Database Revolution

New Databases for Data Warehouse and BI

Georg Morsing
SAS Institute A/S
Database Revolution

Situation today

- The database market is dominated by relational database management systems (RDBMS)
- Optimized for transaction processing

Problem

- Users need fast access to subsets of information in an enterprise data mountain
- New data is constantly being added, but storage infrastructure cannot keep up
Database Revolution

- Fast search engines revolutionized the internet
- Think of the same revolution when using databases for DW and BI
  - "Google" your database and get results in seconds

Database revolution is needed
1. Remove the transactional overhead
2. Use data partitioning
3. Exploit parallel processing
4. Optimize the database for DW and BI
Make the best use of your budget...

“Most enterprises must manage IT budget growth rates at levels below the expected rate of business growth.

They must do more with less.“

Source: The Gartner
Scenario 2005: IT Leaders' Next Big Decisions
Intelligence Database

Analysts agree that a business view of the data is needed ...

Butler Group

“It is becoming increasingly important to develop a business view of data, which is translated into the type of storage that the data resides on.”

Source: Managing the Storage of Information Report – January 2005
Intelligence Database

Analysts agree that a business view of the data is needed ...

“"The administration overhead of storage is relentlessly escalating. Enterprises must look to do more with less by optimizing the storage environment”

Source: Managing the Storage of Information Report – January 2005
Database Revolution – New Database Category

Intelligence Database
RDBMS versus Intelligence Database

**RDBMS**
- Many updates
- One row at a time
- Low data volume per operation
- Many users
- Query limited data volumes

**Intelligence Database**
- Few batch updates
- Full table scan happens
- High data volume per operation
- Few users
- Query all data volumes

Bloor Research – 2004
Data Warehouse Architecture

Source Systems
- Oracle
- SAP
- DB2

Data Staging

Enterprise Data Warehouse

Data Marts
- OLAP

Reports & Analysis

Metadata
Data Warehouse Architecture

Source Systems | Data Staging | Enterprise Data Warehouse | Data Marts | Reports & Analysis

- Oracle
- SAP
- DB2

But how does it really look at many sites?

Metadata
Data Warehouse Architecture

*How does it really look at many sites?*

1. Policy to use the preferred RDBMS

2. No BI user must access EDW

Terabytes
Data Warehouse Architecture

How does it really look at many sites?

1. Policy to use the preferred RDBMS

2. No BI user must access EDW

Enterprise Data Warehouse

Terabytes

Data mart chaos!
Requirement to the New DW & BI Database

Intelligence Database
- Use open standards
- Access all data sources

- Provide access to detail data
- Provide logging and security

- Minimize disk consumption
- Minimize cost to run and maintain the database

- Scale linear and not exponential
- Load new data fast
- Response time in seconds
New Database for Enterprise Intelligence

- Information
  - From large data volumes
  - Very fast
  - To many concurrent users

- Scalable through
  - Data partitioning
  - Parallel processing
New Database for Enterprise Intelligence

- Remove transaction overhead in RDBMS
- New database technology
  - Smart data organization
  - Use hardware efficiently
New Database for Enterprise Intelligence

- Many concurrent users
- Small task

- Fewer users
- Huge tasks
- Exploit all hardware in a short moment

Designed for the task at hand
New Database for Enterprise Intelligence

Remove transaction overhead in RDBMS
Data partitioning & parallel processing

- Disk usage: Reduce 2-4 times
- Response times: Up to 40 times faster
- Fast batch updates: Up to 100 times faster
SAS Intelligence Database

Database only for Enterprise Intelligence

- Fast
- Robust
  - 24/7, dedicated database
- Scalable
  - Constant response times
  - With growth in data, queries and users
- Low “total cost of ownership”
  - Reduced hardware demand
  - Exploit hardware efficiently
  - Simple administration and maintenance
SAS Intelligence Database

*The technology behind the new database*

1. Data partitioning
2. Parallel processing
3. Advanced hybrid indexing
SAS Intelligence Database

Data partitioning

<table>
<thead>
<tr>
<th>Traditional table</th>
<th>SAS IDB table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptor</td>
<td>Metadata</td>
</tr>
<tr>
<td>Data</td>
<td>Data</td>
</tr>
<tr>
<td>System Table</td>
<td></td>
</tr>
</tbody>
</table>
SAS Intelligence Database

Data partitioning

Operating system

RAM

CPU

CPU

CPU

CPU

I/O controller

I/O controller

SAS IDB table

Metadata

Data
SAS Intelligence Database

Data partitioning

Operating system

RAM

CPU

CPU

CPU

CPU

I/O controller

I/O controller

SAS IDB table

Metadata

Data
SAS Intelligence Database

Data partitioning – Cluster database

ETL processes running in parallel

Data A ➔ Load process ➔ Fact A
Data B ➔ Load process ➔ Fact B
Data C ➔ Load process ➔ Fact C
Data D ➔ Load process ➔ Fact D

SAS IDB tables

Run cluster program
SAS Intelligence Database

Data partitioning – Cluster database

ETL processes running in parallel

Data A → Load process → Fact A
Data B → Load process → Fact B
Data C → Load process → Fact C
Data D → Load process → Fact D

Cluster SAS IDB table

proc spdo library = mylib;
cluster create
   ThirdQuarter2003
   mem = Oct2006
   mem = Nov2006
   mem = Dec2006;
quit;

Run cluster program
SAS Intelligence Database

Data partitioning – Cluster database

- Time-based cluster database

SAS IDB table

Cluster SAS IDB table

Metadata

Data

Cluster metadata

Table1

Table2

Table3

Table4

Table5

Table6

Table7

Table8

Table9

Table10

Table11

Table12
SAS Intelligence Database

Data partitioning – Dynamic cluster database

- Prepare new data
- Run “cluster add”
SAS Intelligence Database

Partitioning of data – Dynamic cluster database

- Very fast to add new data to a running database
SAS Intelligence Database
*The technology behind the new database*

1. Data partitioning
2. Parallel processing
3. Advanced hybrid indexing

**Diagram:**
- **RAM**
- **CPU**
- **Application**
- **I/O controller**
- **Cluster database**

---

Copyright © 2006, SAS Institute Inc. All rights reserved.
SAS Intelligence Database

*The technology behind the new database*

1. Data partitioning
2. Parallel processing
3. Advanced hybrid indexing

Cluster database

![Diagram of SAS Intelligence Database](image)
SAS Intelligence Database
Parallel processing

Exploit the hardware efficiently

- Built-in parallel program logic
  - Where processing
  - Sorting
  - Group by
  - Table join
  - Multi-index builds & updates

- Query optimizer – SQL planner
SAS Intelligence Database

Parallel processing

Partial data files
WHERE, KEEP, SORT, SUMMARIZE, GROUP BY ...

Partial results

Collect results

Finished results to client
SAS Intelligence Database
Parallel processing – Table join

Parallel sorting of partitioned data files

Table A and B sorted

Parallel table join

Threaded Sort on Table A

Threaded Sort on Table B

Output table
SAS Intelligence Database

Parallel processing – Index creation

Partitioned data files

Parallel creation of index

Index ready - also partitioned
## SAS Intelligence Database

**Advanced hybrid index technology**

<table>
<thead>
<tr>
<th>Traditional table</th>
<th>SAS IDB table</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptor</strong></td>
<td><strong>Metadata</strong></td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td><strong>Data</strong></td>
</tr>
<tr>
<td><strong>Index</strong></td>
<td><strong>Index</strong></td>
</tr>
</tbody>
</table>

- Parallel index evaluations
- Threads processes WHERE at the same time
- Index metadata
- Index segments
SAS Intelligence Database
Advanced hybrid index technology

- Parallel index evaluations
- Threads processes WHERE at the same time
- Index metadata and segments

where column_a in ('A', 'B');
SAS Intelligence Database – Role-based applications

- Information User
- Analyst
- Business User

Access to the SAS IDB just like to a RDBMS

- Web server
- Server tier
- SAS Intelligence Database

Security, logging and integration

- Technical Administrator

Create and update the SAS IDB

Data Sources

Data Integration Developer

Metadata

SAS Intelligence Database

Create and update the SAS IDB

Copyright © 2006, SAS Institute Inc. All rights reserved.
SAS Intelligence Database
– Case study from Denmark

- Using DB2 on z/OS
- 941,000,000 rows
- 1% new data every day

Problems

- Batch update expensive
- Batch takes long time, cannot always be completed during the night
- User queries take long time (up to 2 hours)
- Want to expand amount of data with 200%
- DB2 partitions runs full – have to delete data
SAS Intelligence Database
– Case study from Denmark

Old solution

- Move data from the online DB2 system
- FTP data to the BI platform on z/OS
- Run batch jobs on z/OS 20.00 – 8.00
- Give users access to the DB2 data on z/OS 8.00 – 20.00

941,000,000 rows
DB2 online system

New solution

- Move data from the online DB2 system
- FTP data to the BI platform on AIX
- Prepare data on AIX
- Update the (metadata) SAS Intelligence Database
- Users have access to the SAS Intelligence Database 24/7 on AIX
SAS Intelligence Database – Case study from Denmark

Results

- Batch update
  - From X hours to 3 min.
- Response time
  - From 2 hours to 2 min.
- Expand data amount
  - Now 1,700,000,000 rows
- Return Off Investment
  - Save CPU time on z/OS
    - 1:10

Copyright © 2006, SAS Institute Inc. All rights reserved.
Database Revolution – SAS Intelligence Database

For further information, please contact:

- Georg.morsing@sdk.sas.com