datapartitions b, syscat.tablespace c where a.tabschema
not like 'SYS%' and a.tabschema = b.tabschema and a.tabname =
b.tabname and b.datapartitionid = 0 and b.tbpaceid = c.tbpaceid
```

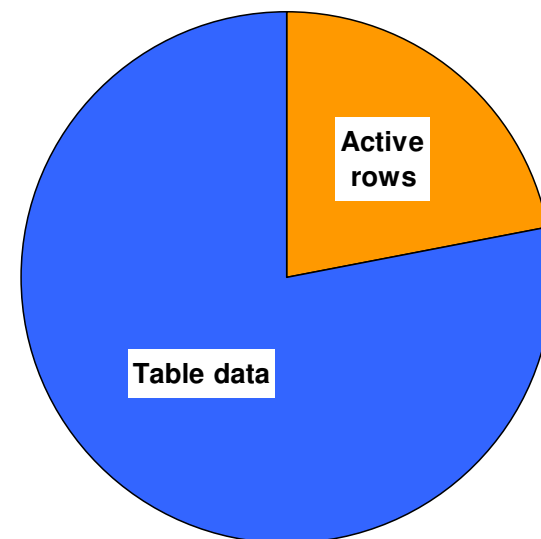
Determine the Size of Active Uncompressed Table Data

▪ Step (a) – Limit to active rows

- Review how hot/warm/cold are the data rows
- Focus on what percentage of the range of data in the database a query typically touches
- Example
 - Database stores 7 years of data
 - If typical query accesses
 - > All 7 years → 100% of rows are active
 - > 3/7 years → 43% of rows are active
 - > 1/7 years → 14% of rows are active

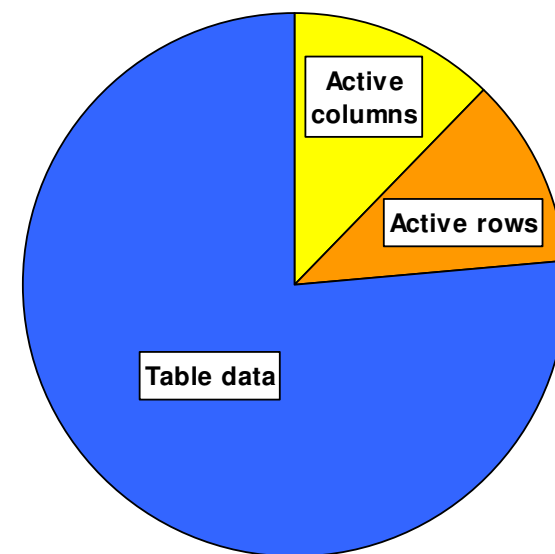
▪ Default assumption

- Common rule of thumb estimate in warehouses is that 25-30% of all rows are active at any given time



Determine the Size of Active Uncompressed Table Data (cont.)

- **Step (b) – Limit to active columns**
 - Review how hot/warm/cold are the columns
 - Focus on what percentage of the columns in the tables a query typically touches
 - EXPLAIN plans will show column access per query
 - Use `EXPLAIN_OBJECT.COLUMN_COUNT` and `EXPLAIN_STREAM.COLUMN_COUNT` to determine ratio per query
- **Default assumption**
 - Common rule of thumb estimate in warehouses is that less than 50% of all columns are active at any given time



BLU Acceleration Compression

- **To determine amount of memory required for BLU Acceleration, need to estimate the BLU compression rate**
- **Conservative rule of thumb: 7-8x compression of table data**
- **Over-estimating the compression rate will result in an under-sized system**

How Much Memory Does BLU Need?

- **Memory is allocated within DB2 in three areas**
 - Bufferpool
 - Sort memory (for joins and sorts)
 - Other memory (for locking, utility heap, etc)
- **To take advantage of in-memory optimizations in BLU Acceleration, it is recommended to target 40-50% of the size of the compressed **active** table data to fit in the bufferpool**
- **Calculate the amount of memory required for the desired bufferpool configuration**

How Much Memory Does BLU Need? (cont.)

- **BLU Acceleration best practice recommendations have two different distribution recommendations (bufferpool/sort/other)**
 - Low concurrency: 40/40/20
 - High concurrency: 25/55/20

- 1. **For a given bufferpool size, determine the total amount of memory required for the desired distribution (see above)**

- 2. **Calculate the required cores to support the memory distribution using an 8GB/core ratio (**this is the minimum recommended**)**

- 3. **Round the number of cores up to the nearest appropriate socket value and recalculate memory requirements at 8GB/core**

- 4. **Increase the memory/core ratio as required to address concurrency/complexity requirements**

Putting it All Together

Total Uncompressed Data **10TB**

Use 7.5x compression
(be conservative here) **1.3TB**

Reduce by Active Row % **400GB** Assume 30%

Reduce by Active Column % **200GB** Assume 50%

Fit Half the active data in memory **100GB**

Size total memory assuming above is 40% of total

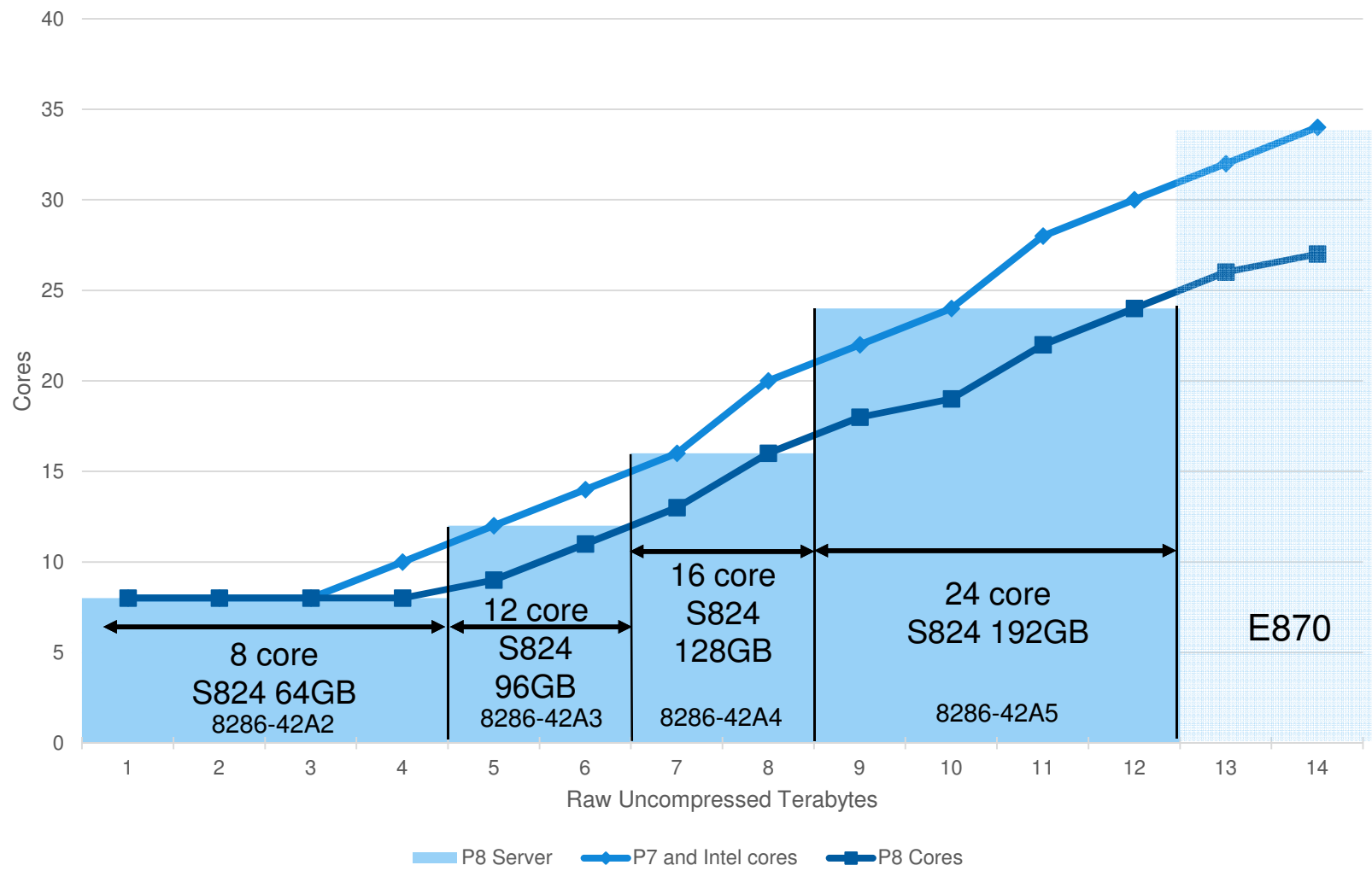
100GB **100GB** **50GB** Assume 40% BP = 250GB

Divide by 8GB/core to get core count

250 / 8 = 32 cores

Power 8 Minimum Sizing Made Simple

BLU Power 8 Sizing

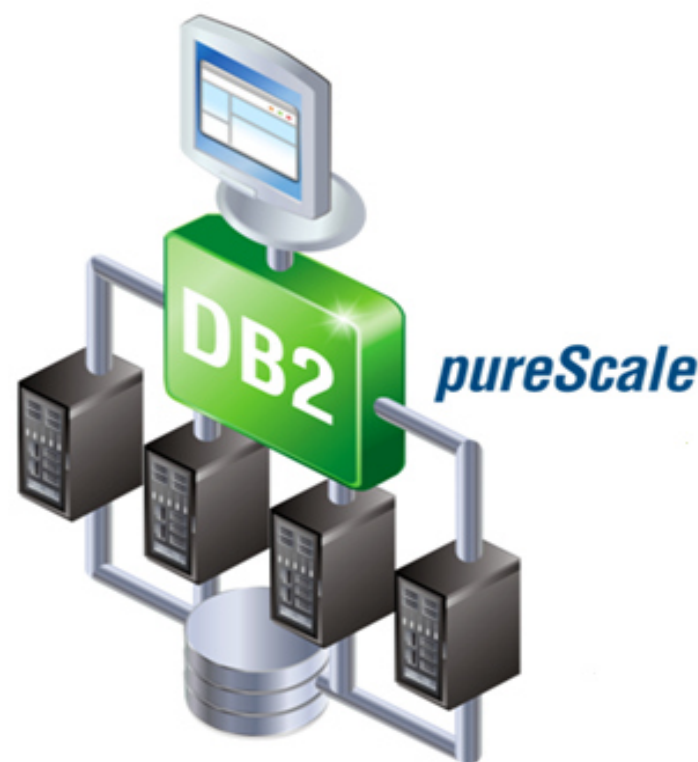


DB2 Brings New Economics to Continuous Availability

More Flexible and More Affordable

New

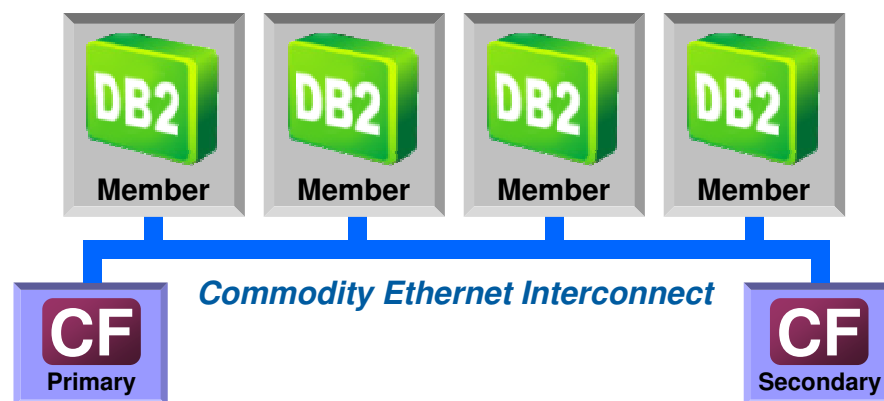
- **Leverage commodity hardware and network adapters**
- **Continuous availability**
 - Deliver uninterrupted data access with consistent performance
- **Extreme scalability**
 - Add capacity as your needs grow, without over provisioning
- **Application transparency**
 - Avoid the time, risk and cost of application changes



DB2 pureScale with TCP/IP Interconnect

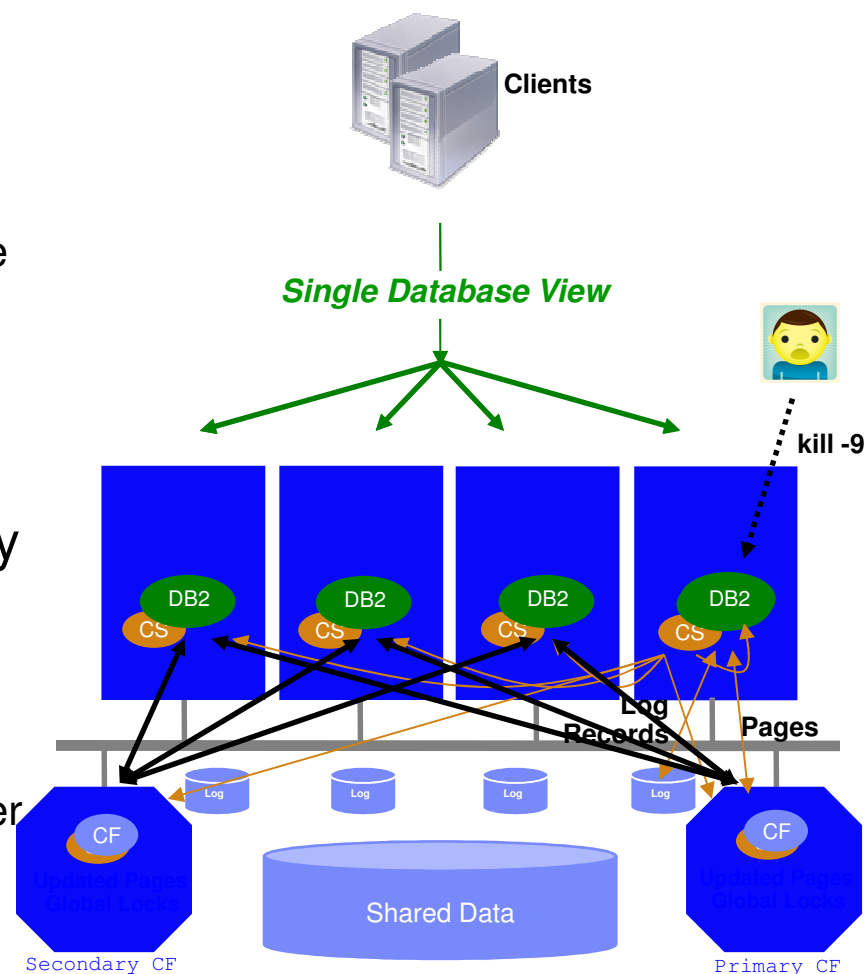


- TCP/IP (sockets) interconnect for **faster cluster setup and lower cost deployments** using commodity network hardware
- Provides **exactly the same level of high availability** as RDMA-based pureScale environments
- Appropriate for small clusters with moderate workloads where availability is the primary motivator for pureScale
- 10 Gigabit Ethernet (10GE) strongly recommended for production installations



Member Failure Summary

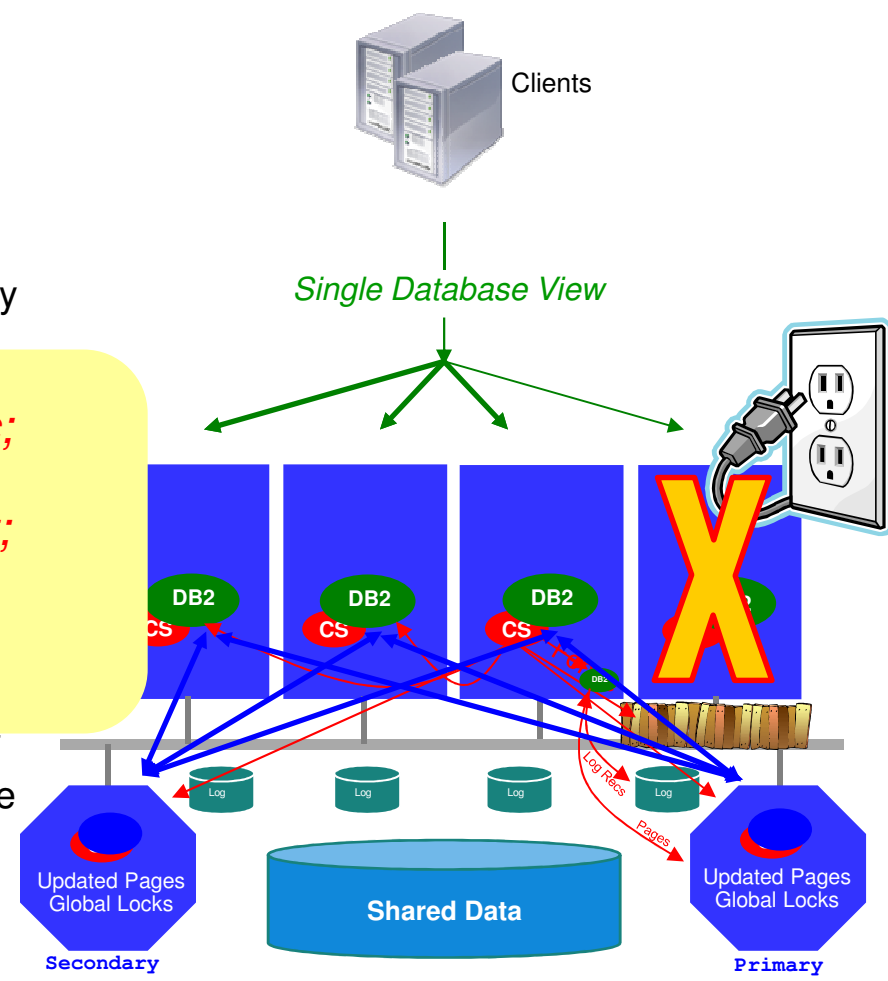
- Member Failure
- DB2 Cluster Services automatically detects member's death
 - Inform other members, and CFs
 - Initiates automated member restart on same or remote host
 - Member restart is like crash recovery in a single system, but is much faster
 - Redo limited to in-flight transactions
 - Benefits from page cache in CF
- Client transparently re-routed to healthy members
- Other members fully available at all times – *“Online Failover”*
 - CF holds update locks held by failed member
 - Other members can continue to read and update data not locked for update by failed member
- Member restart completes
 - Locks released and all data fully available



Member HW Failure : “Restart on Guest Host”

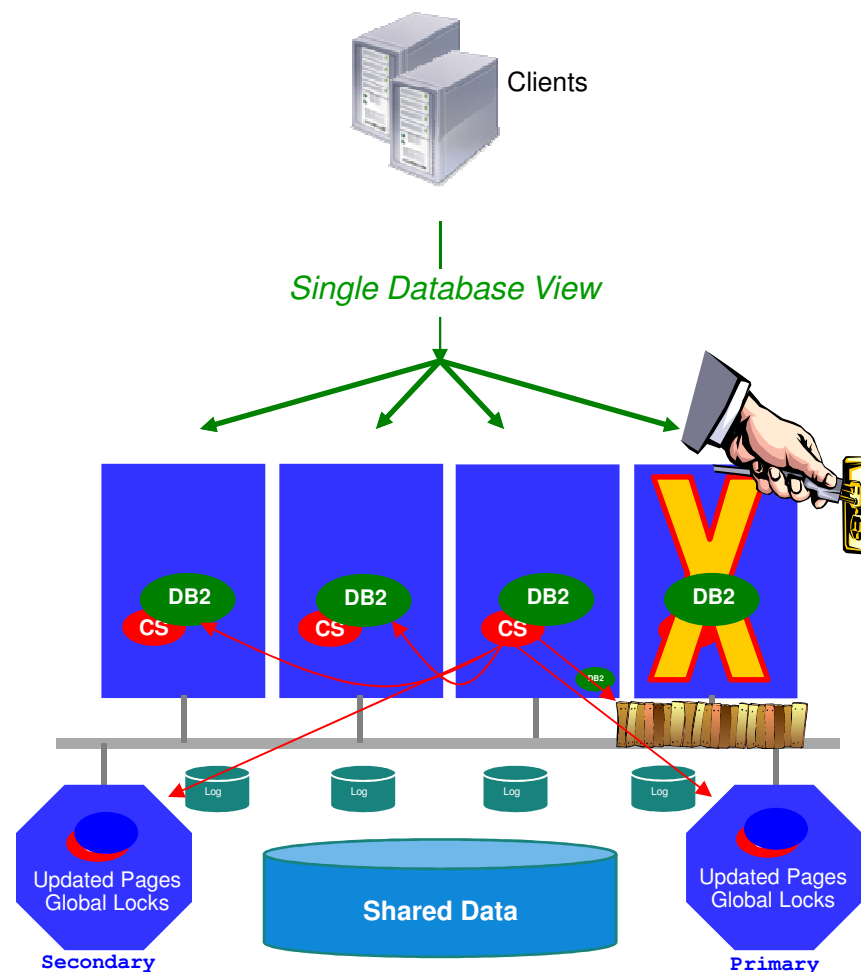
- **Power cord tripped over accidentally**
- **DB2 Cluster Services loses heartbeat and declares member down**
 - Informs other members & CFs
 - Fences member from logs and data
 - Initiates automated member restart on another (“guest”) host
 - Using reduced, and pre-allocated memory model
 - Member restart is like a database crash recovery in a single system database, but is much faster
 - Redo limited to inflight transactions
 - Benefits from page cache in Primary
- **In the mean-time, client connections automatically re-routed to other members**
 - Based on least load (by default)
 - Pre-designated failover mechanism
- **Other members remain fully available throughout – “Online Failover”**
 - Primary retains update locks held by members at the time of failure
 - Other members can continue to read and update data not locked for write access by failed member
- **Member restart completes**
 - Retained locks released and all data available

Automatic;
Ultra Fast;
Online



Member Failback

- **Power restored and system re-booted**
- **DB2 Cluster Services automatically detects system availability**
 - Informs other members and CFs
 - Removes fence
 - Brings up member on home host
- **Client connections automatically re-routed back to member**



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

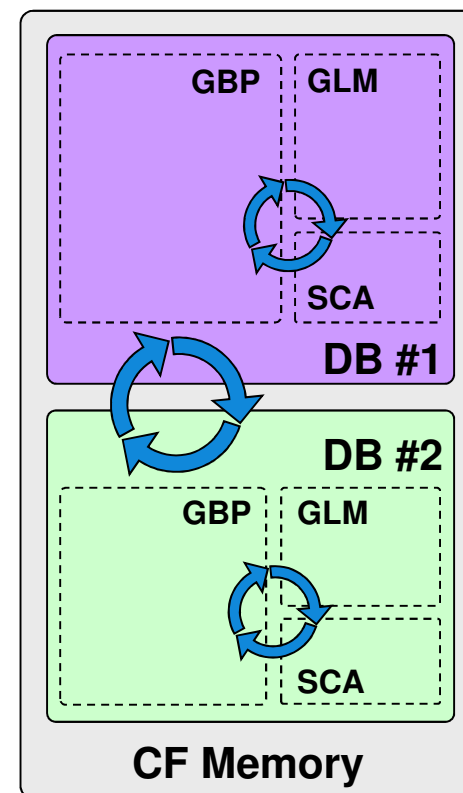
- **KVM supported with**

- Any x64 system that is supported by both RHEL 6.2 and DB2 pureScale
- RHEL 6.2 and higher

CF Self-Tuning Memory

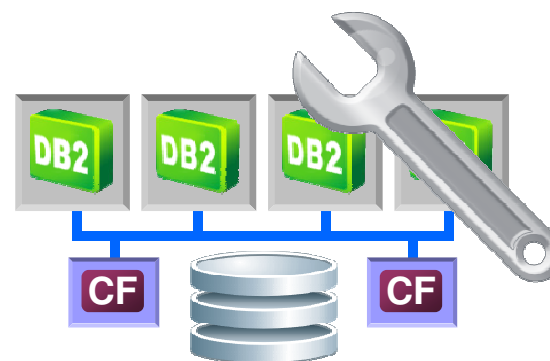
A green circular icon with the word "New" in white text.

- **CF memory is optimally distributed between consumers based on workload**
 - Less administrative overhead for DBA, with reduction in memory monitoring and management
- **Can function at two levels**
 - Dynamic distribution of CF memory between multiple databases in an instance
 - Dynamic distribution of database's CF memory between its consumers
 - Group buffer pool (GBP)
 - Global lock manager (GLM)
 - Shared communication area (SCA)
- **Enabled by setting:**
 - Registry variable `DB2_DATABASE_CF_MEMORY=AUTO`

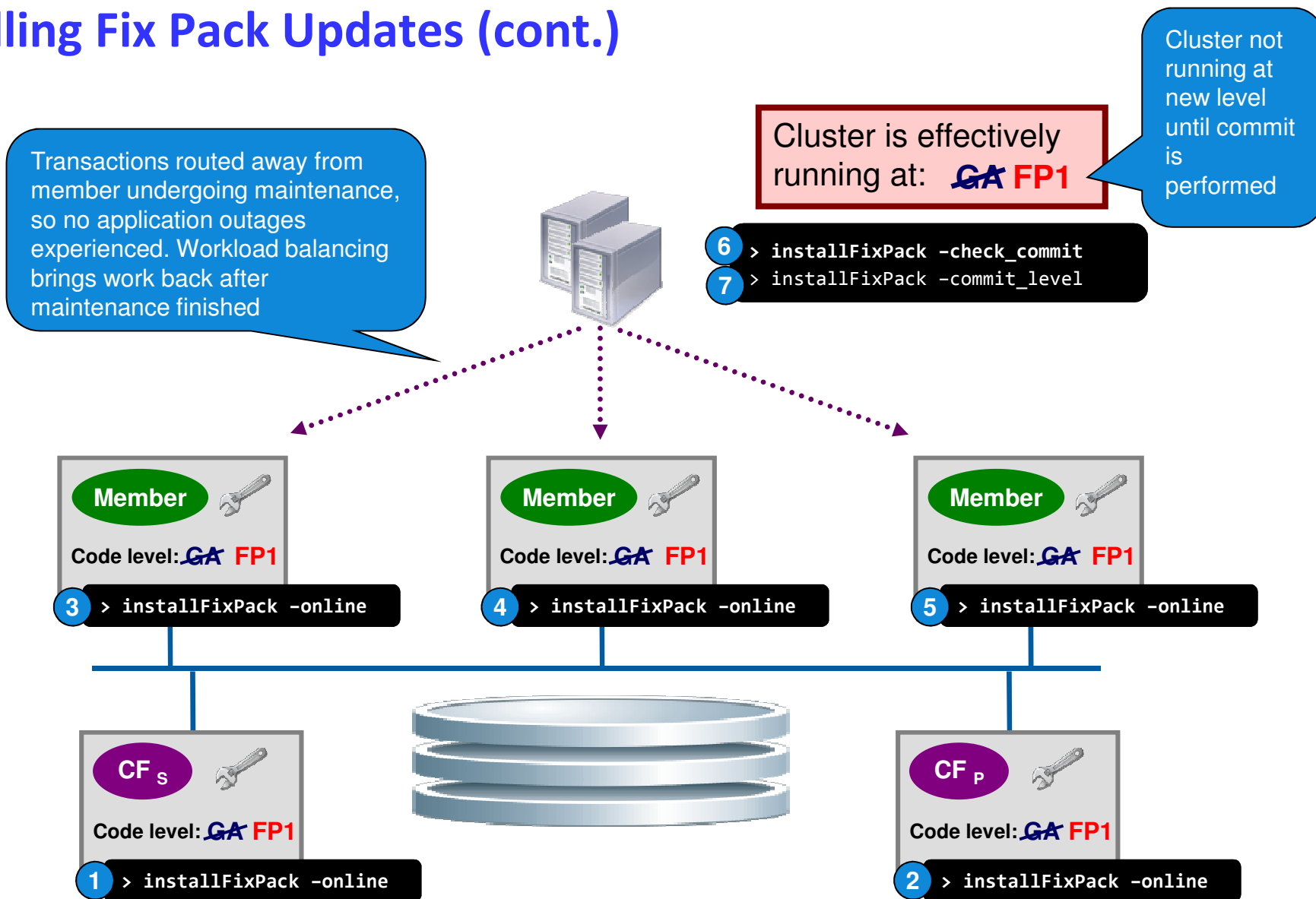


Rolling Database Fix Pack Updates

- **Transparently install pureScale fix packs in an online rolling fashion**
- **No outage experienced by applications**
- **Single `installFixPack` command run on each member/CF**
 - Quiesces member
 - Existing transactions allowed to finish (configurable timeout, default is 2 minutes)
 - New transactions sent to other members
 - Installs binaries
 - Updates instance
 - Member still behaves as if running on previous fix pack level
 - Unquiesces member
- **Final `installFixPack` command to complete and commit updates**
 - Instance now running at new fix pack level



Rolling Fix Pack Updates (cont.)



Extreme Performance via Deep Power8 Exploitation

- **Faster performance for financial calculations**
 - Decimal arithmetic using new vector based instructions
 - Row based tables benefit from vector processing on decimal data

- **Improved integrity and reliability**
 - Leverage new Power8 algorithms for high speed memory integrity checking
 - Increased processing performance while ensuring a higher level of integrity for data pages

- **Optimizations for increased concurrency**
 - Power8 will support twice the threading of Power7
 - Can result in software contention if not optimized for
 - DB2 10.5 FP4 exploits low level P8 AIX latching algorithms to improve the concurrency of these extremely highly threaded servers

What does the next generation database look like?

- ✓ **The most advanced in-memory technology on the market today**
- ✓ **Super fast for transactional and analytic workloads**
- ✓ **Available, reliable, resilient**
- ✓ **Simple, intelligent and agile**
- ✓ **Easy to deploy, cloud ready**



The Next Generation Data Platform

Transact

- Provide high levels of service without the high price
- Seamlessly expand or contract as needed, paying only for what you use when you use it
- Enhance customer experience by delivering data when and where it's needed



Analyze

- Leverage next generation in-memory technology
- Get instant insight into operational and warehouse data without compromising performance of either
- Deploy analytics solutions faster with “load and go” and operational simplicity

Innovate

- Free your applications from database complexity for faster delivery
- Transform your ability to make business decisions with 35x to 73x faster analytics, with some queries running more than 1400x faster^{1,2}
- Optimize IT resources and utilization with built-in simplicity and autonomies