IBM DB2 Analytics Accelerator use cases

Ciro Puglisi
Netezza Europe
+41 79 770 5713
cpug@ch.ibm.com
Traditional systems landscape

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

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
Visionary systems landscape

Benefits
- Consolidating all the components into a single system
- Uniform access to any data
- Efficient data movement within the system (ideally, no network)
- Opportunity to remove, i.e. consolidate some of the layers

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

Approaches
- Columnar stores
- In-memory databases
- Hardware acceleration, special purpose processors
- Appliances

Building on proven technology base
- System z Data Sharing and Parallel Sysplex technology provides all the needed characteristics except one: special purpose processing for analytical workloads to minimize the need for manual tuning
Information Management

Optimization use case
IDAA for existing BI workload on System z

Benefits
- Consolidate reporting where the majority of data being analyzed resides (z/OS)
- Accelerate long running DB2 for z/OS queries from minutes to seconds for greater business value
- Reduce processing cost by freeing up MIPS and storage space for indexing
- Address performance challenges with complex and ad hoc queries
- Avoid costs and efforts to tune the system
- The forgotten query: consider queries previously set aside due to performance challenges?

- Accelerate long running DB2 for z/OS queries from minutes to seconds for greater business value
- Reduce processing cost by freeing up MIPS and storage space for indexing
- Address performance challenges with complex and ad hoc queries
- Avoid costs and efforts to tune the system
- The forgotten query: consider queries previously set aside due to performance challenges?

- Address performance challenges with complex and ad hoc queries
- Avoid costs and efforts to tune the system
- The forgotten query: consider queries previously set aside due to performance challenges?

- Address performance challenges with complex and ad hoc queries
- Avoid costs and efforts to tune the system
- The forgotten query: consider queries previously set aside due to performance challenges?

- Avoid costs and efforts to tune the system
- The forgotten query: consider queries previously set aside due to performance challenges?
Swiss Re

- Acceleration of reports based on mainframe data

- IBM DB2 Analytics Accelerator (Netezza 1000-6) set up within 2 days including HW installation and DB2 pairing
  - Tables loaded and queries executed on day 2

- **Time to value:** first load results
  - Load rate >800GB/h
  - Compression factor 10x
  - 1st query accelerated after 2h

- **Performance:** first query accelerated 51 times. Tested 80 reports with acceleration up to 90 times
  - Quicker responses so users can sharpen their analyses faster

- **Integration:** no change to application interfaces

- **Optimization:** manage workload more efficiently
  - CPU reduction can lead to fast ROI!
Large North American health insurer

Queries running up to 1908 times faster

<table>
<thead>
<tr>
<th>Query</th>
<th>Total Rows Reviewed</th>
<th>Total Rows Returned</th>
<th>DB2 Only Hours</th>
<th>Sec(s)</th>
<th>DB2 with IDAA Hours</th>
<th>Sec(s)</th>
<th>Times Faster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query 1</td>
<td>2,813,571</td>
<td>853,320</td>
<td>2.39</td>
<td>2.640</td>
<td>0.00</td>
<td>0.00</td>
<td>1,908</td>
</tr>
<tr>
<td>Query 2</td>
<td>2,813,571</td>
<td>588,780</td>
<td>1.90</td>
<td>2.224</td>
<td>0.00</td>
<td>0.00</td>
<td>1,320</td>
</tr>
<tr>
<td>Query 3</td>
<td>2,813,571</td>
<td>1,271</td>
<td>0.55</td>
<td>0.580</td>
<td>0.00</td>
<td>0.00</td>
<td>460</td>
</tr>
<tr>
<td>Query 4</td>
<td>2,813,571</td>
<td>601,197</td>
<td>1.08</td>
<td>1.080</td>
<td>0.00</td>
<td>0.00</td>
<td>810</td>
</tr>
<tr>
<td>Query 5</td>
<td>3,422,705</td>
<td>509</td>
<td>0.55</td>
<td>0.580</td>
<td>0.00</td>
<td>0.00</td>
<td>810</td>
</tr>
<tr>
<td>Query 6</td>
<td>4,280,465</td>
<td>185</td>
<td>0.53</td>
<td>0.510</td>
<td>0.00</td>
<td>0.00</td>
<td>810</td>
</tr>
<tr>
<td>Query 7</td>
<td>3,421,271</td>
<td>582,925</td>
<td>0.55</td>
<td>0.580</td>
<td>0.00</td>
<td>0.00</td>
<td>810</td>
</tr>
<tr>
<td>Query 8</td>
<td>3,422,705</td>
<td>724</td>
<td>0.44</td>
<td>0.436</td>
<td>0.00</td>
<td>0.00</td>
<td>810</td>
</tr>
<tr>
<td>Query 9</td>
<td>4,130,107</td>
<td>137</td>
<td>0.43</td>
<td>0.435</td>
<td>0.11</td>
<td>0.11</td>
<td>193</td>
</tr>
</tbody>
</table>
Workload assessment

- **Customer**
  - Collects information from dynamic statement cache, supported by step-by-step instruction and REXX script (small effort for customer)
  - Upload compressed file (up to some MB) to IBM FTP server

- **IBM**
  - Import data into local database
  - Quick analysis based on known DB2 Analytics Accelerator capabilities

**Key contact:**
Data Warehouse System z/Germany/IBM

Report for a first assessment:
- Acceleration potential for
  - Queries
  - Estimated time
  - CP cost

---

IBM lab Database

Documentation and REXX procedure

Data package (mainly unload data sets)

Pre-process and load

IBM lab Database

Quick Workload Test Tool

Report Assessment
Consolidation use case
IDAA to enable Data Analytics consolidation

Utilize virtualization to optimize the use of resources while reducing costs and gaining new agility

A single platform to manage and administer

DB2 Analytics Accelerator

Consolidate the ever growing proliferation of data marts onto a single, easily managed platform

Benefits
- Simplify management of costly/complex Data integration of System z Data with distributed BI system.
- Enable application queries which you would prefer to run with more real-time data on System z
- Consolidate isolated islands of data on one secure Hub
- Eliminate “Multiple Versions of the Truth”
- Increase time to value to deploy new DataMarts
- Integrated OLTP & BI approach to achieve true Operational BI
Banking customer in Southern Europe

- Started Beta program with one test workload
- Results were so convincing that the program was extended to another, much larger workload
- Results were compared to the results of three competitors

BENCHMARK RESULTS - logarithmic scale

Query # SECONDS

Competitor 1
Competitor 2
Competitor 3
DB2 & IDAA
New BI use case
IDAA for new reporting or Operational BI on System z

Gain a complete and timely understanding of all customer and supplier interactions by deploying operational BI on top of your mainframe enterprise information management infrastructure.

Keep new reporting capabilities onto a single, easily managed platform.

Benefits
- Business benefits of analytics for queries previously set aside
- Price/performance of Netezza with mainframe class reliability and security for BI applications
- Fast time to value with transparent integration into existing applications
- Time and cost avoidance in external systems, processes to offload data and integration efforts
- Integrated OLTP & BI approach to achieve true Operational BI
## Synchronization options supporting low latency BI

<table>
<thead>
<tr>
<th>Full table refresh</th>
<th>Table partition refresh</th>
<th>Incremental update</th>
</tr>
</thead>
<tbody>
<tr>
<td>The entire content of a database table is refreshed for accelerator processing.</td>
<td>Optimization for (time-) partitioned warehouse tables, appending changes “at the end”.</td>
<td>Scattered updates after “bulk” load.</td>
</tr>
<tr>
<td>- Existing ETL process replaces entire table.</td>
<td>- More efficient than full table refresh for larger tables.</td>
<td>- Reporting on continuously updated data (e.g., an ODS), considering most recent changes.</td>
</tr>
<tr>
<td>- Multiple sources or complex transformations.</td>
<td>- Reporting based on consistent snapshot (“check point”).</td>
<td>- More efficient for smaller updates than full table refresh.</td>
</tr>
<tr>
<td>- Smaller, un-partitioned tables.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Reporting based on consistent snapshot (“check point”).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Diagrams:
- **Full table refresh diagram**: Shows the process of refreshing the entire content of a database table for accelerator processing.
- **Table partition refresh diagram**: Illustrates the optimization for partitioned warehouse tables by appending changes at the end.
- **Incremental update diagram**: Visualizes scattered updates after bulk load and their impact on reporting.
Summary

- **Today, IDAA rapid customer adoption demonstrates:**
  - **Financial benefits:**
    - Lower hardware, software and maintenance costs by offloading most expensive reports
    - Reduce need to offload and duplicate data from the mainframe
  - **Business benefits:**
    - Improved BI/Analytics Productivity
    - Fast, predictable response times for “right-time” analysis
    - Improved Service Levels on Reporting Environments
  - **IDAA unique advantages:**
    - Fast time to value
    - Maintain a highly secure environment for sensitive data analysis
    - Transparent to the application

- **Tomorrow, IDAA is positioning DB2 to deliver best performance for both transactional and analytical workloads**
Thank you

Ciro Puglisi
Netezza Europe
+41 79 770 5713
cpug@ch.ibm.com