DataBase Management Systems (DBMS)
Technical Overview and Industry Trends

Eberhard Hechler
Executive Architect, Member IBM Academy of Technology
Information Management Development
IBM Germany R&D Lab

ehechler@de.ibm.com
+49-170-4568453

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Windows Server and the Windows logo are trademarks of the Microsoft group of countries.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

© 2014 IBM Corporation

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

* Other product and service names might be trademarks of IBM or other companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon 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 throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

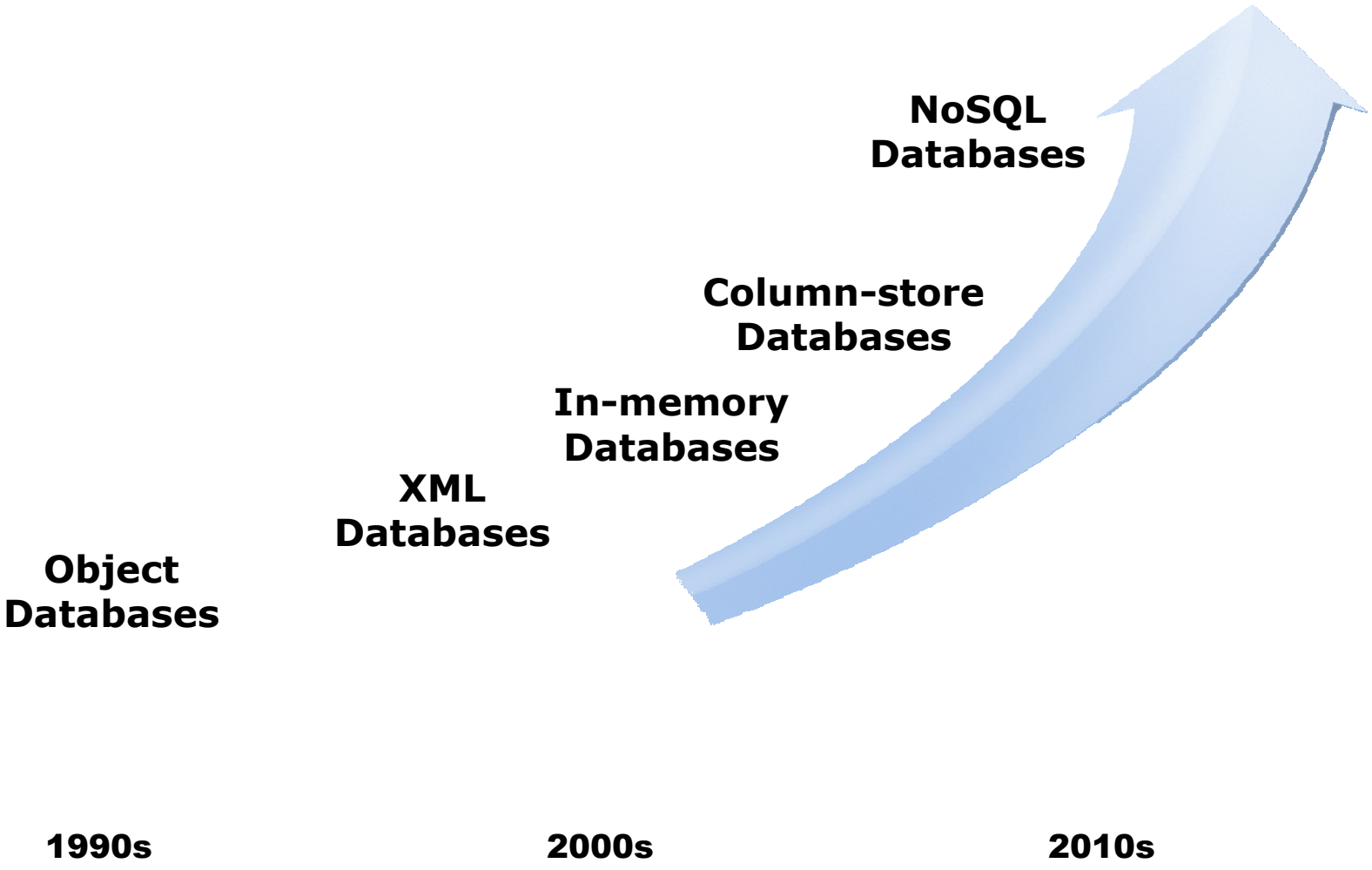
This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g, zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the “Authorized Use Table for IBM Machines” provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html (“AUT”). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Agenda

- **Industry Trends**
  - HTAP
  - Reduction of DW Repositories
  - Simplification of Information Supply Chain
- **Overview of DB2 with BLU Acceleration**
  - Column-oriented DBMS
  - Data Compression
  - Performance Improvements
- **IBM DB2 Analytics Accelerator (IDAA)**
  - DB Appliances
  - DB2 for z/OS Integration

DB2 with BLU Acceleration Technology

Technology Evolution



DB2 10.5 – Processing and analyzing more Data faster

DB2® 10.5
with BLU Acceleration

Multi-workload database software for the era of big data

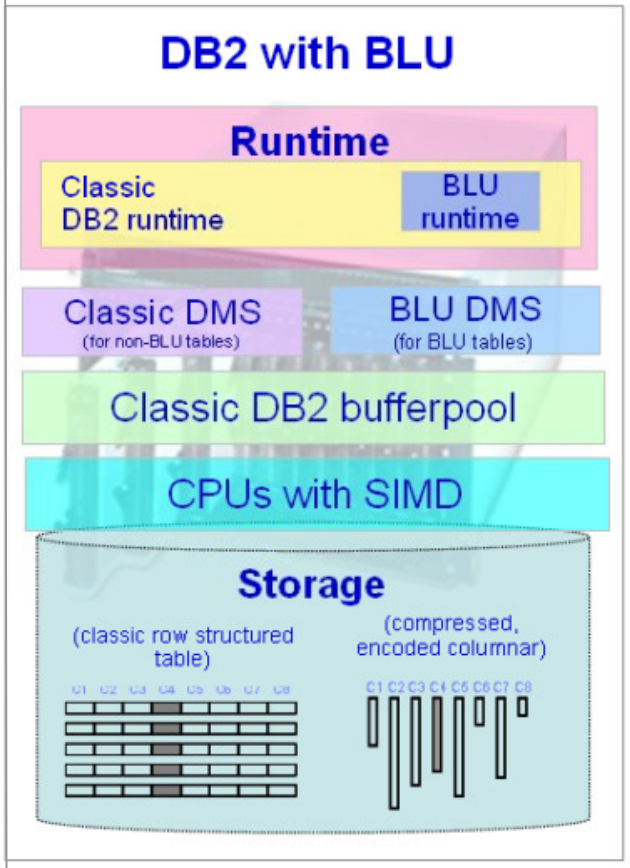
- ***Always Available Transactions***
  99.999% availability with disaster recovery up to 1000s km; means minimal downtime
- ***Faster Analytics***
  In-memory hybrid yields 25x faster analytics without costs or limits of in-memory only
- ***Unprecedented Affordability***
  ~ 1/3 less than Oracle with > 99% Oracle Database application compatibility
- ***Future-Proofed Infrastructure***
  NoSQL and IBM Mobile Database allows clients to expand and modernize their apps

What is DB2 with BLU Acceleration?

- **Large order of magnitude benefits**
  - Performance
  - Storage savings
  - Time to value

- **New technology in DB2 for analytic queries**
  - CPU-optimized unique runtime handling
  - Unique encoding for speed and compression
  - Unique memory management
  - Columnar storage, vector processing
  - Built directly into the DB2 kernel

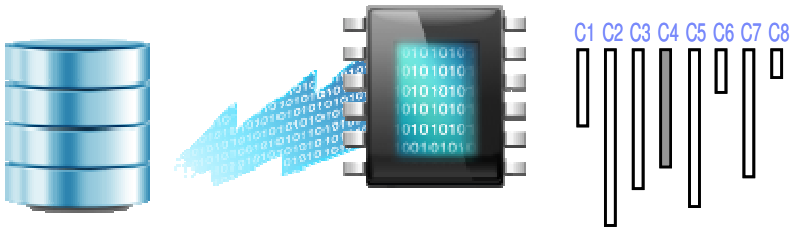
- **Revolution or evolution**
  - BLU tables coexists with traditional row tables - in same schema, storage, and memory
  - Query any combination of row or BLU tables
  - Easy conversion of tables to BLU tables
    - Change everything, or change incrementally



Why is BLU Acceleration a different Technology

Dynamic In-Memory

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

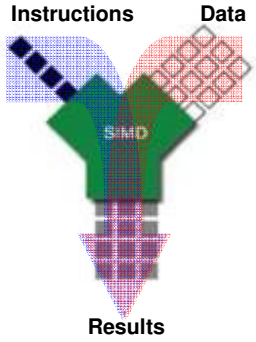


Actionable Compression

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

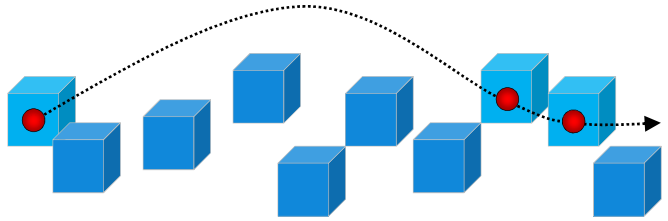
Parallel Vector Processing

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



Data Skipping

Skips unnecessary processing of irrelevant data



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

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

How fast is DB2 BLU Acceleration

Customer	Performance Gains
BNSF	Up to 137x
Handelsbanken	Avg 25x – Up to 100x
Triton Consulting	46x
Yonyou	40x
Coca-Cola Bottling	4x - 15x

10x-25x
speedup
is common

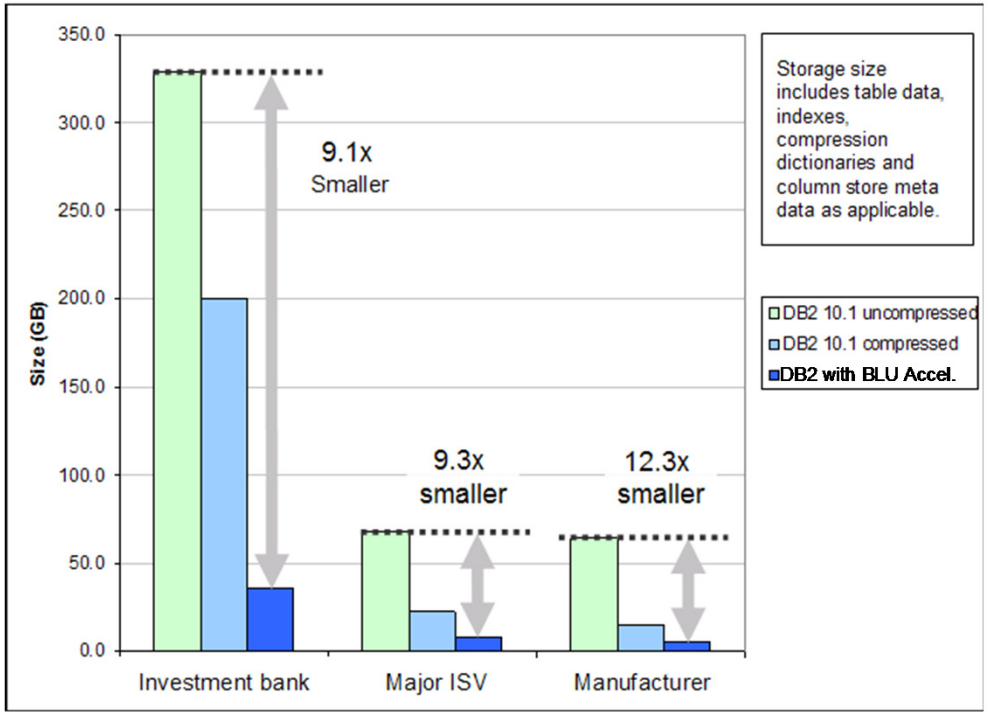
BNSF RAILWAY

"It was amazing to see the faster query times compared to the performance results with our row-organized tables. ***The performance of four of our queries improved by over 100-fold! The best outcome was a query that finished 137x faster by using BLU Acceleration."***

- Kent Collins, Database Solutions Architect, BNSF Railway

Storage Savings

- **Multiple examples of data requiring substantially less storage**
  - 95% smaller than uncompressed data size
  - Fewer objects required – no storage required for indexes, aggregates, etc
- **Multiple compression techniques**
  - Processing takes place on compressed data
- **Compression algorithm adapts to the data**



Seamless Integration into DB2

- **Built seamlessly into DB2 – integration and coexistence**
  - Column-organized tables can coexist with existing, traditional, 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 or combinations of column-organized 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 it's optimized query processing and fewer database objects required
  - Simple migration
    - Conversion from traditional row table to BLU Acceleration is easy
    - DB2 Workload Manager identifies workloads to tune
    - Optim Query Workload Tuner recommends BLU Acceleration table transformations
    - Users only notice speed up; DBA's only notice less work!
  - Management of single server solutions less expensive than clustered solutions

Simplification of Analytic Operations

Traditional Warehouse
Database Design and Tuning

1. Decide on partition strategies
2. Select Compression Strategy
3. Create Table
4. Load data
5. Create Auxiliary Performance Structures
   - Materialized views
   - Create indexes
     - B+ indexes
     - Bitmap indexes
6. Tune memory
7. Tune I/O
8. Add Optimizer hints
9. Statistics collection

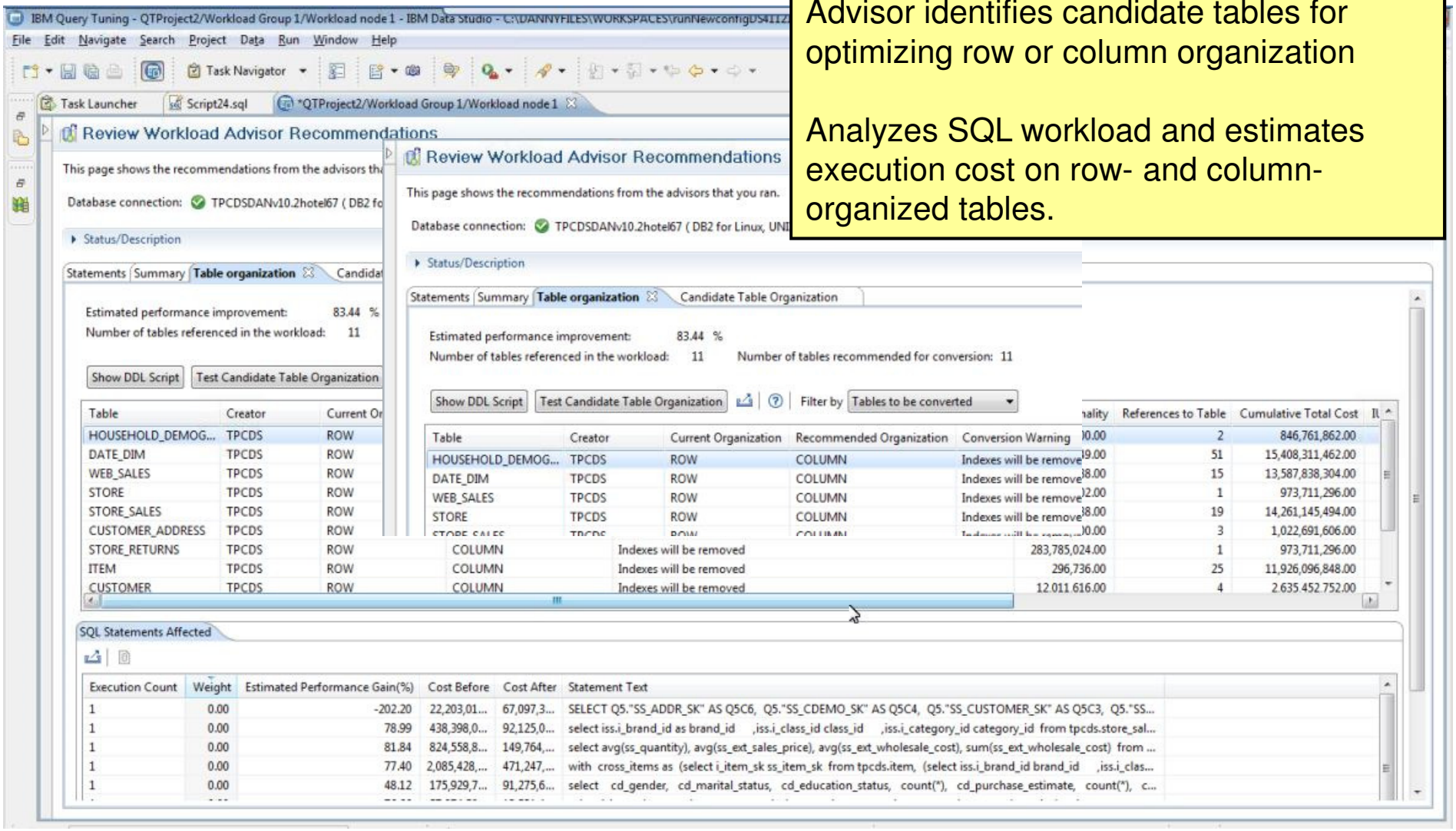
Repeat

AFTER
DB2 with BLU Acceleration

1. Create Table
2. Load data

Create
Load
GO!

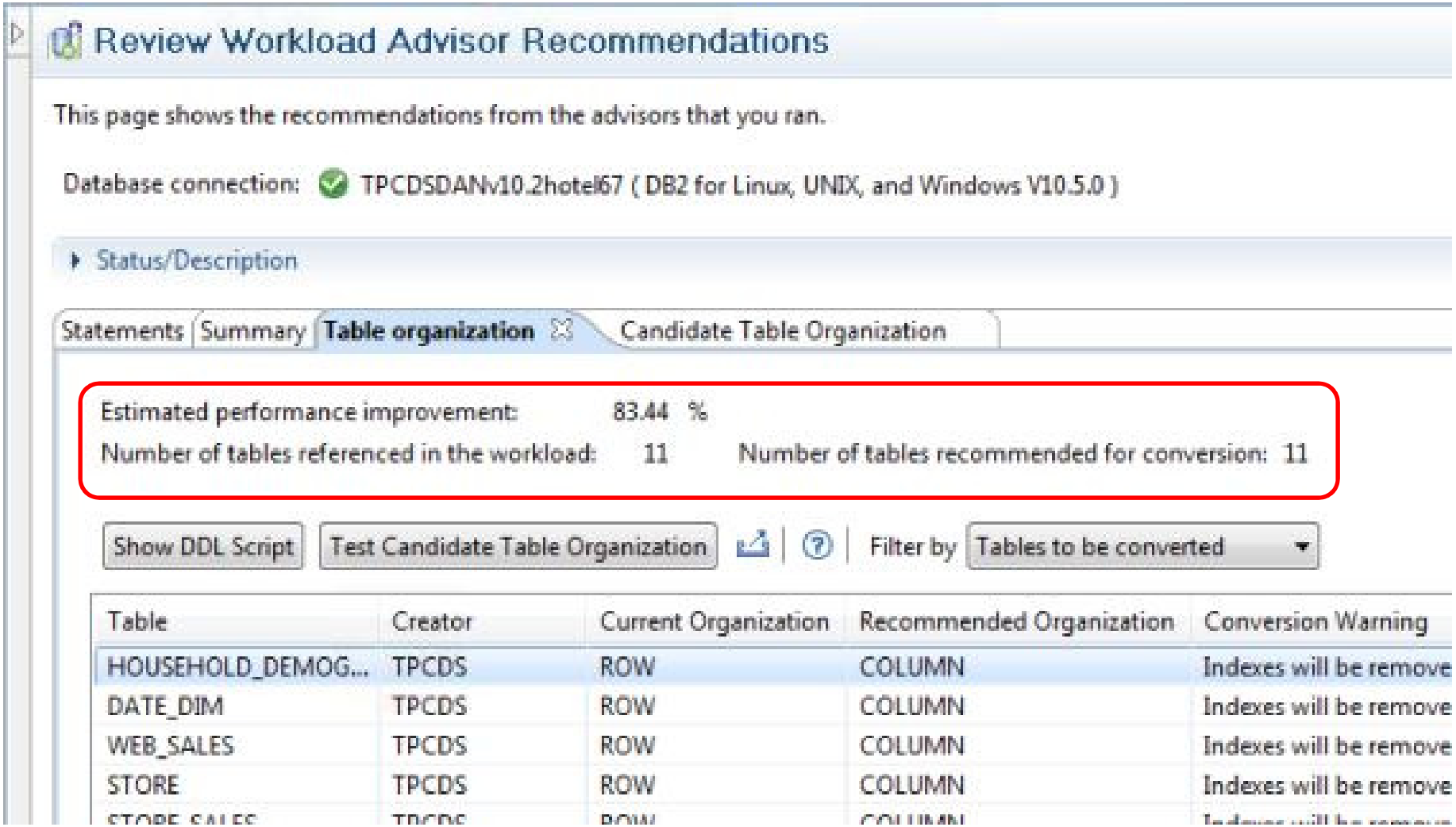
IBM Optim Query Workload Tuner



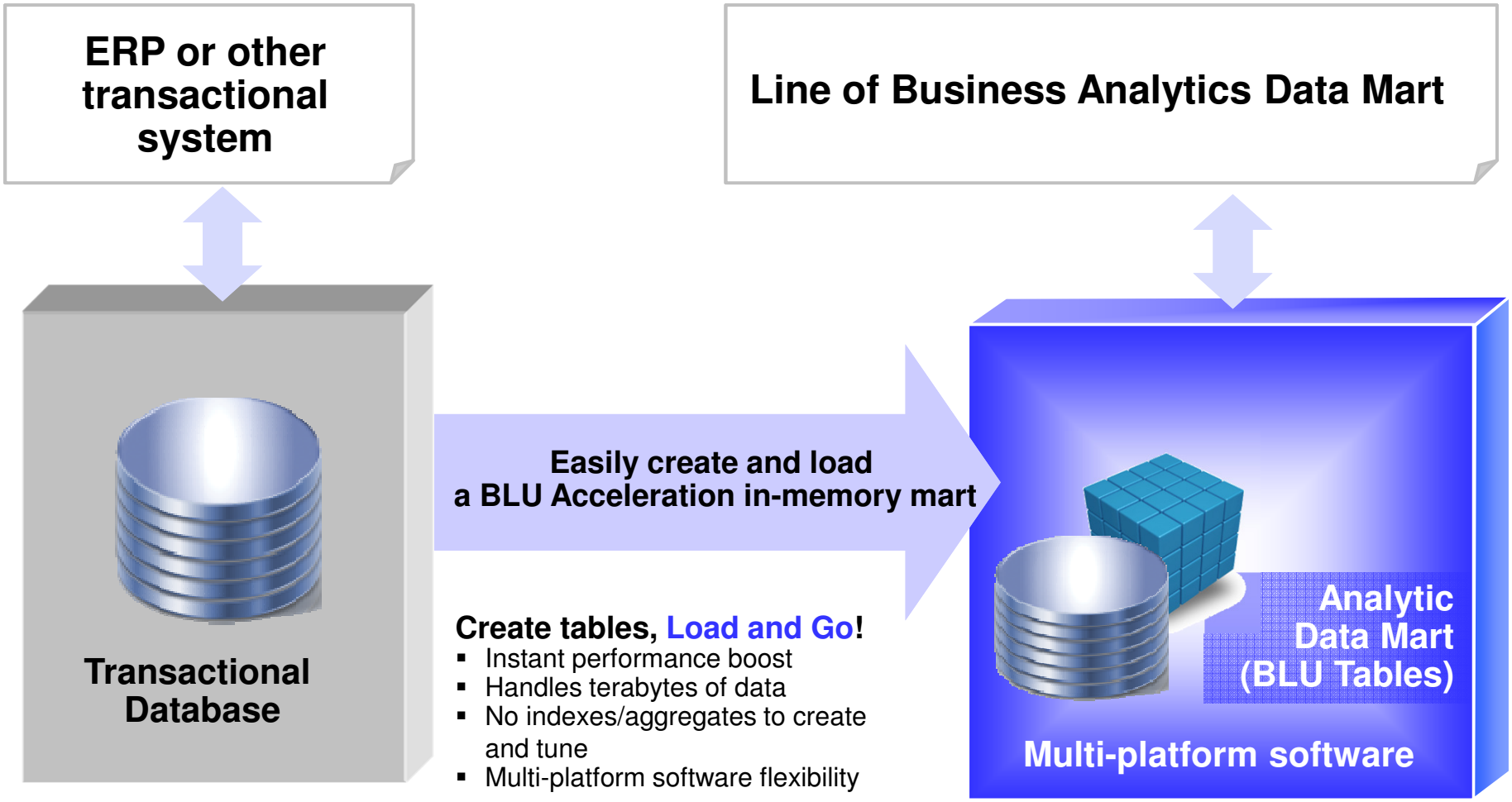
Advisor identifies candidate tables for optimizing row or column organization

Analyzes SQL workload and estimates execution cost on row- and column-organized tables.

IBM Optim Query Workload Tuner



Analytics Data Mart
From Transactional Database



7 Big Ideas: 1 Simple to Implement and Use

- **LOAD and then… run queries**
  - No indexes
  - No `REORG` (it's automated)
  - No `RUNSTATS` (it's automated)
  - No MDC or MQTs or Materialized Views
  - No partitioning
  - No statistical views
  - No optimizer hints

- **It is just DB2!**
  - Same SQL, language interfaces, administration
  - Reuse DB2 process model, storage, utilities

"The BLU Acceleration technology has some obvious benefits: It makes our analytical **queries run 4-15x faster** and **decreases the size of our tables by a factor of 10x**. But it's when I think about **all the things I don't have to do with BLU**, it made me appreciate the technology even more: **no tuning, no partitioning, no indexes, no aggregates**."
-Andrew Juarez, Lead SAP Basis and DBA

7 Big Ideas: 1 Simple to Implement and Use

- **One setting optimized the system for BLU Acceleration**
  - Set `DB2_WORKLOAD=ANALYTICS`
  - Informs DB2 that the database will be used for analytic workloads

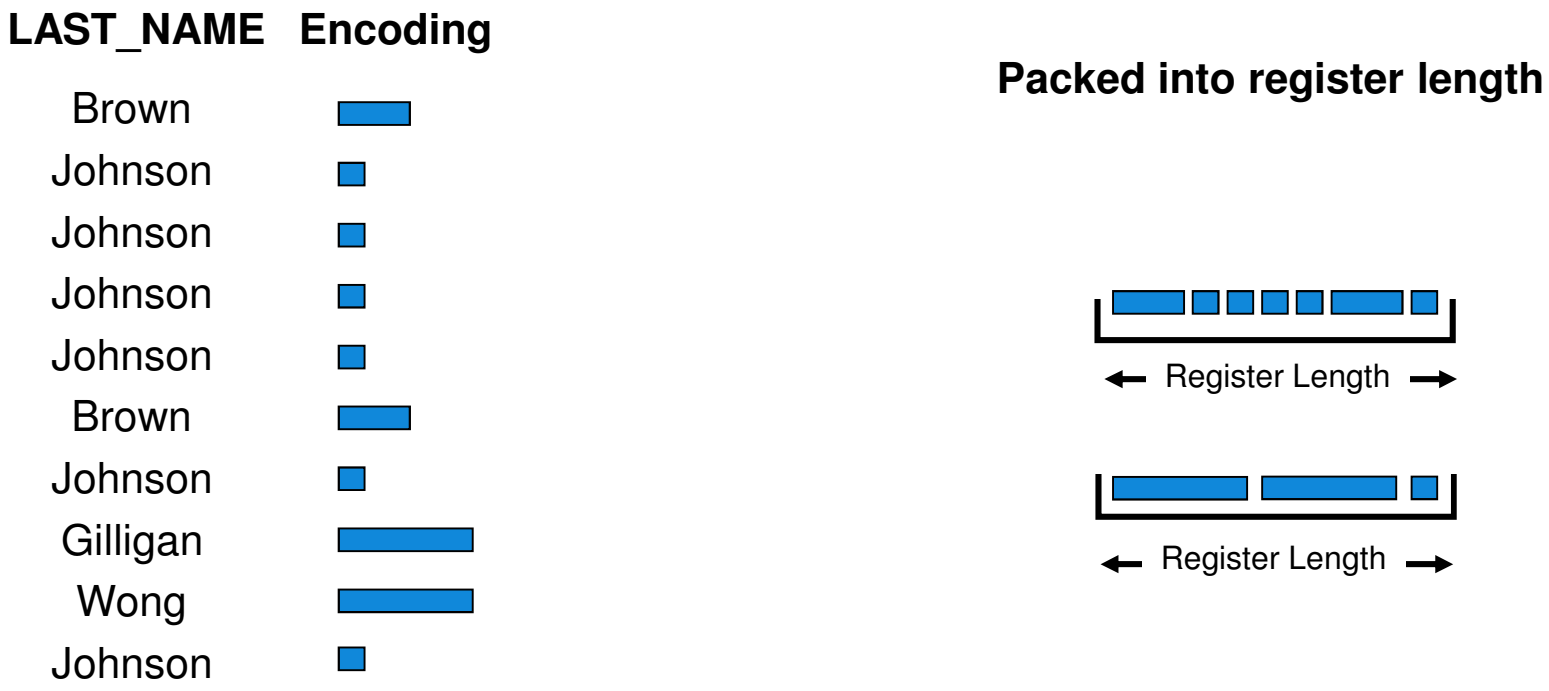
- **Automatically configures DB2 for optimal analytics performance**
  - Makes column-organized tables the default table type
  - Enables automatic workload management
  - Enables automatic space reclaim
  - Page and extent size configured for analytics
  - Memory for caching, sorting and hashing, utilities are automatically initialized based on the server size and available RAM

- **Simple Table Creation**
  - If `DB2_WORKLOAD=ANALYTICS`, tables will be created column organized automatically
  - For mixed table types can define tables as `ORGANIZE BY COLUMN` or `ROW`
  - Compression is always on – No options

- **Easily convert tables from row-organized to column-organized**
  - `db2convert` utility

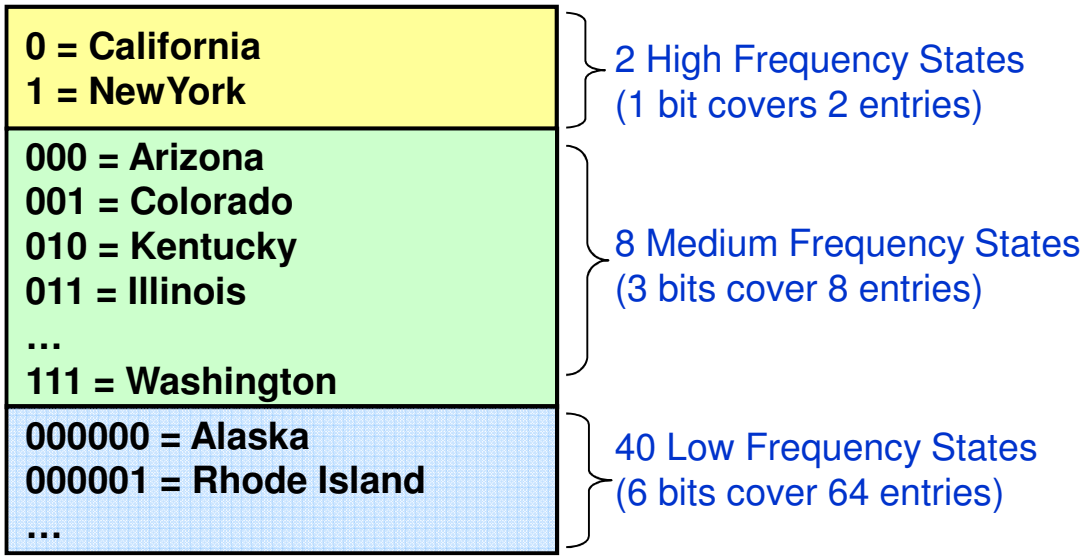
7 Big Ideas: 2 Compute-friendly Encoding/Compression

- **Massive compression with approximate Huffman encoding**
  - More frequent the value, the fewer bits it takes
- **Register-friendly encoding dramatically improves efficiency**
  - Encoded values packed into bits matching the register width of the CPU
  - Fewer I/Os, better memory utilization, fewer CPU cycles to process



BLU uses multiple Compression Techniques

- **Approximate Huffman-Encoding ("frequency-based compression"), prefix compression, and offset compression**
- **Frequency-based compression: Most common values use fewest bits**

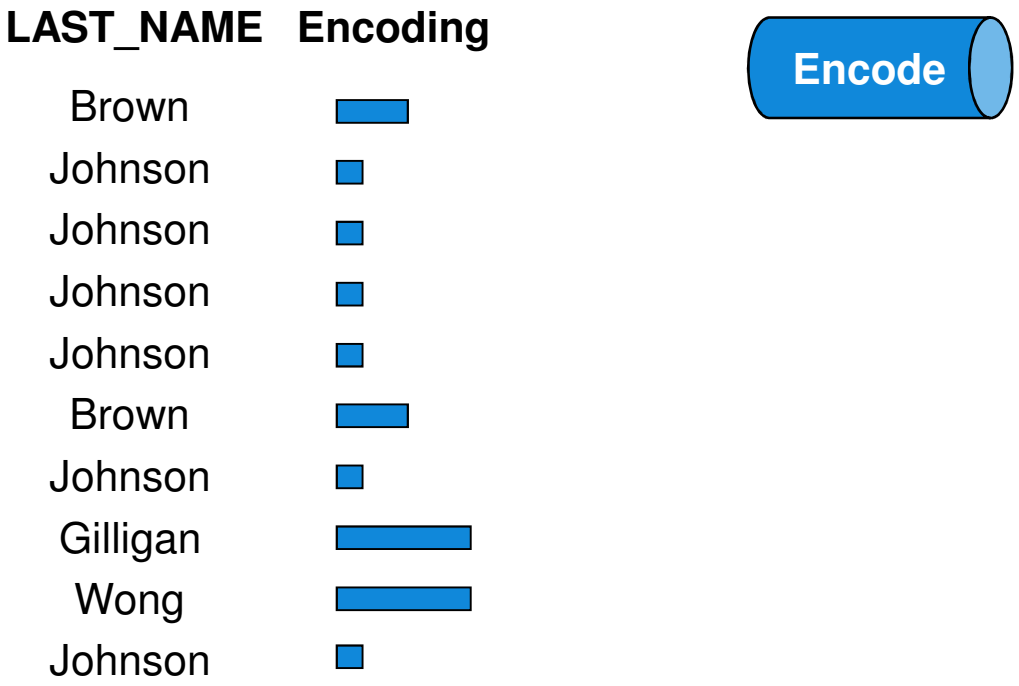


- Exploiting skew in data distribution improves compression ratio
- Very effective since all values in a column have the same data type
- Maps entire values to dictionary codes

7 Big Ideas: 2 Data remains compress. during Evaluation

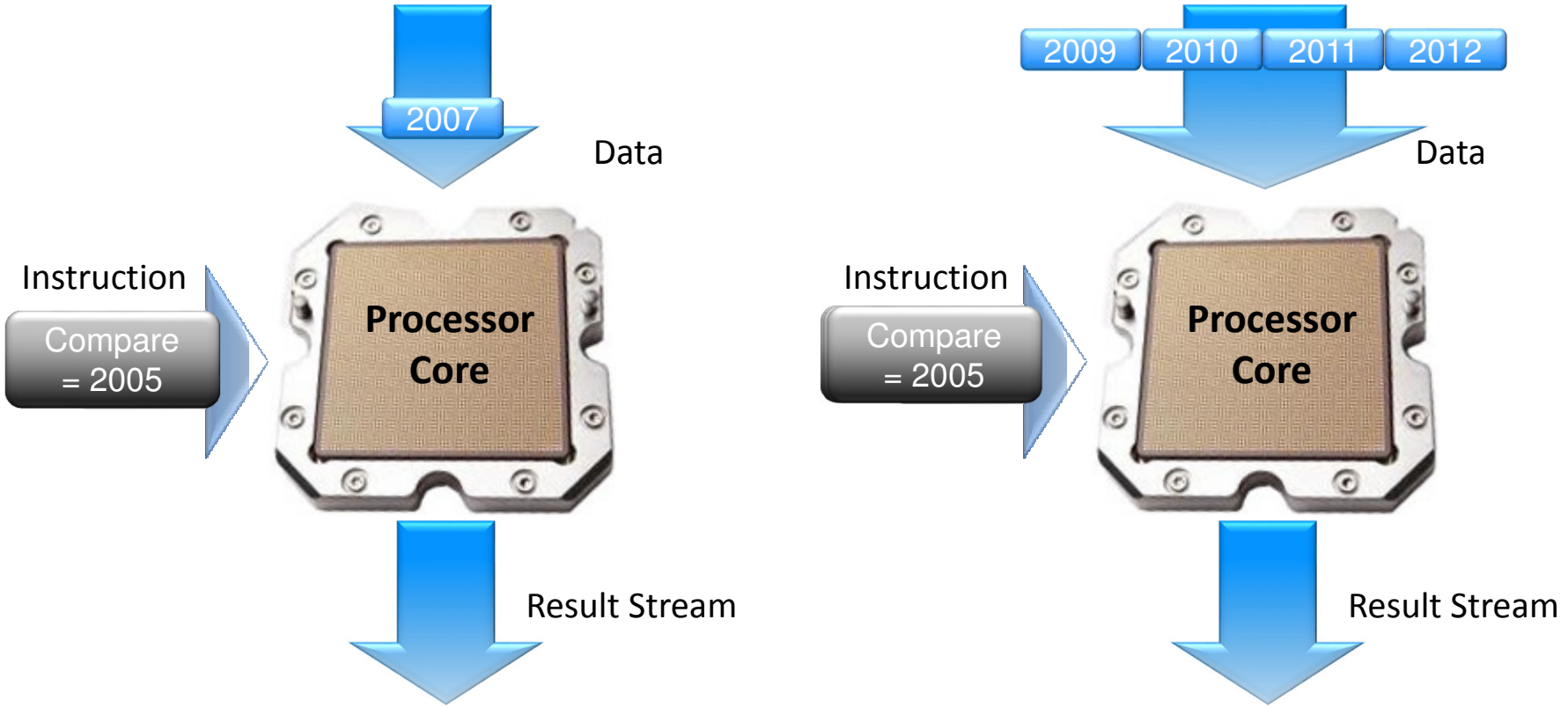
- **Encoded values do not need to be decompressed during evaluation**
  - Predicates (=, <, >, >=, <=, Between, etc), joins, aggregations and more work directly on encoded values

```
SELECT COUNT(*) FROM T1 WHERE LAST_NAME = 'Johnson'
```



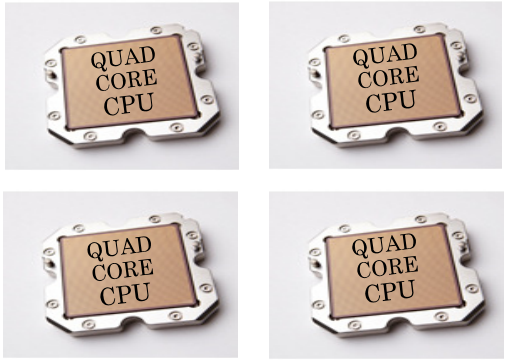
7 Big Ideas: 3 Multiply the Power of the CPU

- **Performance increase with Single Instruction Multiple Data (SIMD)**
- **Using hardware instructions, DB2 with BLU Acceleration can apply a single instruction to many data elements simultaneously**
  - Predicate evaluation, joins, grouping, arithmetic



7 Big Ideas: 4 Core-friendly Parallelism

- **Careful attention to physical attributes of the server**
  - Queries on BLU Acceleration tables automatically parallelized
- **Maximizes CPU cache, cacheline efficiency**



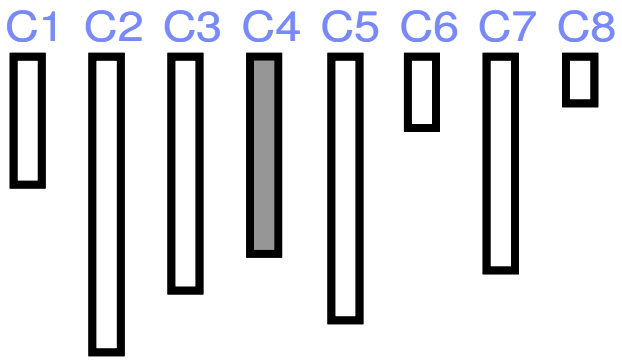
BNSF RAILWAY

"During our testing, we couldn't help but notice that ***DB2 10.5 with BLU Acceleration is excellent at utilizing our hardware resources****. The core-friendly parallelism that IBM talks about was clearly evident and* ***I didn't even have to partition the data across multiple servers****."*

- Kent Collins, Database Solutions Architect, BNSF Railway

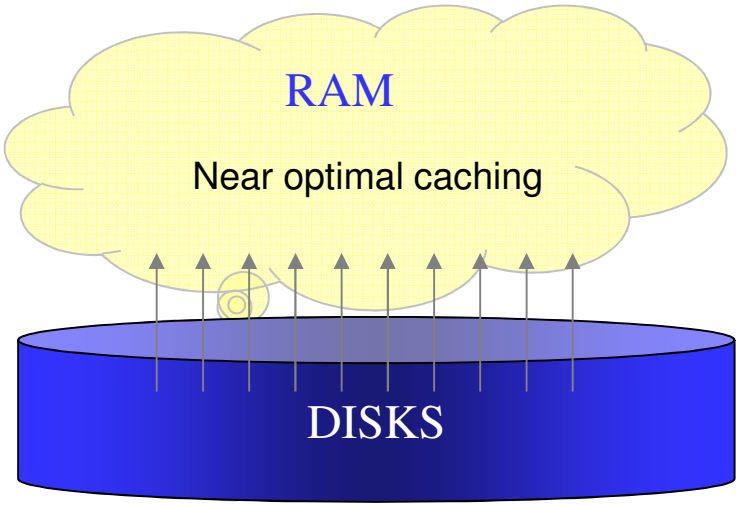
7 Big Ideas: 5 Column Store

- **Minimal I/O**
  - Only perform I/O on the columns and values that match query
  - As queries progresses through a pipeline the working set of pages is reduced
- **Work performed directly on columns**
  - Predicates, joins, scans, etc. all work on individual columns
  - Rows are not materialized until absolutely necessary to build result set
- **Improved memory density**
  - Columnar data kept compressed in memory
- **Extreme compression**
  - Packing more data values into very small amount of memory or disk
- **Cache efficiency**
  - Data packed into cache friendly structures



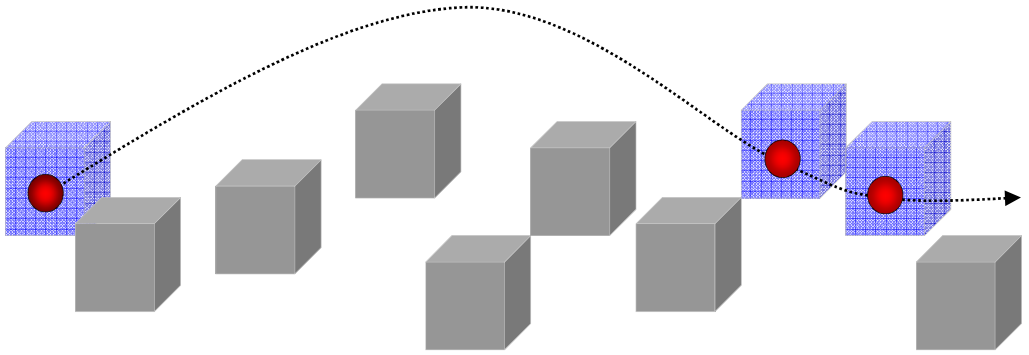
7 Big Ideas: 6 Scan-friendly Memory Caching

- **New algorithms cache in RAM effectively**
- **High percent of interesting data fits in memory**
  - We leave the interesting data in memory with the new algorithms
- **Data can be larger than RAM**
  - No need to ensure all data fits in memory
  - Optimization for in memory and I/O efficiency



7 Big Ideas: 7 Data Skipping

- **Automatic detection of large sections of data that do not qualify for a query and can be ignored**
- **Order of magnitude savings in all of I/O, RAM, and CPU**
- **No DBA action to define or use – truly invisible**
  - Persistent storage of min and max values for sections of data values



Synopsis Table

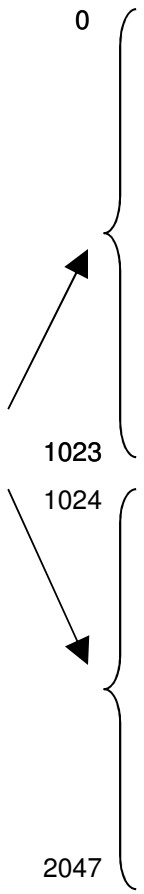
- Meta-data that describes which *ranges* of values exist in which parts of the user table

SYN13033016521627515 2_SALES_COL

TSNMIN	TSNMAX	S_DATEMIN	S_DATEMAX	. . .
0	1023	2005-03-01	2006-10-17	. . .
1024	2047	2006-08-25	2007-09-15	. . .
. . .				

TSN = Tuple Sequence Number

- Enables DB2 to skip portions of a table when scanning data to answer a query
- Benefits from data clustering, loading pre-sorted data



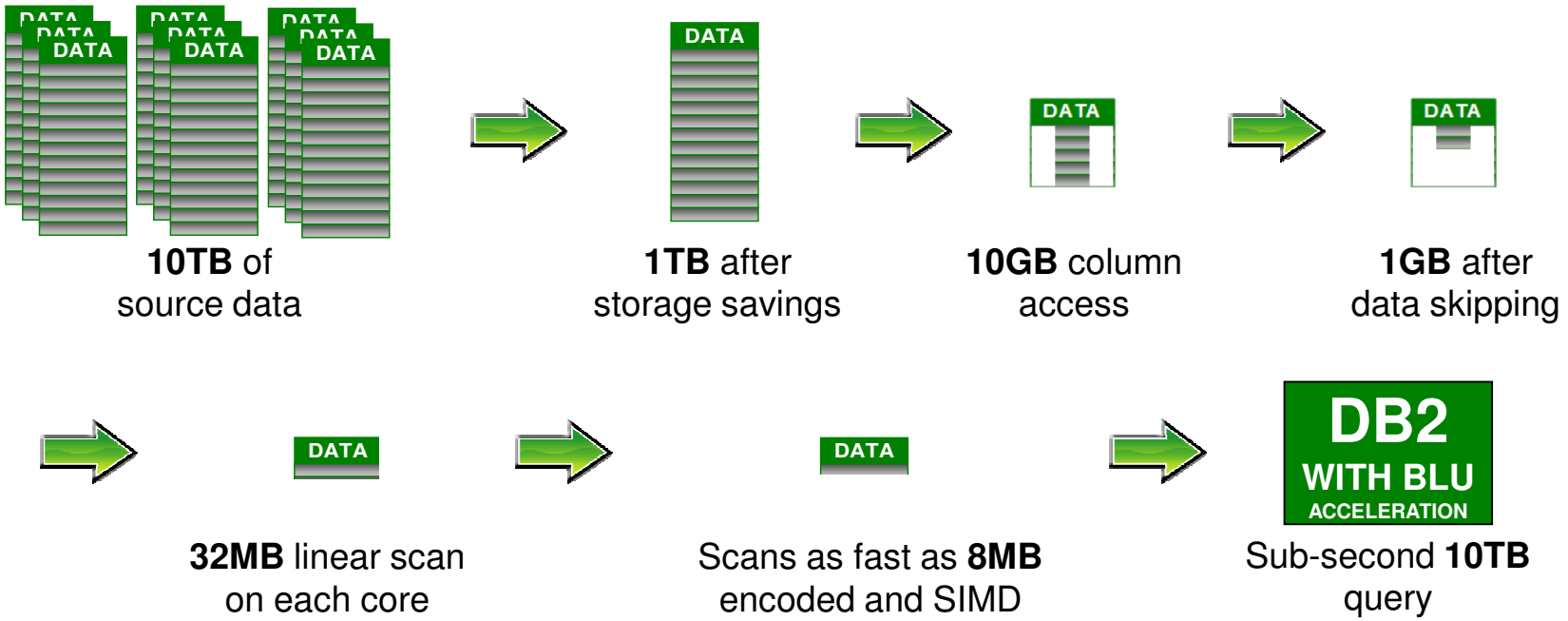
User table: **SALES_COL**

S_DATE	QTY	. . .
2005-03-01	176	. . .
2005-03-02	85	. . .
2005-03-03	267	
2005-03-04	231	
. . .		
. . .		
. . .		

7 Big Ideas: How DB2 with BLU Acceleration Helps

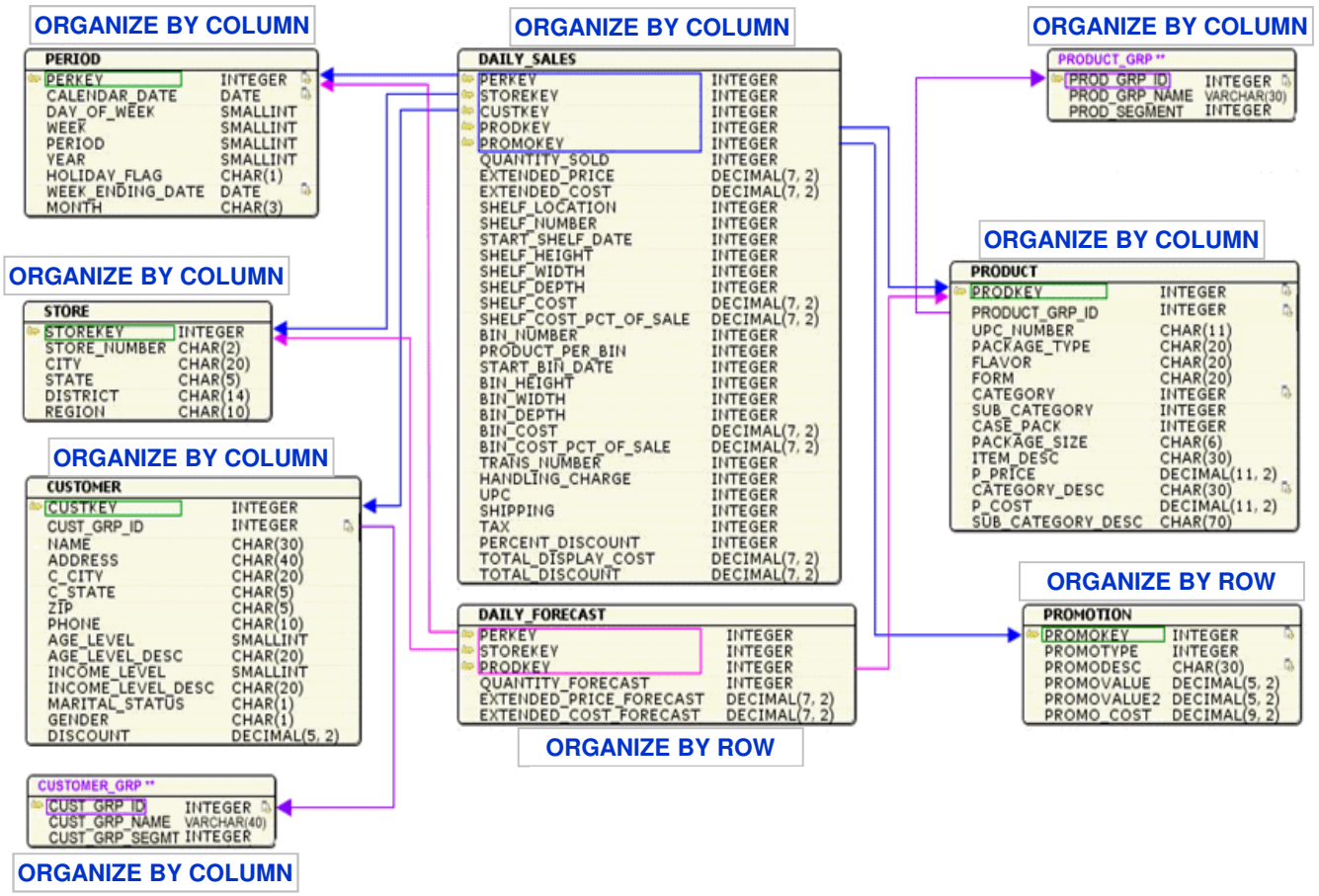
~Sub second 10TB query

- **The system – 32 cores, 10TB table with 100 columns, 10 years of data**
- **The query: `SELECT COUNT(*) from MYTABLE where YEAR = '2010'`**
- **The result: sub second 10TB query! Each CPU core examines the equivalent of just 8MB of data**



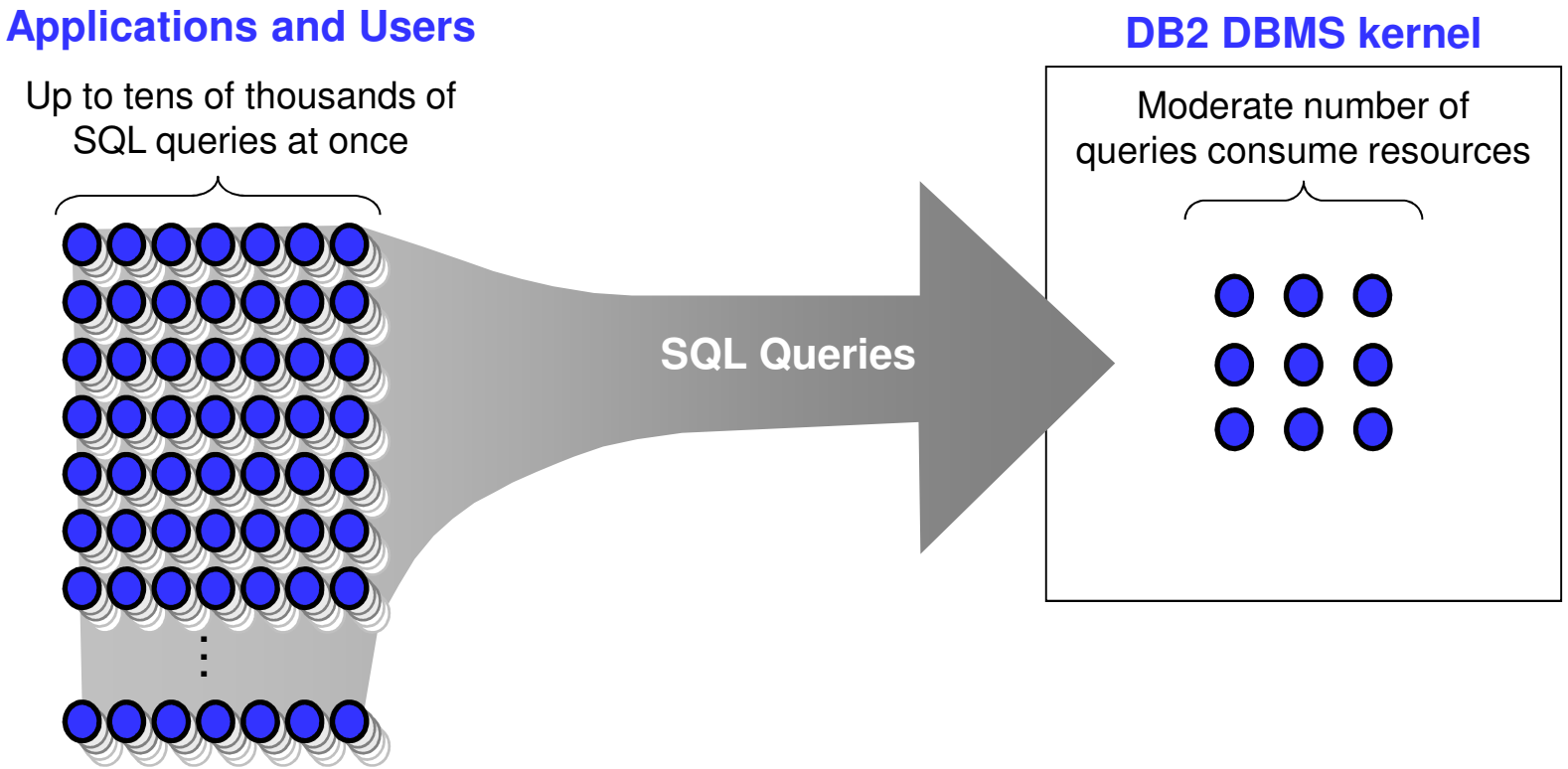
Mixing Row and Columnar Tables

- **DB2 10.5 supports mixing row and columnar tables seamlessly**
  - In the same tablespace and bufferpools
  - In the same query
- **Best query performance for analytic queries usually occurs with all tables columnar**
- **Mixing row and columnar can be necessary**
  - Point queries (highly selective access) favor row-organized tables with index access
  - Small, frequent, write operations favor row-organized tables



Automatic Workload Management

- **Built-in and automated query resource consumption control**
- **Enabled automatically when DB2_WORKLOAD=ANALYTICS**
- **Many queries can be submitted, but limited number get executed concurrently**



Optimize the Entire Hardware Stack

In-Memory Optimized

- **Memory latency optimized for**
  - Scans
  - Joins
  - Aggregation
- **More useful data in memory**
  - Data stays compressed
  - Scan friendly caching
- **Less to put in memory**
  - Columnar access
  - Late materialization
  - Data skipping

CPU Optimized

- **CPU acceleration**
  - SIMD processing for
    - Scans
    - Joins
    - Grouping
    - Arithmetic
- **Keeping the CPUs busy**
  - Core friendly parallelism
- **Less CPU processing**
  - Operate on compressed data
  - Late materialization
  - Data skipping

I/O Optimized

- **Less to read**
  - Columnar I/O
  - Data skipping
  - Late materialization
- **Read less often**
  - Scan friendly caching
- **Efficient I/O**
  - Specialized columnar prefetching algorithm

Value of DB2 BLU Acceleration?

BLU Acceleration

Next Generation Database for Analytics

- **Extreme performance out-of-the-box**
- **Massive storage savings**
  - No indexes required
- **Lower cost of operational analytics**

Seamlessly Integrated

- **Built seamlessly into DB2**
- **Consistent SQL, interfaces, administration**
- **Dramatic simplification**
  - Less to design
  - Less to tune
  - **Just Load and Go**

Hardware Optimized

- **In Memory Optimized**
  - Compressed in memory
- **Modern CPU Exploitation**
- **I/O Optimized**
  - Only read columns of interest

IBM DB2 Analytics Accelerator (IDAA)

Analytics and "HTAP"
Hybrid Transaction/Analytical Processing

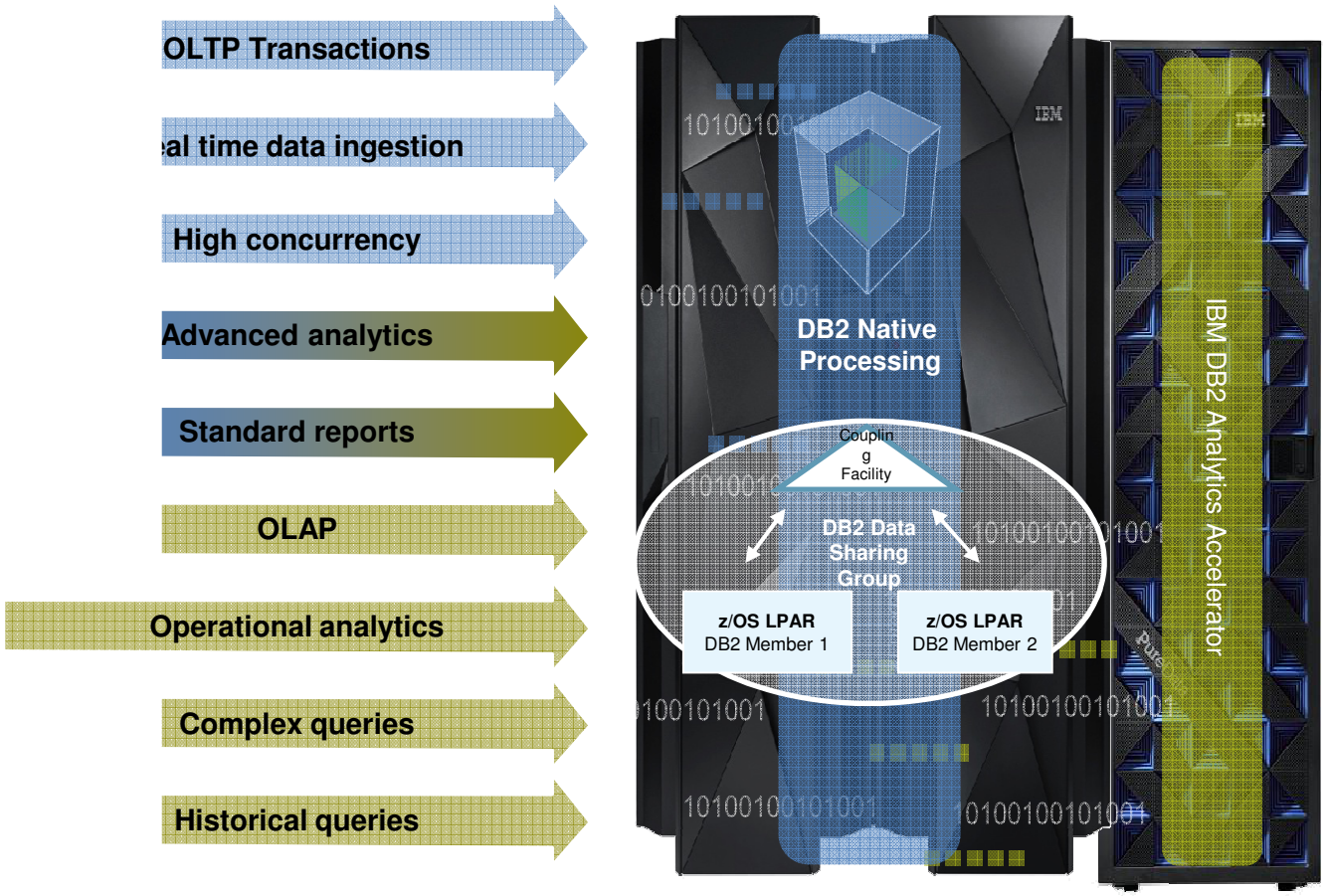
- Much of the world's operational data resides on z/OS systems
  - Many customers ETL this data to other platforms for analytics
  - Information supply chain and ETL: expensive and complex, error prone, security concerns, data not current, multiple copies

- Grow z/OS analytics capabilities so that ETL can be significantly reduced, even avoided

- ***Hybrid Transaction and Analytics Processing*** (HTAP)
  - OLTP and analytics (incl. predictive) in the same database system

- "Big Data" integration

Operations and Analytics Coexistence

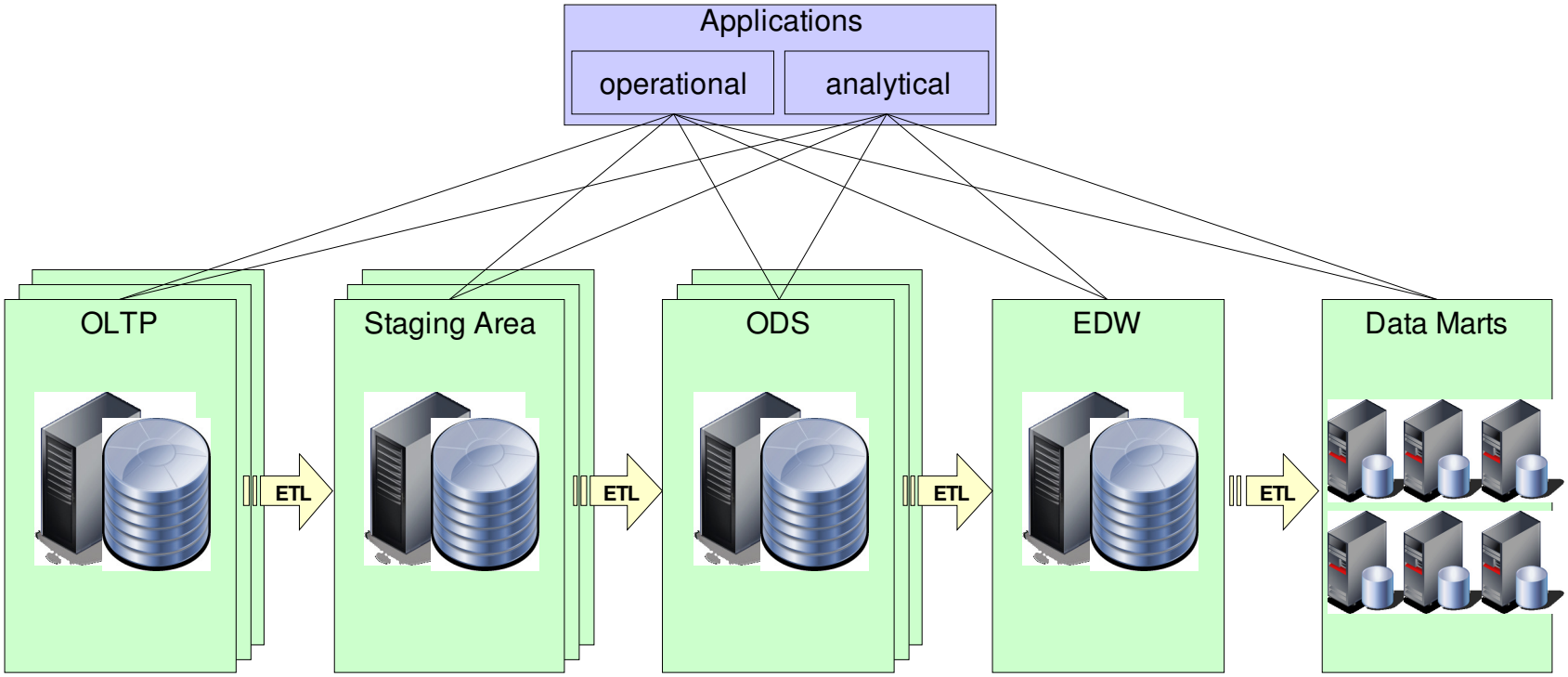


z/OS LPAR Configuration: *zEC12, 8 CPUs; 120 GB Memory; zOS 1.13; DB2 V11 for z/OS*
IBM D2 Analytics Accelerator Configuration: *Striper model, 112 cores; 240 data slices*
IBM D2 Analytics Accelerator Configuration updated via batch processes such as nightly ETL

Two use cases:

- **Operational Priority**
  - Keeping operational throughput constant, add analytics load to the system
  - Data used for analysis can be slightly out of sync with operations
- **Data Priority**
  - Data used for operations and analytics must be in close synchronization
  - Higher latency of operational throughput is acceptable

Traditional Systems Landscape

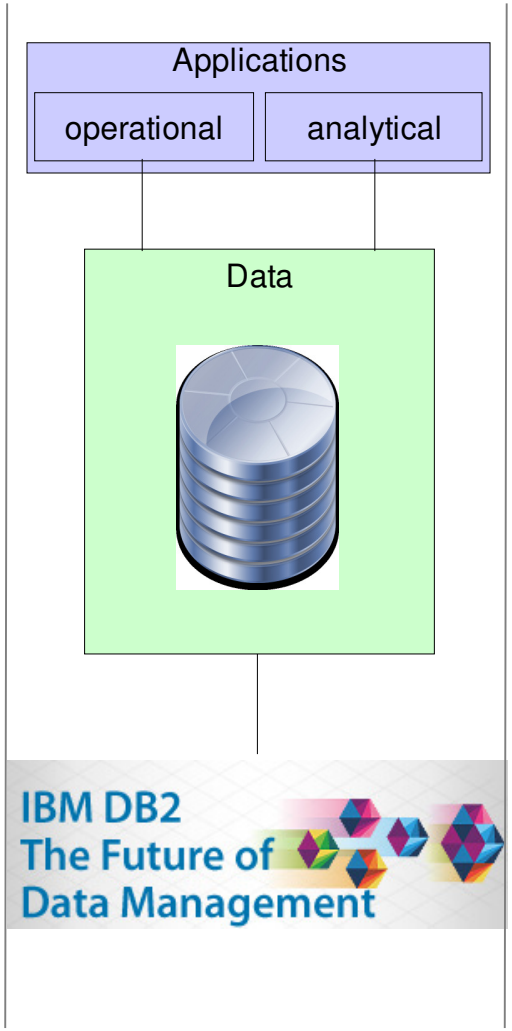


- ***Negative ramifications***:
  - Complexity – both in systems management and in applications
  - Difficulties in supporting real time analytics
  - Inability to match ever more demanding SLA requirements
  - High total cost of ownership

- ***Historical reasons***:
  - Different access patterns → impact on performance
  - EDW as the data integration hub → again, impact on performance
  - Different life-cycle characteristics → and again, impact on performance
  - Different Service Level Agreements (SLA)
    - Lack of broadly available workload management capabilities
    - Choice of lower cost-of-acquisition offerings

Visionary Systems Landscape

- ***Benefits***
  - Uniform policies and procedures for security, HA, DR, monitoring, same tools, same skills, ...
  - Efficient data movement within the system, often not involving network (ELT vs. ETL)
  - Uniform access to any data for types of applications
  - Opportunity to remove, i.e. consolidate some of the layers, ultimately leading to a single database
- ***Approaches***
  - Large RAM
  - 'In-memory' databases
  - Massively parallel processing
  - Large number of sockets, cores, servers
  - Vector processing
  - Hardware acceleration through special purpose processors
  - FPGA, GPU, ...
  - Columnar stores
  - Appliances



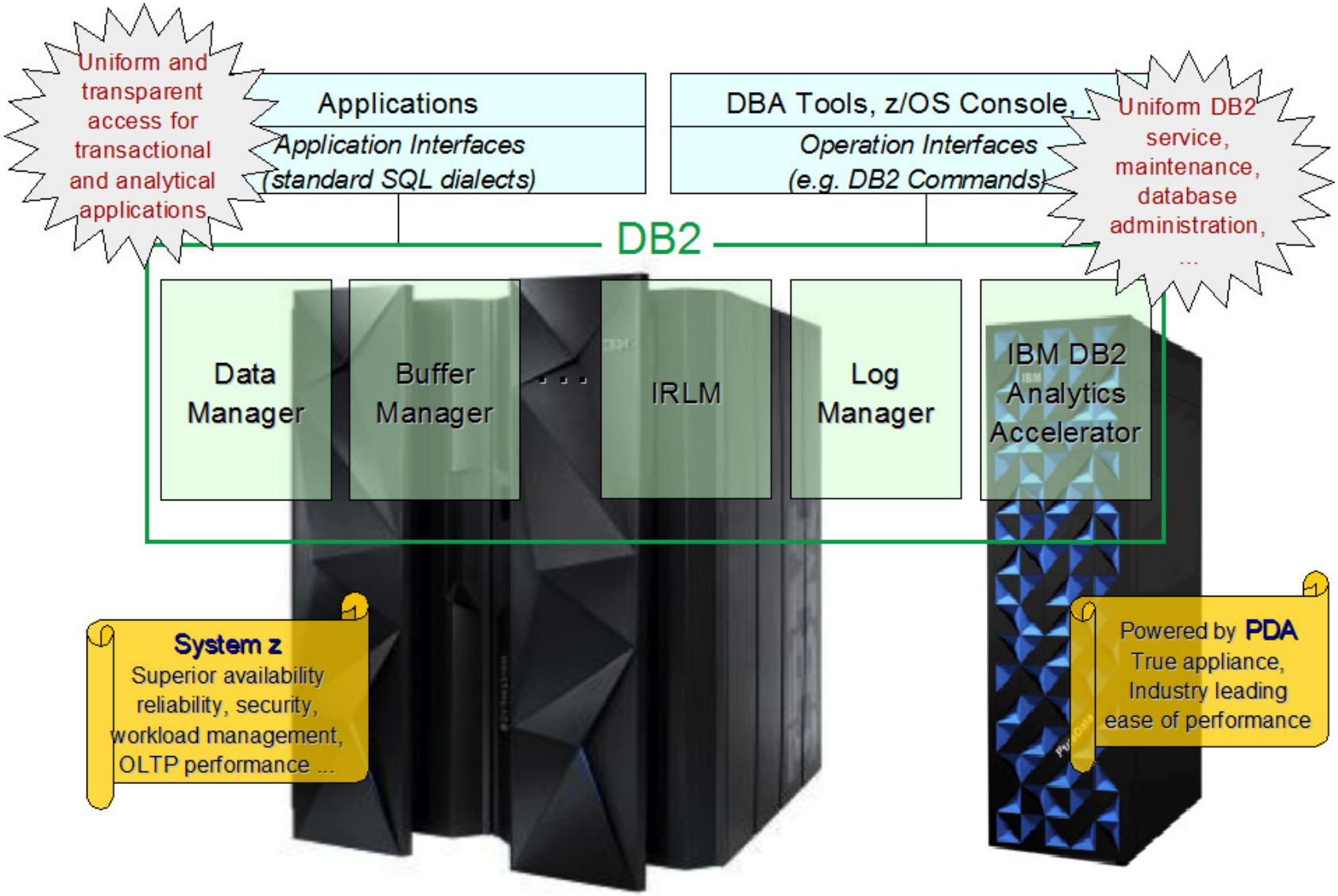
- ***Challenges***
  - ✔ Mixed workload management capabilities
  - ✔ Ensuring continuous availability, security and reliability
  - ✔ Providing seamless scale-up and scale-out
  - ⛔ Providing universal processing capabilities to deliver best performance for both transactional and analytical workloads **without the need for excessive tuning**
- ***Building on proven technology base***
  - DB2 (both z/OS and LUW) already provide superior technology to address most of the challenges
  - The remaining challenge is addressed by adding special purpose processing component for analytical workloads
    - DB2 for z/OS: **IBM DB2 Analytics Accelerator**
    - DB2 for LUW: **BLU**

IBM DB2 Analytics Accelerator (IDAA)

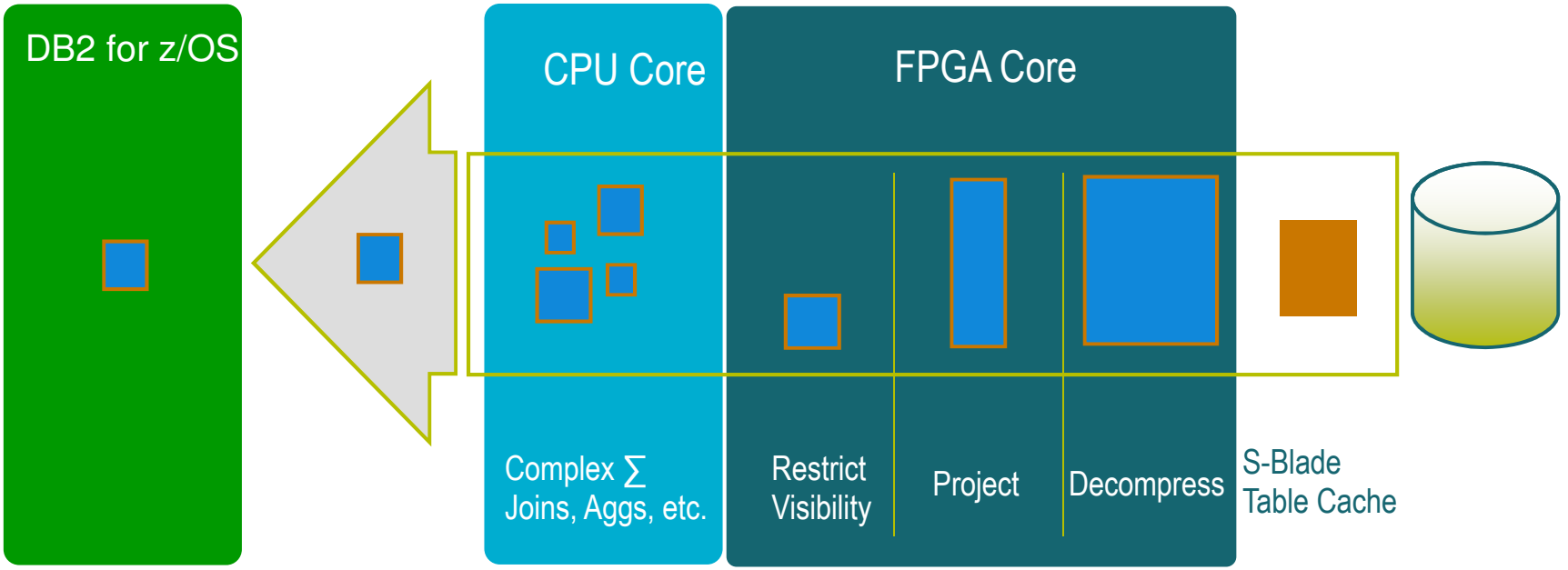
- ***What is it?***
  - A high performance appliance that integrates Netezza technology with zEnterprise technology, to deliver dramatically faster business analytics on **relational** / structured data

- ***What does it do?***
  - Accelerates complex SQL queries, up to 2000x faster (days, hours reduced to seconds)
  - Improves access to and lowers the cost of storing, managing and processing historical data
  - Minimizes latency
  - Reduces zEnterprise capacity requirements
  - Improves security & reduces risk

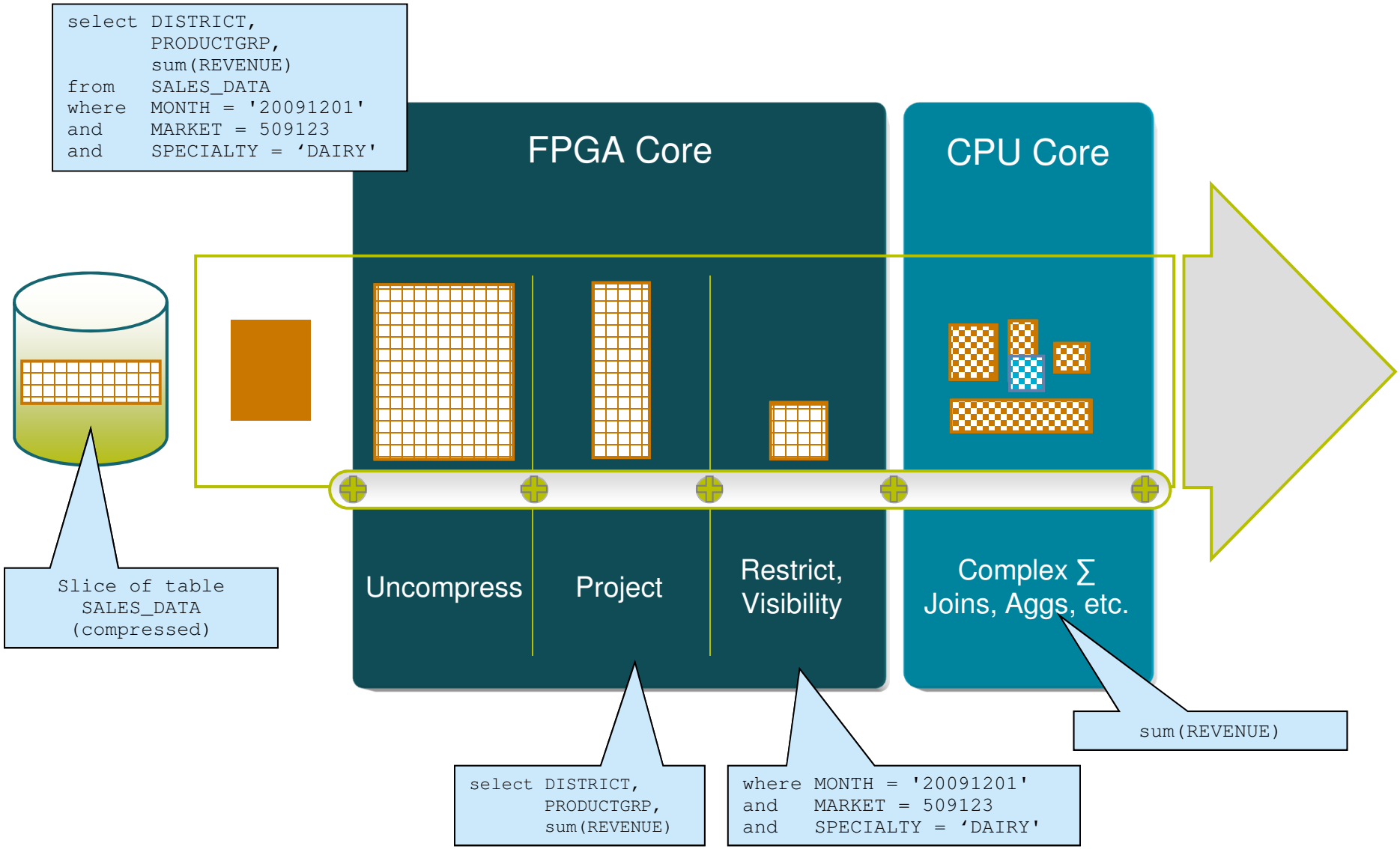
DB2 for z/OS becomes a Hybrid DBMS



Applying Data Stream Processing to DB2 Queries



Field Programmable Gate Arrays (FPGAs)



IDAA executes complex Queries significantly faster

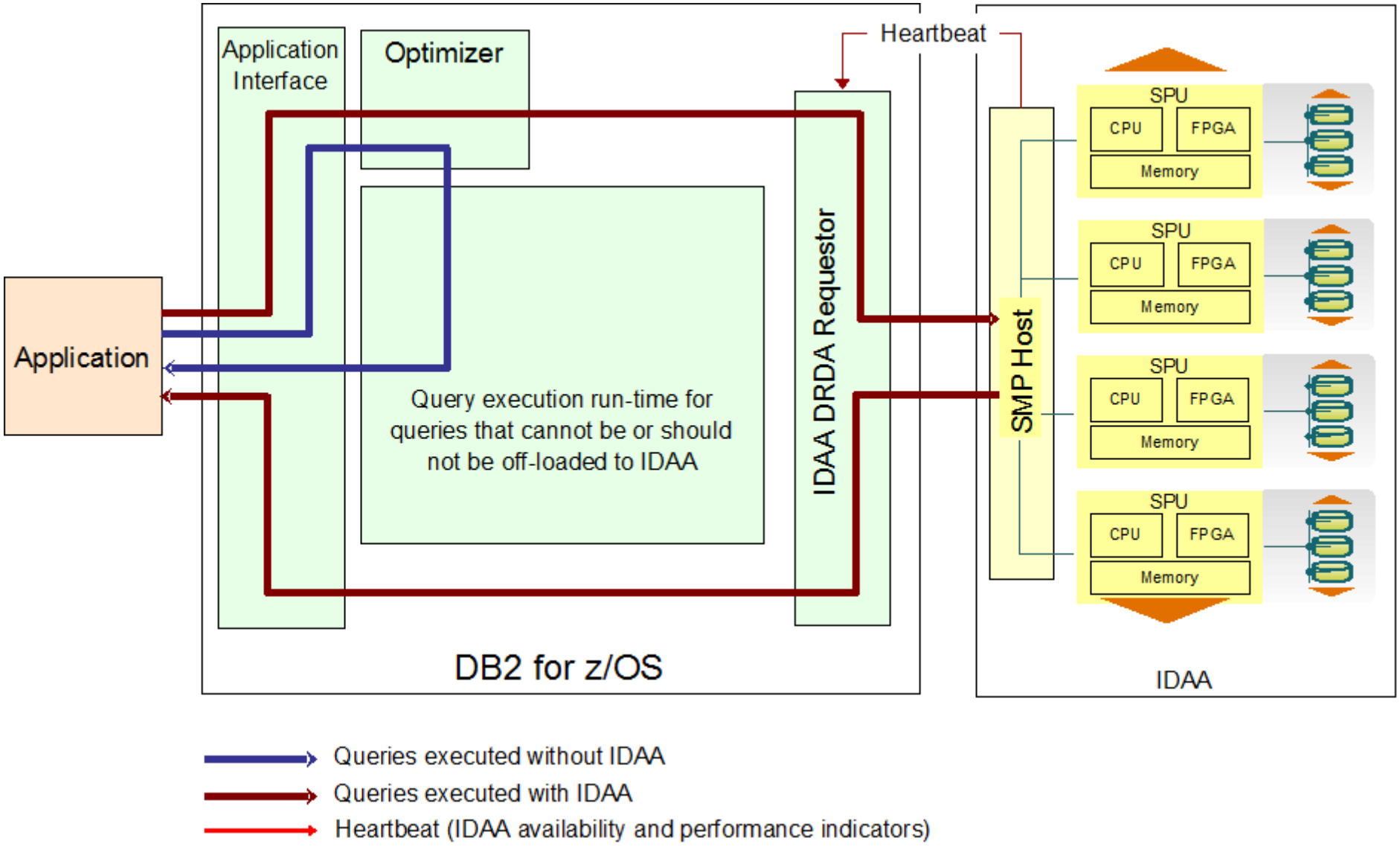
Query	DB2 (Secs)	DB2 + Analytics Accelerator (Secs)	Speed Up	Rows Reviewed	Rows Returned
Query 1	9,540	5	1,908x	2,813,571	853,320
Query 2	8,220	5	1,644x	2,813,571	585,780
Query 3	4,560	6	760x	8,260,214	274
Query 4	4,080	5	816x	2,813,571	601,197
Query 5	4,080	70	58x	3,422,765	508
Query 6	3,180	6	530x	4,290,648	165
Query 7	3,120	4	780x	361,521	58,236
Query 8	2,640	2	1,320x	342,529	724
Query 9	2,520	193	13x	4,130,107	137

Run analytic workloads on the same platform as the operational data

- IBM DB2 Analytics Accelerator based on Netezza technology
- Integrated with DB2 for z/OS, transparent to the application
- Unprecedented response times – complex queries run in seconds instead of hours

Actual customer results, October 2011

DB2 for z/OS: Query Execution Process Flow



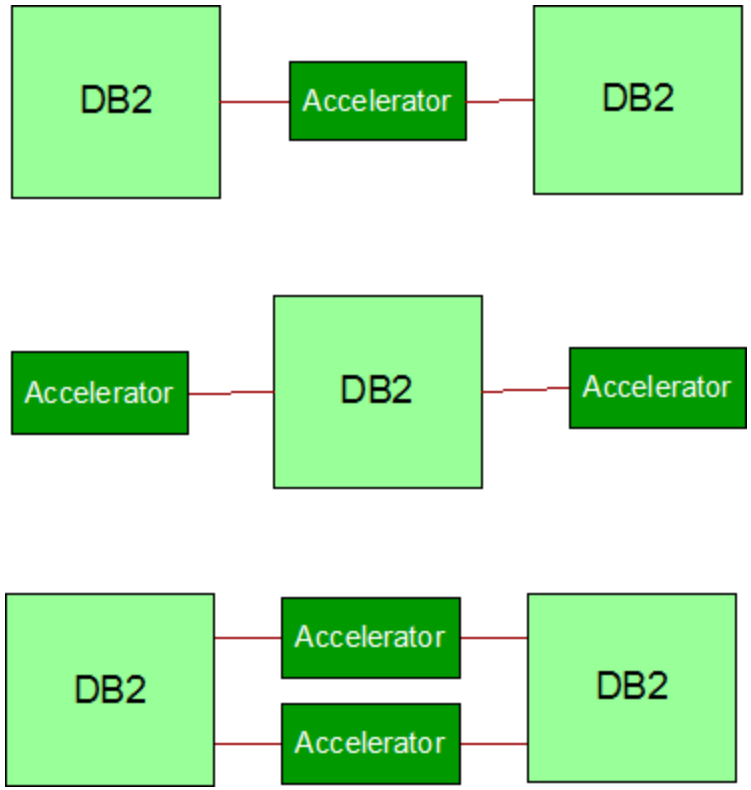
Connectivity Options

1. **Multiple DB2 systems can connect to a single accelerator**

2. **A single DB2 system can connect to multiple accelerator**

3. **Multiple DB2 systems can connect to multiple accelerator**

Policy based workload management
Better utilization of accelerator resources
Scalability
High availability

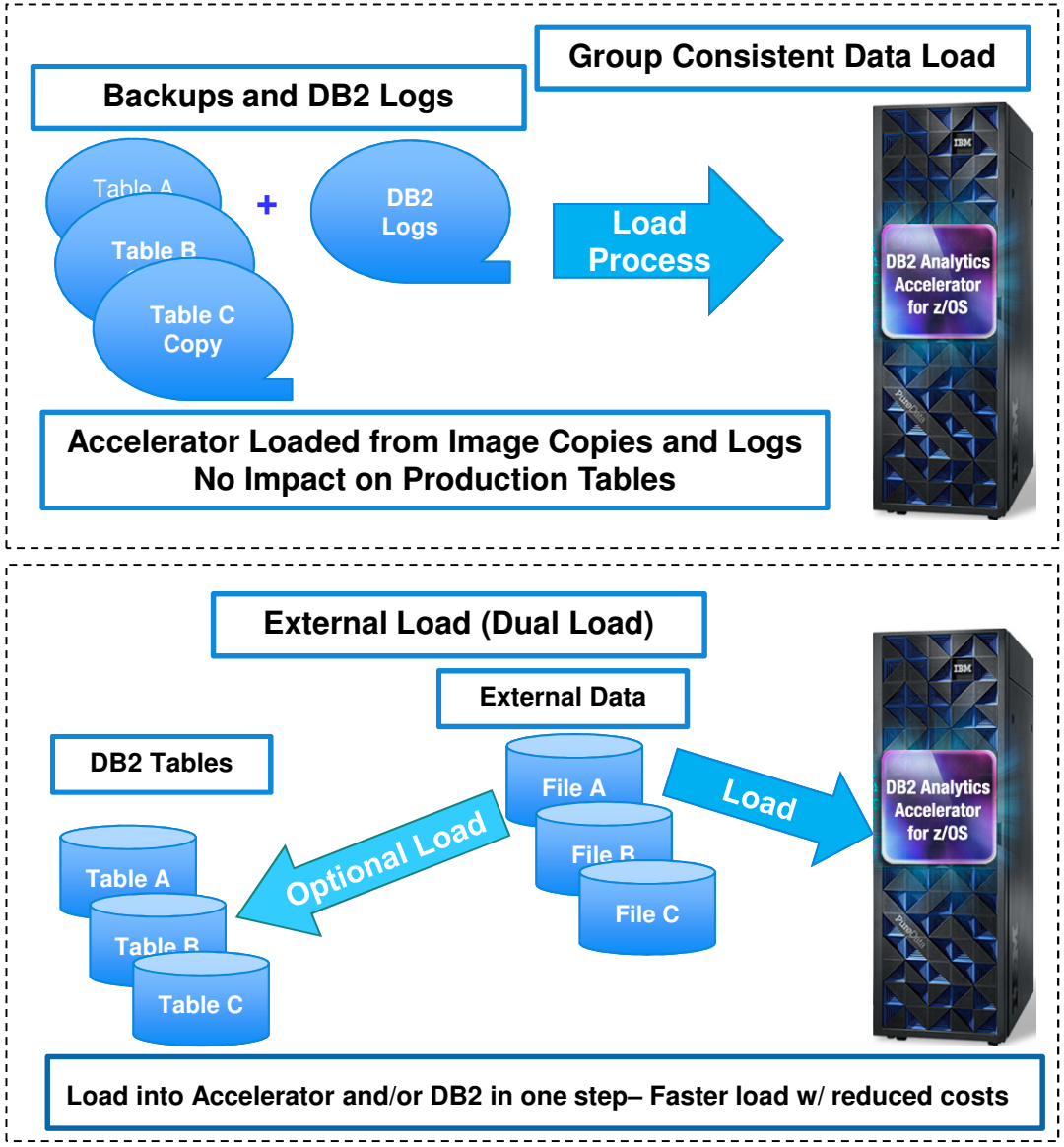


- **Full flexibility for DB2 systems:**
  - Residing in the same LPAR
  - Residing in different LPARs
  - Residing in different CECs
  - Being independent (non-data sharing)
  - Belonging to the same data sharing group
  - Belonging to different data sharing groups

In-Memory Database

- **In-memory DBMS has existed for many years**
  - IBM divested SolidDB to UNICOM on June 30, 2014
- **In-memory concepts apply for both row and column store formats**
- **DB2 for z/OS incorporates extensive in-memory technology and operates almost exclusively on in-memory data**
  - DB2 keeps frequently accessed data in memory (buffer pools)
    - Avoids disk I/O: > 90% of data accessed in memory without I/O
    - Prefetch mechanisms avoid I/O waits
    - Option to pin a table in memory
  - DB2 writes all data changes (INSERT, UPDATE, DELETE) to memory
    - Persistently writes log records to disk by commit time
      - Same behavior as In-Memory Databases
- **System z servers support TB sized memories, and we expect memory sizes to increase – DB2 will evolve accordingly**
- **Coupling Facility main memory is a unique opportunity for System z**

IDAA Loader for z/OS – Flexible Load Options



- Loading data from a file to Accelerator and/or DB2:
  - Building a new data warehouse
    - Possibly bringing data to DB2 on z
  - Users desire to load file into DB2, Accelerator, or both
  - Can be extracted from DB2 or other sources
- Loading data to Accelerator from operational DB2 tables
  - Operational data store (ODS) and EDW w/ daily point in time refresh
  - Require transaction consistent data
  - Historical point-in-time load

Data Synchronization Options

Synchronization options	Use cases, characteristics and requirements
Full table refresh The entire content of a database table is refreshed for accelerator processing	• Existing ETL process replaces entire table • Multiple sources or complex transformations • Smaller, un-partitioned tables • Reporting based on consistent snapshot • Need for refresh automatically detected
Table partition refresh For a partitioned database table, selected partitions can be refreshed for accelerator processing	• Optimization for partitioned warehouse tables, typically appending changes “at the end” • More efficient than full table refresh for larger tables • Reporting based on consistent snapshot • Need for refresh automatically detected
Incremental update Log-based capturing of changes and propagation to IDAA with low latency (typically few minutes)	• Scattered updates after “bulk” load • Reporting on continuously updated data (e.g., an ODS), considering most recent changes • More efficient for smaller updates than full table refresh • Applications can request reporting on committed data only
IBM Analytics LOAD utility IDAA load utility used to manage IDAA	• Permits to load DB2 tables and IDAA in the same time • Permits to load on IDAA only sequential file • Permits to load IDAA table only using Image Copy and LOG