# Deep Dive into IBM DB2 Analytics Accelerator Query Acceleration

Maryela Weihrauch and Ruiping Li

DB2 for z/OS, IBM Silicon Valley

February 8, 2013

Session 12762

# Agenda

- IBM DB2 Analytic Accelerator design objectives
- Overall architecture and usage cycle
- Query acceleration
- Disaster recovery
- High Performance Storage Saver
- Workload assessment

# Fit The Solution To The Workload – Access to all Data

| Business Analytics / ODS | OLTP |
|---|---|
| - Complex queries against a data warehouse<br>- Read only<br>- Star schema design often used to speed up queries<br>- Different modes of operation | - Transactions against operational data<br>- Reads and writes<br>- Normalized schema design to eliminate redundancy<br>- Multiple user throughput operation |

Tightly integrated solution optimized for workloads

Simplification

Single entry point to the true data

High fidelity data

Performance

| IBM DB2 Analytics Accelerator<br>DB2 for z/OS ODS members | DB2 for z/OS OLTP members |
|---|---|

# DB2 Analytics Accelerator

*Accelerating decisions to the speed of business*

*Blending System z and Netezza technologies to deliver unparalleled, mixed workload performance for complex analytic business needs.*

**Get more insight from your data timely**

- Fast, predictable response times for “right-time” analysis
- Accelerate analytic query response times
- Improve price/performance for analytic workloads
- Minimize the need to create data marts for performance
- Highly secure environment for sensitive data analysis
- Transparent to the application

# IBM DB2 Analytic Accelerator Design Objectives: *it's an accelerator*

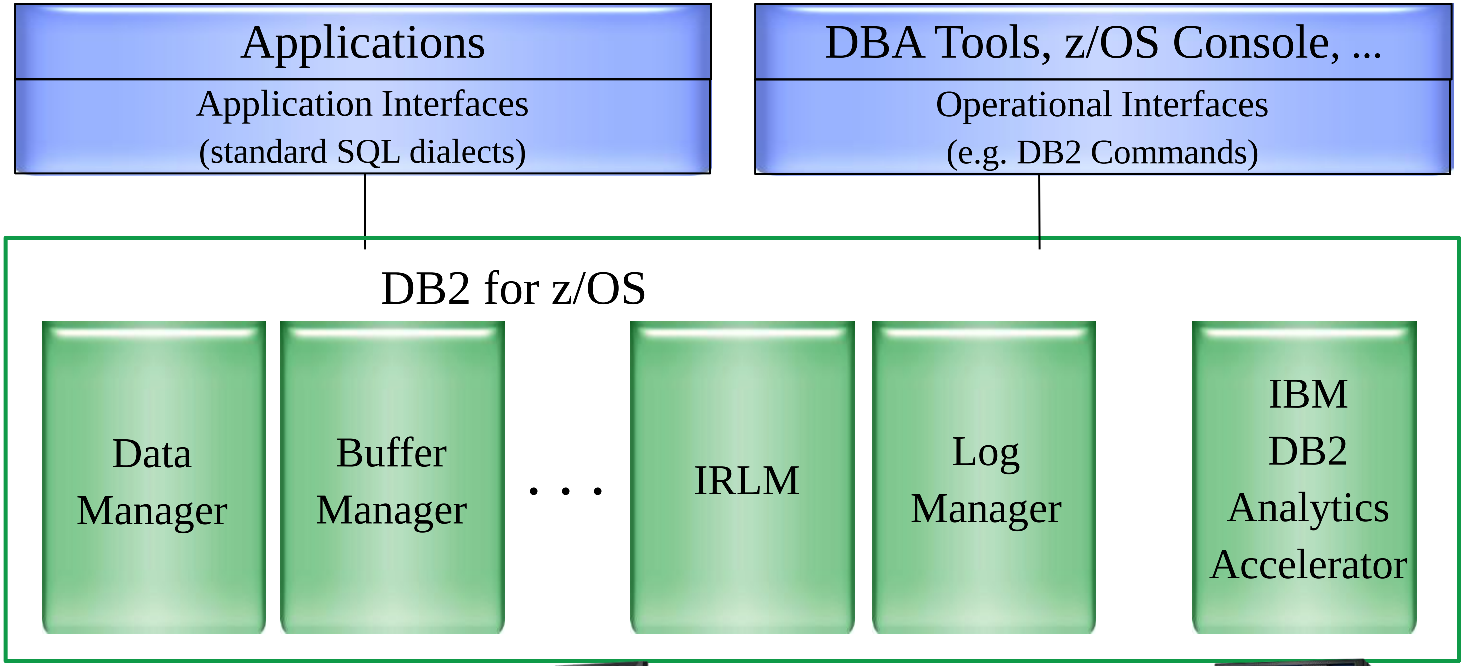
- **DB2 continues to own data (both OLTP and DW)**
  - Access to data (authorization, privileges, …)
  - Data consistency and integrity (backup, recovery, …)
  - Enables extending System z QoS characteristics to BI/DW data as well
- **Applications access data (both OLTP and DW) only through DB2**
  - User or DB2 controls whether to execute query in DB2 mainline or route to IBM DB2 Analytic Accelerator
  - DB2 returns results to the calling application
  - Existing applications do not have to change.
  - Enables mixed workloads and selection of optimal access path (within DB2 or IBM DB2 Analytic Accelerator) depending on access pattern
- **IBM DB2 Analytic Accelerator as a virtual DB2 component**
  - DB2 provides key IBM DB2 Analytic Accelerator status and performance indicators as well as typical administration tasks by standard DB2 interfaces and means
  - No direct access (log-on) to IBM DB2 Analytic Accelerator accelerator
  - Enables operational cost reduction through skills, tools and processes consolidation

# Feedback from Customers: Fast Time to Value

- IBM DB2 Analytics Accelerator
  - ➔ Production ready - 1 person, 2 days
- Table Acceleration Setup … **2 Hours**
  - DB2 "Add Accelerator"
  - Choose a Table for "Acceleration"
  - Load the Table (DB2 copy to Netezza)
  - Knowledge Transfer
  - Query Comparisons
- Initial Load Performance …
  - ➔ 5.1 GB in 1 Min 25 Seconds (24M rows)
    400 GB in 29 Min (570M rows)
- Actual Query Acceleration … **up to 1908x as fast**
  - ➔ 2 Hours 39 Minutes to 5 Seconds
- CPU Utilization Reduction … **99% less CPU**
  - ➔ 24M rows: 56.5 CPU seconds to 0.4 CPU seconds

Actual customer results, October 2011

# Deep DB2 Integration



Applications

Application Interfaces
(standard SQL dialects)

DBA Tools, z/OS Console, ...

Operational Interfaces
(e.g. DB2 Commands)

DB2 for z/OS

Data Manager

Buffer Manager

. . .

IRLM

Log Manager

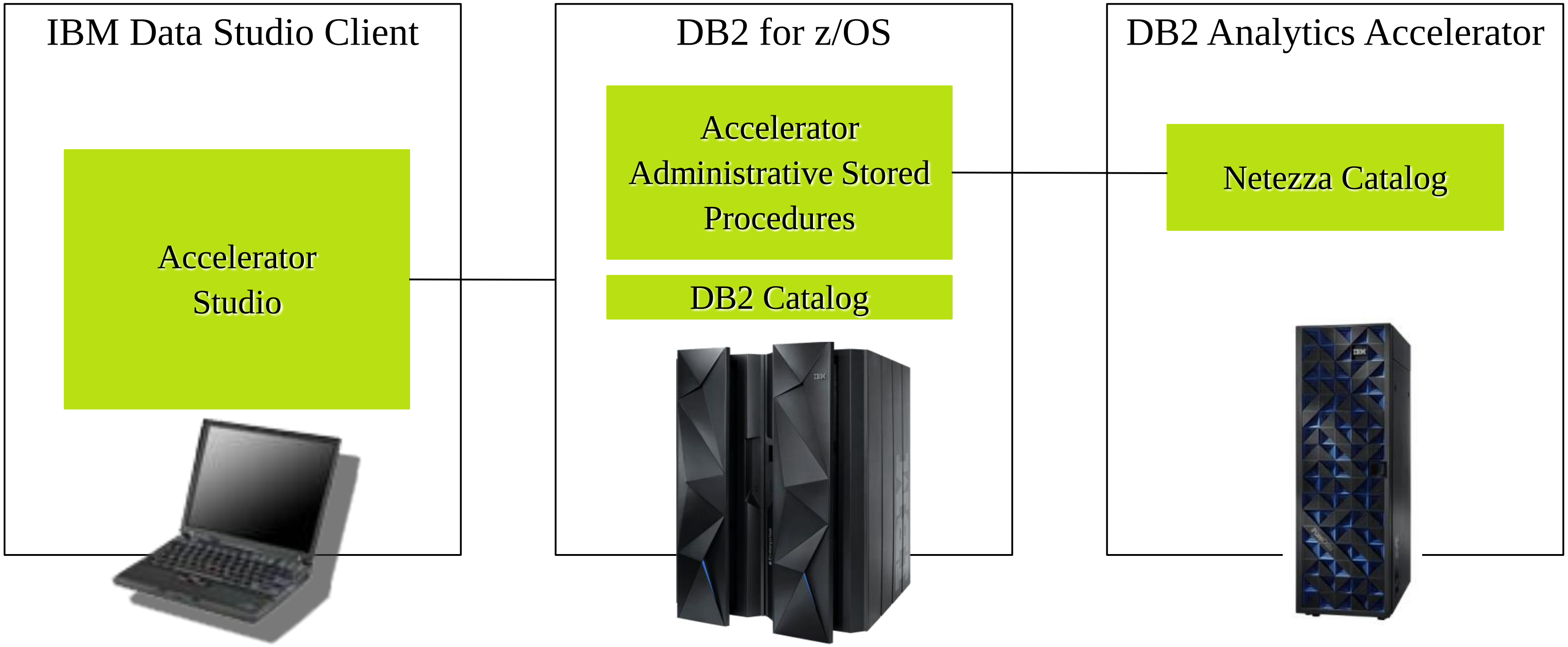
IBM DB2 Analytics Accelerator

Superior availability reliability, security, Workload management

z/OS on System z

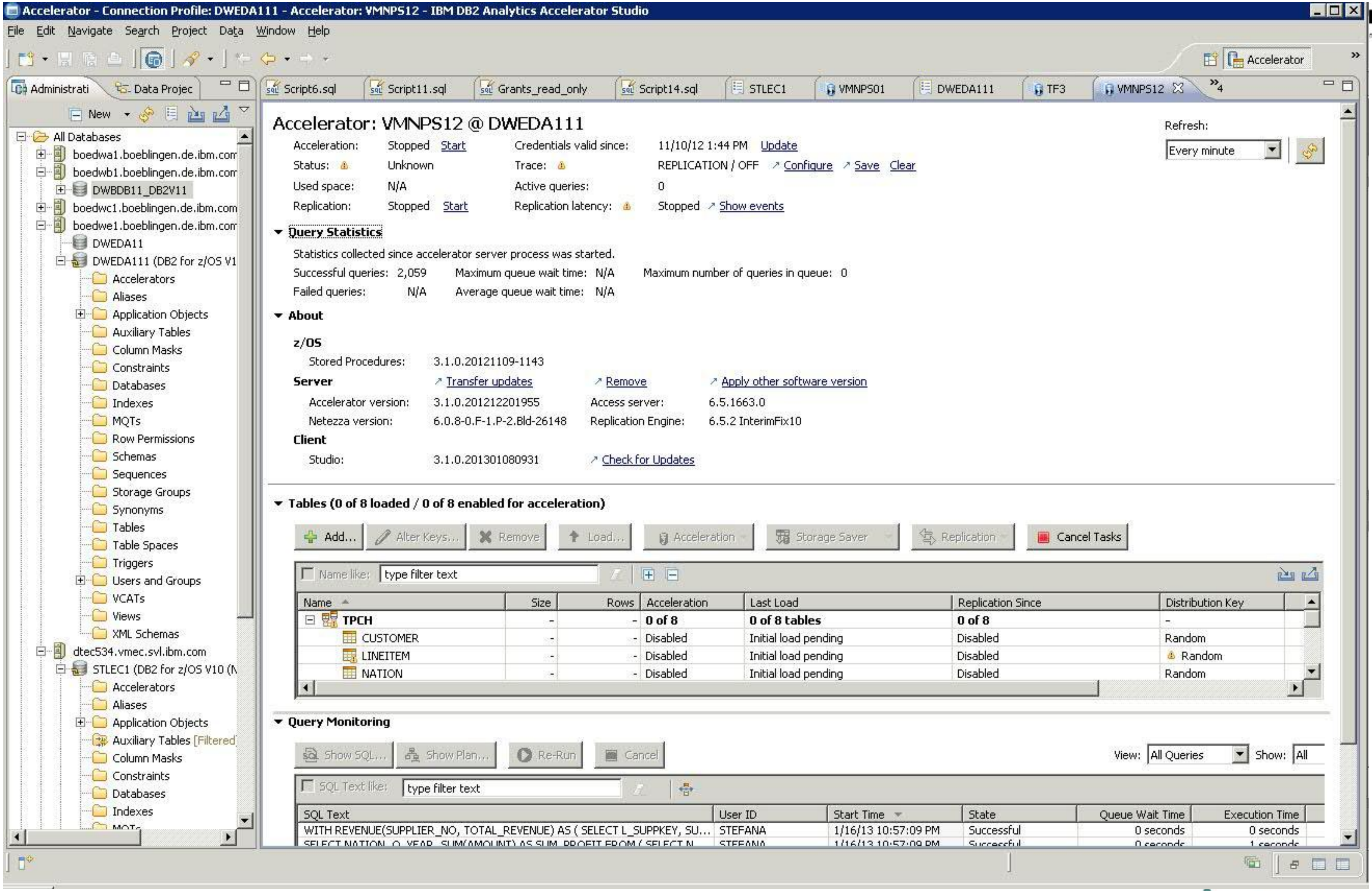
Superior performance on analytic queries

# Accelerator Table Definition and Deployment



- The tables need to be defined and deployed to the Accelerator before data is loaded and queries sent to it for processing.
  - Definition: identifying tables for which queries need to be accelerated
  - Deployment: making tables known to DB2, i.e. storing table meta data in the DB2 and Netezza catalog.
- IBM DB2 Analytics Accelerator Studio guides you through the process of defining and deploying tables, as well as invoking other administrative tasks.
- IBM DB2 Analytics Accelerator Stored Procedures implement and execute various administrative operations such as table deployment, load and update, and serve as the primary administrative interface to the Accelerator from the outside world including Accelerator Studio.

# Studio GUI Overview


Accelerator - Connection Profile: DWEDA111 - Accelerator: VMNPS12 - IBM DB2 Analytics Accelerator Studio
File Edit Navigate Search Project Data Window Help
Accelerator
Administrati
Data Projec
Script6.sql
Script11.sql
Grants_read_only
Script14.sql
STLEC1
VMNPS01
DWEDA111
TF3
VMNPS12
New

- All Databases
  - boedwa1.boeblingen.de.ibm.com
  - boedwb1.boeblingen.de.ibm.com
    - DWBDB11_DB2V11
  - boedwc1.boeblingen.de.ibm.com
  - boedwe1.boeblingen.de.ibm.com
    - DWEDA11
    - DWEDA111 (DB2 for z/OS V1
      - Accelerators
      - Aliases
      - Application Objects
      - Auxiliary Tables
      - Column Masks
      - Constraints
      - Databases
      - Indexes
      - MQTs
      - Row Permissions
      - Schemas
      - Sequences
      - Storage Groups
      - Synonyms
      - Tables
      - Table Spaces
      - Triggers
      - Users and Groups
      - VCATs
      - Views
      - XML Schemas
  - dtec534.vmec.svl.ibm.com
    - STLEC1 (DB2 for z/OS V10 (N
      - Accelerators
      - Aliases
      - Application Objects
      - Auxiliary Tables [Filtered]
      - Column Masks
      - Constraints
      - Databases
      - Indexes
      - MQTs

Accelerator: VMNPS12 @ DWEDA111

Refresh: Every minute

Acceleration: Stopped Start  
Status: Unknown  
Used space: N/A  
Replication: Stopped Start

Credentials valid since: 11/10/12 1:44 PM Update  
Trace: REPLICATION / OFF Configure Save Clear  
Active queries: 0  
Replication latency: Stopped Show events

**Query Statistics**

Statistics collected since accelerator server process was started.

Successful queries: 2,059 Maximum queue wait time: N/A Maximum number of queries in queue: 0  
Failed queries: N/A Average queue wait time: N/A

**About**

**z/OS**  
Stored Procedures: 3.1.0.20121109-1143

**Server** Transfer updates Remove Apply other software version  
Accelerator version: 3.1.0.201212201955 Access server: 6.5.1663.0  
Netezza version: 6.0.8-0.F-1.P-2.Bld-26148 Replication Engine: 6.5.2 InterimFix10

**Client**  
Studio: 3.1.0.201301080931 Check for Updates

**Tables (0 of 8 loaded / 0 of 8 enabled for acceleration)**

Add... | Alter Keys... | Remove | Load... | Acceleration | Storage Saver | Replication | Cancel Tasks

☐ Name like: type filter text

| Name | Size | Rows | Acceleration | Last Load | Replication Since | Distribution Key |
|---|---|---|---|---|---|---|
| TPCH | - | - | 0 of 8 | 0 of 8 tables | 0 of 8 | - |
| CUSTOMER | - | - | Disabled | Initial load pending | Disabled | Random |
| LINEITEM | - | - | Disabled | Initial load pending | Disabled | Random |
| NATION | - | - | Disabled | Initial load pending | Disabled | Random |

**Query Monitoring**

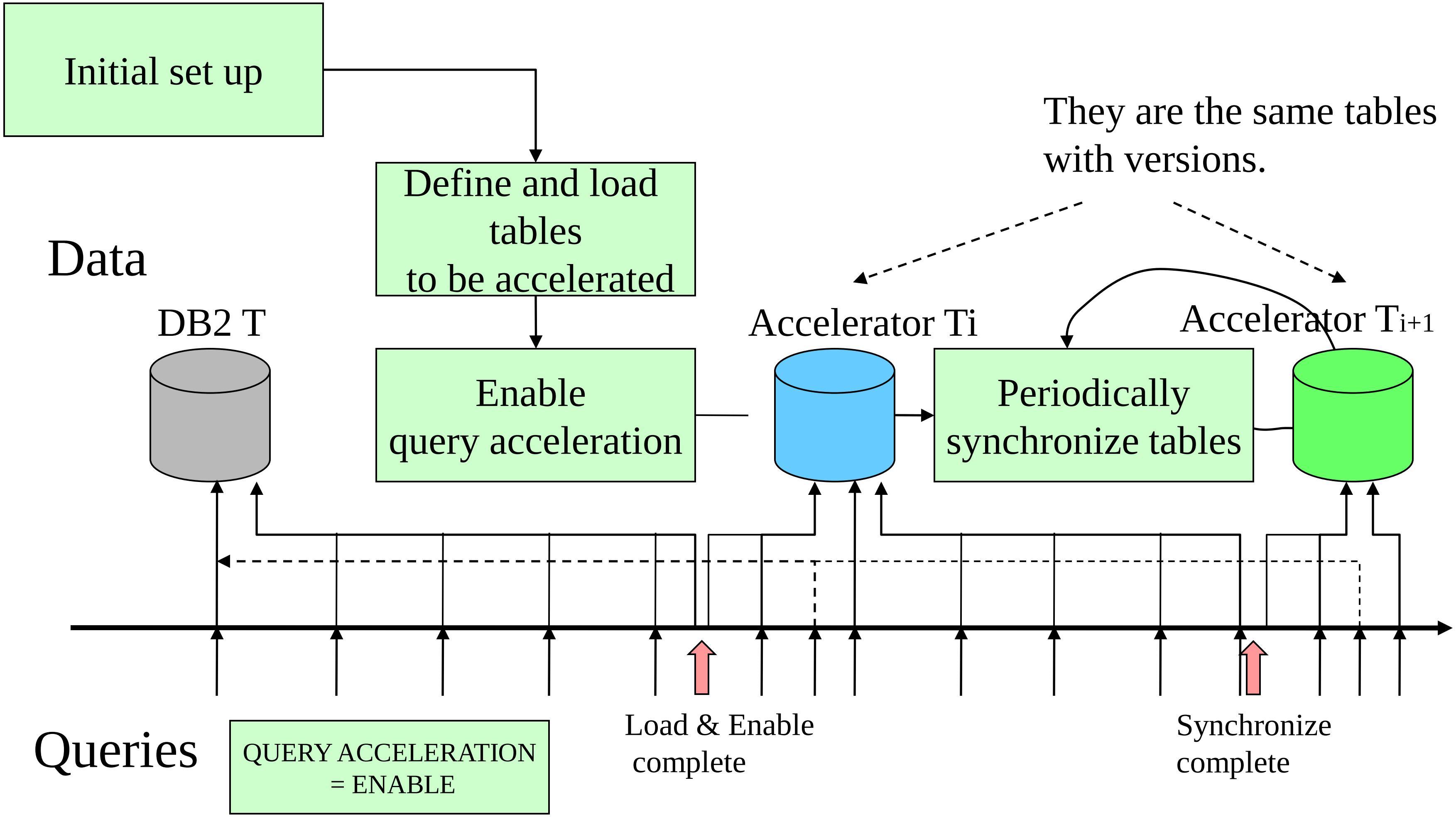
Show SQL... | Show Plan... | Re-Run | Cancel

View: All Queries Show: All

☐ SQL Text like: type filter text

| SQL Text | User ID | Start Time | State | Queue Wait Time | Execution Time |
|---|---|---|---|---|---|
| WITH REVENUE(SUPPLIER_NO, TOTAL_REVENUE) AS ( SELECT L_SUPPKEY, SU... | STEFANA | 1/16/13 10:57:09 PM | Successful | 0 seconds | 0 seconds |
| SELECT NATION, O_YEAR, SUM(AMOUNT) AS SUM_PROFIT FROM ( SELECT N... | STEFANA | 1/16/13 10:57:09 PM | Successful | 0 seconds | 1 seconds |

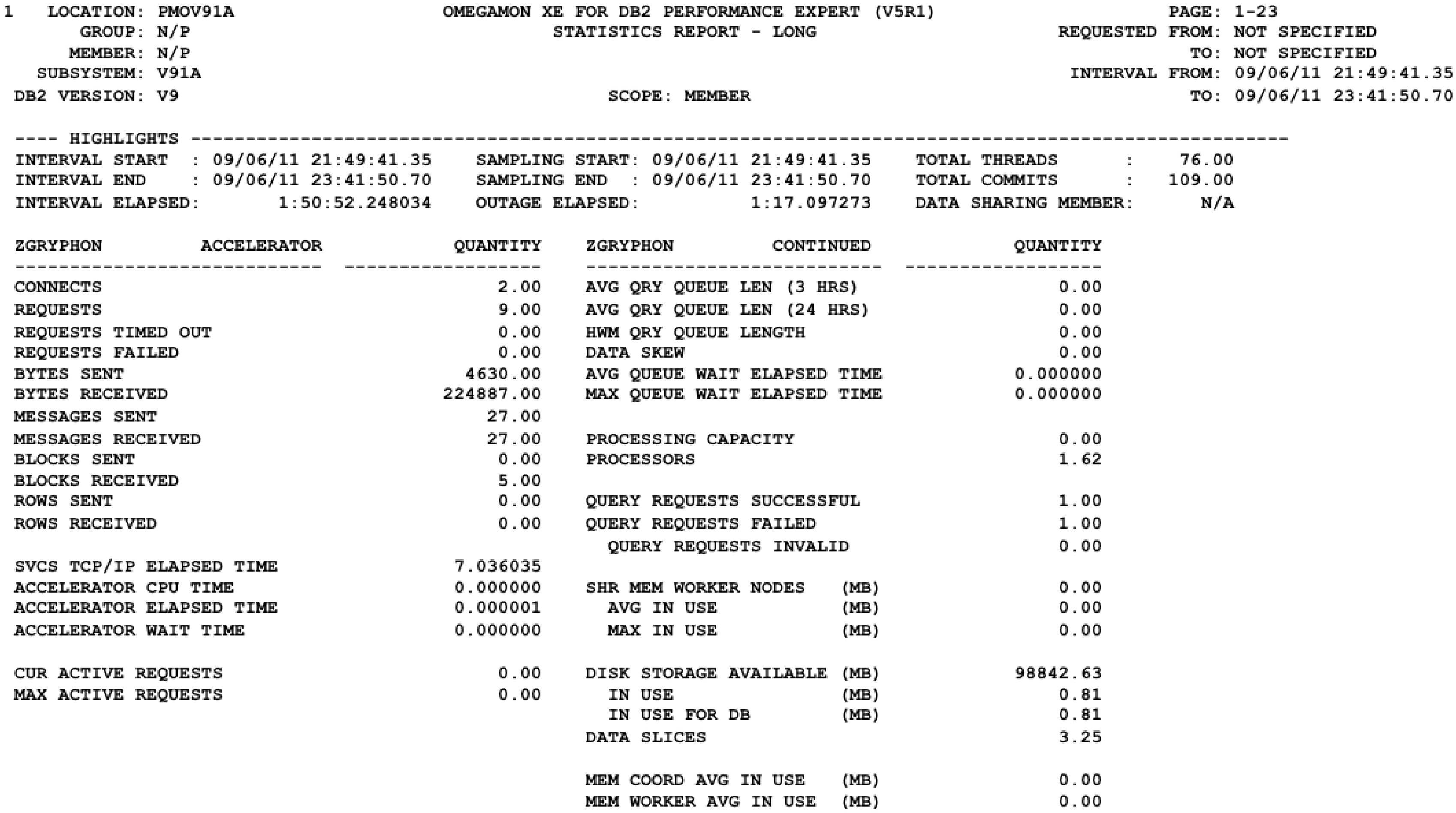
# Usage Cycles: Deploy, Enable and Refresh


Initial set up
Define and load tables to be accelerated
They are the same tables with versions.
Data
DB2 T
Accelerator Ti
Accelerator Ti+1
Enable query acceleration
Periodically synchronize tables
Queries
QUERY ACCELERATION = ENABLE
Load & Enable complete
Synchronize complete

# Synchronization Options

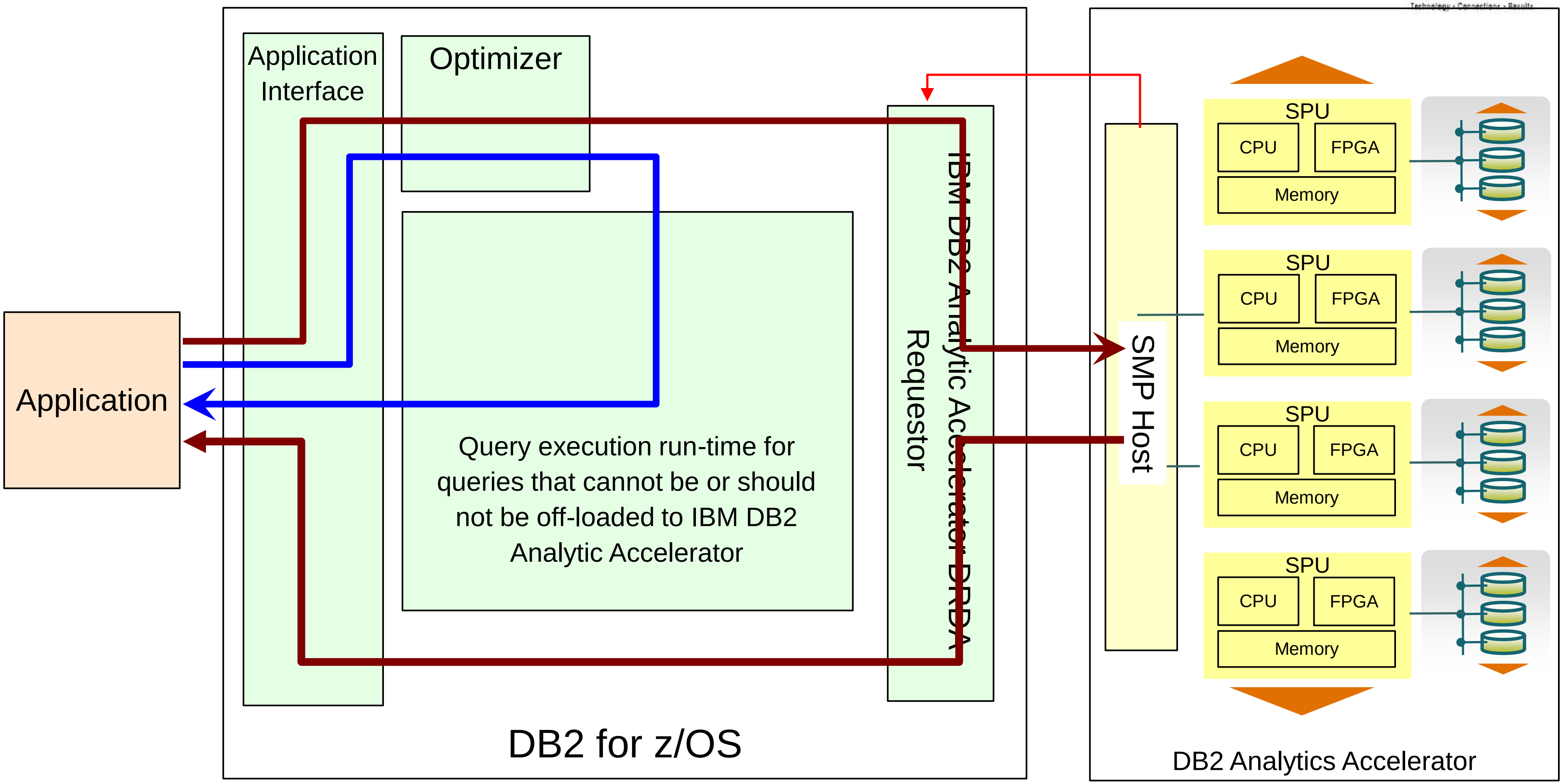
| Synchronization options | Use cases, characteristics and requirements |
|---|---|
| **Full table refresh**<br>The entire content of a database table is refreshed for accelerator processing | ▪ Existing ETL process replaces entire table<br>▪ Multiple sources or complex transformations<br>▪ Smaller, un-partitioned tables<br>▪ Reporting based on consistent snapshot |
| **Table partition refresh**<br>For a partitioned database table, selected partitions can be refreshed for accelerator processing | ▪ Optimization for partitioned warehouse tables, typically appending changes "at the end"<br>▪ More efficient than full table refresh for larger tables<br>▪ Reporting based on consistent snapshot |
| **Incremental update**<br>Log-based capturing of changes and propagation to the accelerator with low latency (typically few minutes) | ▪ Scattered updates after "bulk" load<br>▪ Reporting on continuously updated data (e.g., an ODS), considering most recent changes<br>▪ More efficient for smaller updates than full table refresh |

# Instrumentation Enhanced for Statistics and Accounting



```
1    LOCATION: PMOV91A                    OMEGAMON XE FOR DB2 PERFORMANCE EXPERT (V5R1)                    PAGE: 1-23
        GROUP: N/P                                  STATISTICS REPORT - LONG                    REQUESTED FROM: NOT SPECIFIED
       MEMBER: N/P                                                                                         TO: NOT SPECIFIED
    SUBSYSTEM: V91A                                                                              INTERVAL FROM: 09/06/11 21:49:41.35
 DB2 VERSION: V9                                         SCOPE: MEMBER                                      TO: 09/06/11 23:41:50.70

 ---- HIGHLIGHTS ----------------------------------------------------------------------------------------------------
 INTERVAL START   : 09/06/11 21:49:41.35    SAMPLING START: 09/06/11 21:49:41.35    TOTAL THREADS      :    76.00
 INTERVAL END     : 09/06/11 23:41:50.70    SAMPLING END  : 09/06/11 23:41:50.70    TOTAL COMMITS      :   109.00
 INTERVAL ELAPSED:       1:50:52.248034    OUTAGE ELAPSED:          1:17.097273    DATA SHARING MEMBER:      N/A

 ZGRYPHON        ACCELERATOR           QUANTITY    ZGRYPHON        CONTINUED            QUANTITY
 ---------------------------  ------------------    ---------------------------  ------------------
 CONNECTS                                  2.00    AVG QRY QUEUE LEN (3 HRS)                 0.00
 REQUESTS                                  9.00    AVG QRY QUEUE LEN (24 HRS)                0.00
 REQUESTS TIMED OUT                        0.00    HWM QRY QUEUE LENGTH                      0.00
 REQUESTS FAILED                           0.00    DATA SKEW                                 0.00
 BYTES SENT                             4630.00    AVG QUEUE WAIT ELAPSED TIME           0.000000
 BYTES RECEIVED                       224887.00    MAX QUEUE WAIT ELAPSED TIME           0.000000
 MESSAGES SENT                            27.00
 MESSAGES RECEIVED                        27.00    PROCESSING CAPACITY                       0.00
 BLOCKS SENT                               0.00    PROCESSORS                                1.62
 BLOCKS RECEIVED                           5.00
 ROWS SENT                                 0.00    QUERY REQUESTS SUCCESSFUL                 1.00
 ROWS RECEIVED                             0.00    QUERY REQUESTS FAILED                     1.00
                                                     QUERY REQUESTS INVALID                  0.00
 SVCS TCP/IP ELAPSED TIME              7.036035
 ACCELERATOR CPU TIME                  0.000000    SHR MEM WORKER NODES   (MB)               0.00
 ACCELERATOR ELAPSED TIME              0.000001      AVG IN USE           (MB)               0.00
 ACCELERATOR WAIT TIME                 0.000000      MAX IN USE           (MB)               0.00

 CUR ACTIVE REQUESTS                       0.00    DISK STORAGE AVAILABLE (MB)           98842.63
 MAX ACTIVE REQUESTS                       0.00      IN USE               (MB)               0.81
                                                     IN USE FOR DB        (MB)               0.81
                                                   DATA SLICES                               3.25

                                                   MEM COORD AVG IN USE   (MB)               0.00
                                                   MEM WORKER AVG IN USE  (MB)               0.00
```

# Query Execution Process Flow


Application
Application Interface
Optimizer
Query execution run-time for queries that cannot be or should not be off-loaded to IBM DB2 Analytic Accelerator
IBM DB2 Analytic Accelerator DRDA Requestor
DB2 for z/OS
SMP Host
SPU
CPU
FPGA
Memory
DB2 Analytics Accelerator

→ Heartbeat (DB2 Analytics Accelerator availability and performance indicators)

→ Queries executed without DB2 Analytics Accelerator

→ Queries executed with DB2 Analytics Accelerator

# Query Acceleration Criteria

A query can be routed to IBM DB2 Analytic Accelerator if:

- DB2 in V9 NFM, V10 CM9 or V10  NMF.
- The SQL query can be accelerated, i.e. the unit of acceleration is a whole SQL statement
  - The whole SQL statement will either run in DB2 or in the accelerator
  - All tables referenced to in the SQL must be enabled for acceleration
- The SQL statement is dynamic
- The SQL statement is a SELECT, or subselect part of  INSERT FROM SELECT
- The associated cursor is not defined as a scrollable or a rowset cursor
- The DRDA protocol access is in effect for a remote workload (no private protocol)
- The SQL statement is from a package (not plan DBRM)
- The SQL statement takes long in DB2 (if CURRENT QUERY ACCELEATION = ENABLE, or ENABLE WITH FAILBACK)

# Table Qualification and Column Projection

- A table can be accelerated if it does not have:
  - Security label column
  - A row permission (for V10)
- Some columns are left behind when other columns are copied to IBM DB2 Analytic Accelerator. A column is not copied, if it:
  - Has FIELDPROC
  - Has ASCII encoding
  - Is of type ROWID, LOB or XML, TIMESTAMP(n!=6), BINARY, CHAR FOR BIT DATA
  - Has a column mask (V10)

# Query Routing Heuristics

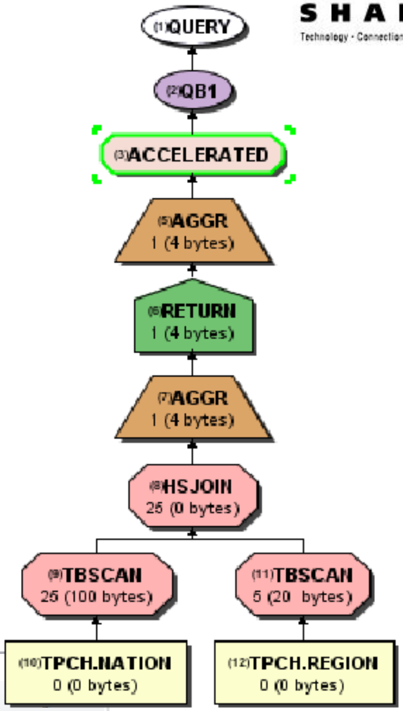
- You get best of both worlds for mixed workloads (OLTP/OLAP)
- DB2 Optimizer uses a set of rules to determine whether a given query is better off being executed in DB2 core engine or routed to the accelerator, such as:
  - typical OLTP access path patterns run in DB2, e,g.
    - *Equal unique or near equal unique access*
    - *One fetch access*
  - typical OLAP/warehousing access patterns run in accelerator
    - *If none of these: WHERE, GROUP BY, ORDER BY, aggregate functions is specified (i.e. all rows are to be returned), the query is not routed*
  - If all the tables referred in the query are “small”, the query is not routed.
    - *ACCEL_TABLE_THRESHOLD is specified by the DB2 Profile_table mechanism. The default value is 1 million rows total*
  - If the estimated cost is lower than the ACCEL_TOTALCOST_THRESHOLD, the query will not be offloaded. The default value is 5,000 .
  - If a huge result set is expected, the query is not routed
    - *ACCEL_RESULTSIZE_THRESHOLD is specified by the DB2 Profile_table mechanism. By default this check is skipped (-1)*
- Recommendation: use the default values. If you have to, change them only after rigorous testing!

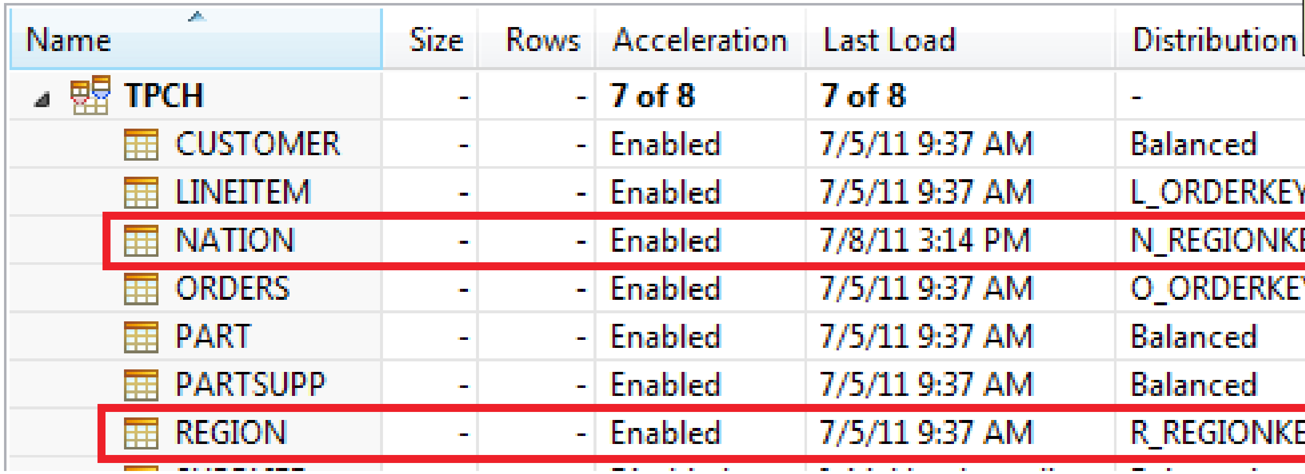
# EXPLAIN

- DB2 EXPLAIN function is enhanced to provide basic information about accelerator usage
  - Whether query qualifies for acceleration and, if not, why
  - The access path details associated with the query execution by Netezza are provided independently of DB2 EXPLAIN by the IBM DB2 Analytic Accelerator Studio.
- For each query (irrespective of the number of query blocks) a row is inserted in the following tables:
  - in both PLAN_TABLE and DSN_QUERYINFO_TABLE, if the query is accelerated
    - PLAN_TABLE's ACCESSTYPE column is set to a value of 'A'
    - DSN_QUERYINFO_TABLE's QI_DATA column shows the converted query text
  - in DSN_QUERYINFO_TABLE only, if the query is not qualified
    - REASON_CODE and QI_DATA columns provide details
- Note that the EXPLAIN tables can be populated with above described information even if there is no accelerator connected to DB2
  - Specifying EXPLAINONLY on START ACCEL command does not establish any communications with an actual accelerator, but enables DB2 to consider its presence in EXPLAIN.

# Netezza EXPLAIN integration

- The Accelerator GUI is able to receive the Netezza Plan files for query executions that happened on the accelerator side. These files are parsed and embedded into DB2 Visual Explain.
- Distribution and Organizing keys can be altered “on the fly” based on the Explain output. The accelerator is redistributing table data in the background.



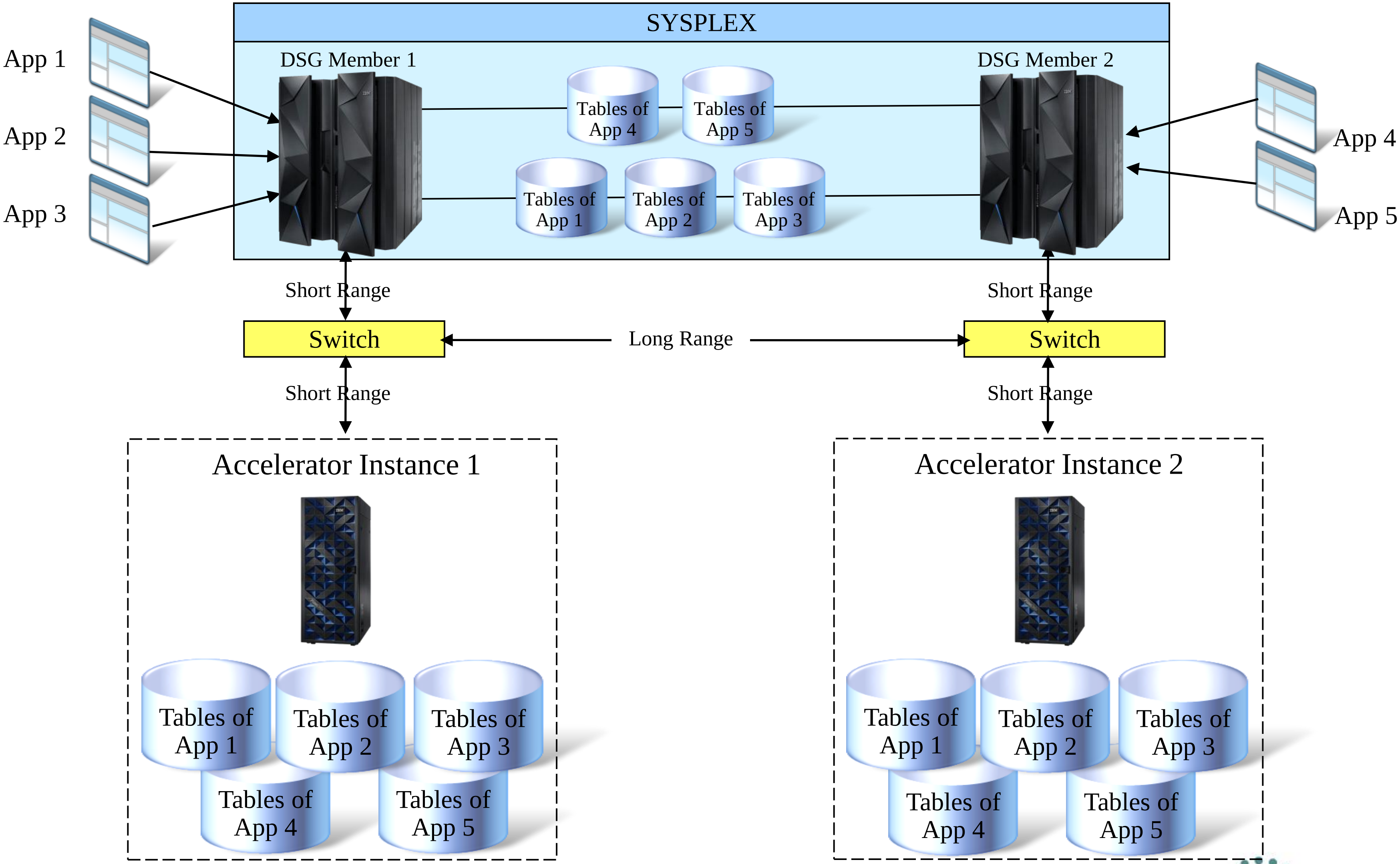


| Name | Size | Rows | Acceleration | Last Load | Distribution |
|---|---|---|---|---|---|
| TPCH | - | - | 7 of 8 | 7 of 8 | - |
| CUSTOMER | - | - | Enabled | 7/5/11 9:37 AM | Balanced |
| LINEITEM | - | - | Enabled | 7/5/11 9:37 AM | L_ORDERKEY |
| NATION | - | - | Enabled | 7/8/11 3:14 PM | N_REGIONKEY |
| ORDERS | - | - | Enabled | 7/5/11 9:37 AM | O_ORDERKEY |
| PART | - | - | Enabled | 7/5/11 9:37 AM | Balanced |
| PARTSUPP | - | - | Enabled | 7/5/11 9:37 AM | Balanced |
| REGION | - | - | Enabled | 7/5/11 9:37 AM | R_REGIONKEY |

# Ways to increase query acceleration chances

- De-correlate a correlated subquery/scalar fullselect into non-correlated.
- Avoid V10 implicit cast feature: numeric_col = ‘123’ => numeric_col = 123
- Avoid multi-row fetch (ROWSET cursor)
- Re-load after DB2 ALTER TABLE on column so the accelerator contains consistent data and types

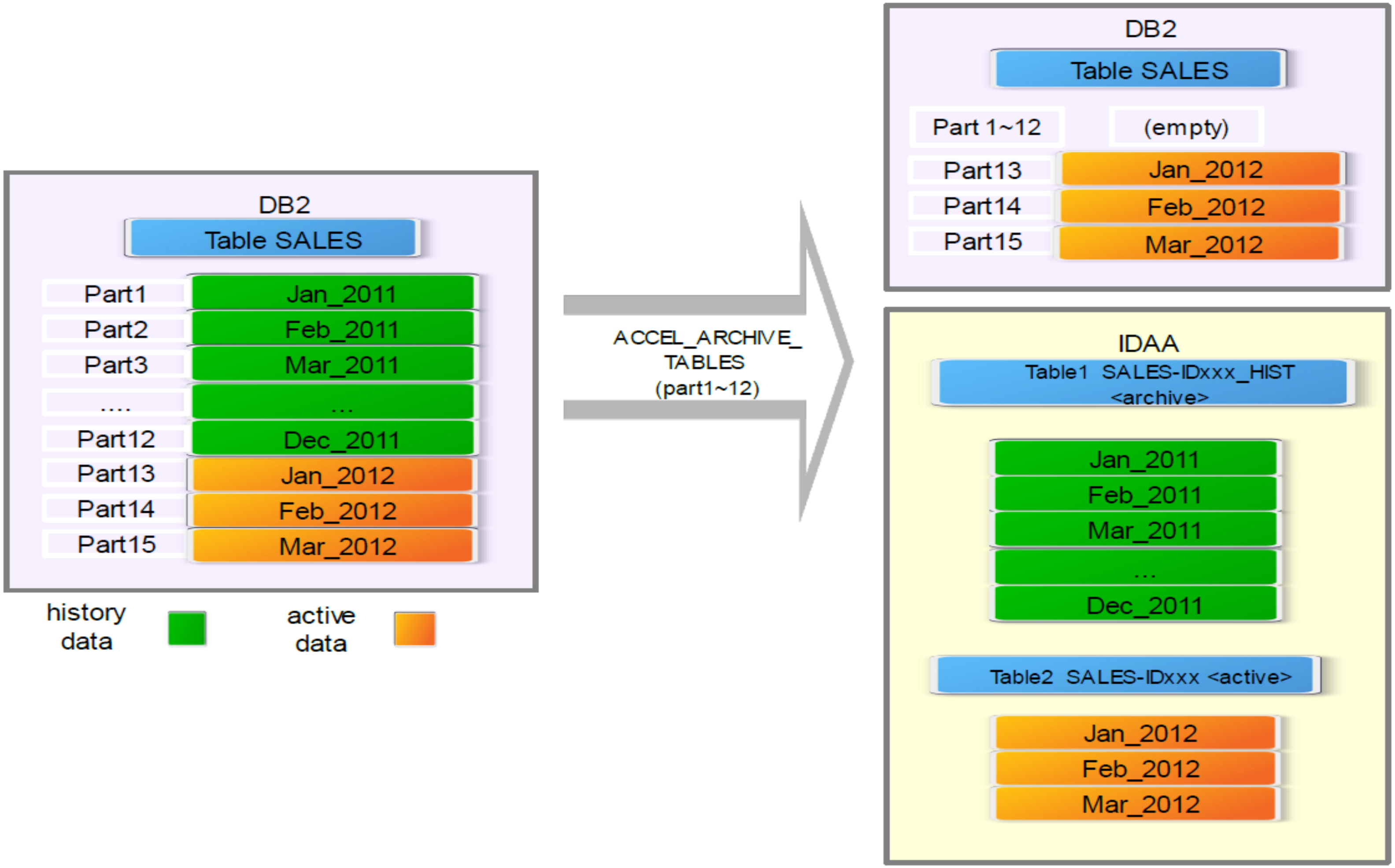
# Disaster Recovery-- Table Loaded in Two Accelerators



# High Performance Storage Saver

- Store historic data on the Accelerator only
- 2 typical ways to keep history of data changes
  - "INSERT-only" approach
    - Any change to data (UPDATE/DELETE) is reflected in database by changing the valid_end timestamp of current row and INSERT of a new row
      - *All data is in a single table which typically grows fast*
  - 2 table approach
    - Any change to data (UPDATE/DELETE) is reflected in database by INSERT of the current row into history table (via trigger or bi-temporal system time in DB2 10) and UPDATE/DELETE of the current row in the active table
- Characteristics of historical data in either approach
  - Data never changes (historical state of current data)
  - Amount of historical data grows quickly

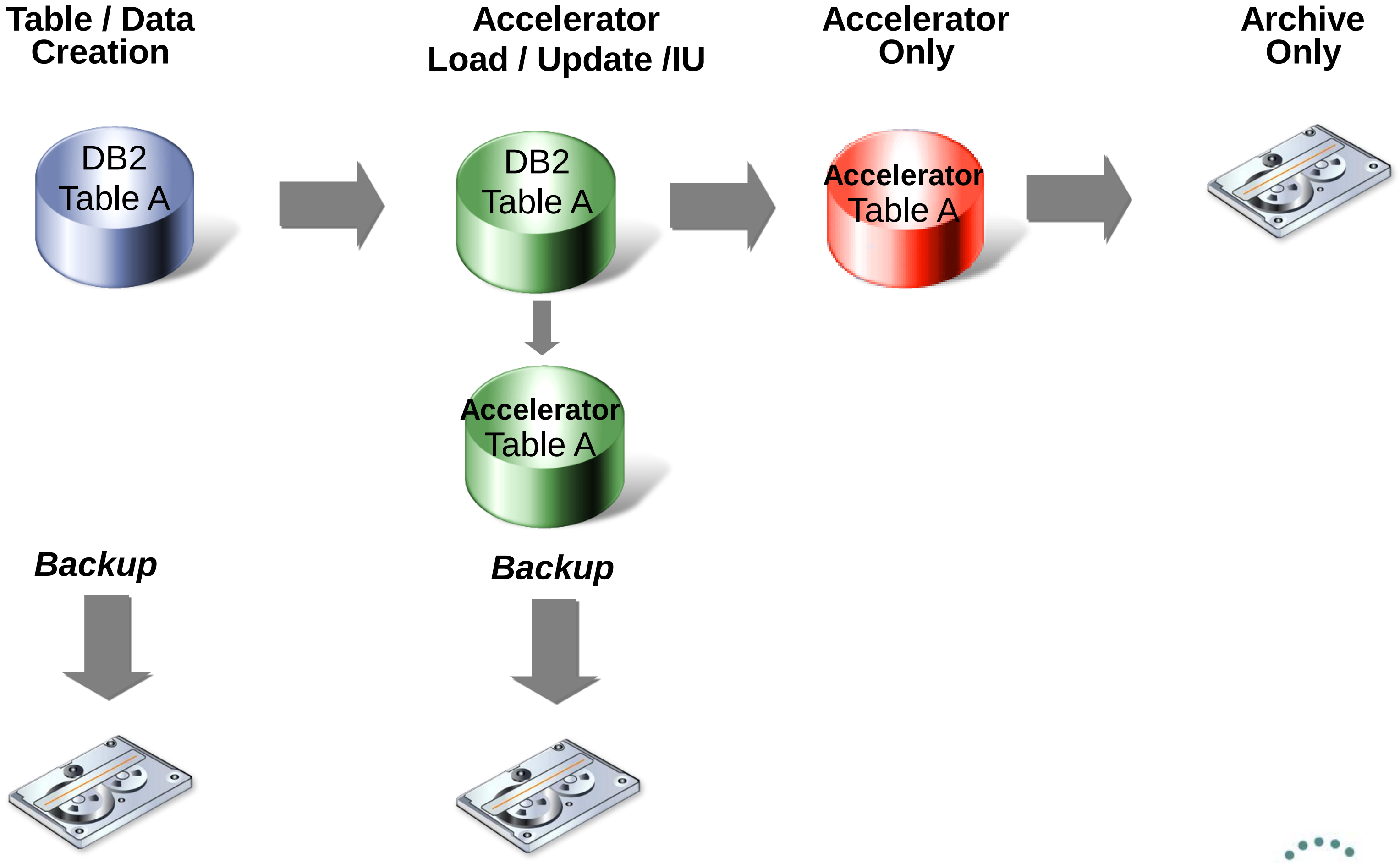
# High Performance Storage Saver



# High Performance Storage Saver

- Data in a partitioned table is divided into two parts:
  - Active data
  - Read-only data
- Save z storage for read-only data by storing them on the accelerator only (purged from DB2 table spaces), query-able through accelerator
- Queries against active data by default, can be accelerated.
- By SET CURRENT GET_ACCEL_ARCHIVE = YES, queries will include the accelerator-archived data, and be executed in accelerators only

# The Evolution of a High Performance Storage Saver



# Workload Self-assessment without an accelerator (1/2)

- **Step1: Setup**
  - Set system parameter ACCEL as AUTO or COMMAND
  - DB2 9 only: set system parm ACCEL_LEVEL set to a value of *V2*
  - Create pseudo catalog table `"SYSACCEL"."SYSACCELERATORS` and `"SYSACCEL"."SYSACCELERATEDTABLES"`
  - Create explain tables

- **Step2: Generate a virtual accelerator**

  -
    ```
    INSERT INTO "SYSACCEL"."SYSACCELERATORS"
    VALUES('SYSACCEL',NULL);
    ```

- **Step3: Populate SYSACCELERATEDTABLES with intended tables**

  -
    ```
    INSERT INTO "SYSACCEL"."SYSACCELERATEDTABLES"
    (NAME,CREATOR,ACCELERATORNAME,REMOTENAME,REMOTECREATOR,
    ENABLE,CREATEDBY,SUPPORTLEVEL)
    SELECT NAME,CREATOR,'SYSACCEL',NAME,CREATOR,'Y',CREATEDBY, 1
    FROM SYSIBM.SYSTABLES
    WHERE TYPE='T' AND DBNAME= <database name>; -- or your own criteria
    ```

# Workload Self-assessment without an accelerator (2/2)

- **Step4: Start the virtual accelerator with EXPLAINONLY**
  - **Issue command:**

```
-START ACCEL(*) ACCESS(EXPLAINONLY)
```

- **Step5: Explain queries**
  - `SET CURRENT QUERY ACCELERATION = ENABLE` or Set system parameter QUERY_ACCELERATION = ENABLE
  - Execute EXPLAIN statement.
  - Check whether a query is eligible for offloading – you can COUNT it

```
SELECT QUERYNO
FROM SYSADM.PLAN_TABLE WHERE  ACCESSTYPE = 'A';
```

  - To see why a query is not eligible for offloading:

```
SELECT QUERYNO, REASON_CODE, QI_DATA
FROM SYSADM.DSN_QUERYINFO_TABLE;
```

- Or explore V10 new functionality `CURRENT EXPLAIN MODE = 'EXPLAIN'` and execute a normal SQL statement.

# Summary

- IBM DB2 Analytic Accelerator opens unprecedented opportunities to analyze massive data in DB2 for z/OS for competitive advantages
- IBM DB2 Analytic Accelerator is simple to use and very effective
- IBM DB2 Analytic Accelerator brings the competitive price/performance to System z for analytics workloads
- We’ve covered main steps in exploiting IBM DB2 Analytic Accelerator:
  - Identify workloads
  - Define and load data, refresh data
  - Enable query acceleration
  - Monitor and tune the performance – like an appliance

# Deep Dive into IBM DB2 Analytics Accelerator Query Acceleration

Maryela Weihrauch and Ruiping Li

IBM Silicon

February 8, 2013

Session 12762