Know-How Network: SAP BW Data Load Performance Analysis and Tuning

Ron Silberstein
Platinum Consultant
Netweaver RIG US – Business Intelligence
SAP Labs, LLC
Welcome SAP Developer Network!

http://www.sdn.sap.com
Agenda

- Data Load Overview
- Extraction Performance
- PSA, Transfer Rules, Update Rules
- Loading InfoCubes and ODS Objects
- Parallel Data Load
- Aggregate Maintenance
Agenda

- Data Load Overview
- Extraction Performance
- PSA, Transfer Rules, Update Rules
- Loading InfoCubes and ODS Objects
- Parallel Data Load
- Aggregate Maintenance
Goals of performance optimization:
First tune the individual single execution and then the whole load processes.
- Eliminating unnecessary processes
- Reducing data volume to be processed
- Deploying parallelism on all available levels

Parallel processes are fully scalable

The typical BW data load process:
SAP Service API: Extraction/Load Mechanism

- InfoCube
- Communication Structure
- Transfer Rules
- Update rules
- Master Data
- Attributes
- Texts
- ODS
- BW S-API
- Extraction Source Structure
- Transaction Data
- DataSource
- Header
- Item
- Master Data
- ATTR
- TXT
- BW 3.0

© 2003 SAP Labs, LLC, Know-How Netowke, Ron Silberstein6

THE BEST-RUN BUSINESSES RUN SAP
How to identify high Extraction Time?

Determine the extraction time:

- Data load Monitor: Transaction RSMO
- Source System: Business Information Warehouse
- PSALoad
- Process: Data load
- Extractor: InfoCube
- ODS
- ALE
- IDOC
- Update rules
- Transfer rules
- Scheduler
- S-API

Extraction Process:

- Scheduler
- Business Information Warehouse
- Source System
- Extractor
- InfoCube
- ODS
- IDOC
- S-API
### Monitoring Extraction: Resource Utilization

#### Look for Many Long-Running Processes

Further Analysis in case of Resource problems when extracting data...

Check SM51 / SM50 in the source system

<table>
<thead>
<tr>
<th>No</th>
<th>Ty</th>
<th>PID</th>
<th>Status</th>
<th>Reason</th>
<th>Start</th>
<th>Err</th>
<th>Sem CPU</th>
<th>Time</th>
<th>Report</th>
<th>Cl User</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>DIA</td>
<td>020</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>DIA</td>
<td>027</td>
<td>Running</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DIA</td>
<td>1015</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DIA</td>
<td>1548</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>DIA</td>
<td>869</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DIA</td>
<td>635</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DIA</td>
<td>483</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>DIA</td>
<td>934</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DIA</td>
<td>1278</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>DIA</td>
<td>721</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>DIA</td>
<td>590</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>DIA</td>
<td>943</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>DIA</td>
<td>883</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>DIA</td>
<td>1209</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>LFO</td>
<td>573</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>LFO</td>
<td>752</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>ENQ</td>
<td>920</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>EGO</td>
<td>496</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>ECO</td>
<td>1411</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>EGO</td>
<td>917</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>SP0</td>
<td>1975</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>LFO</td>
<td>971</td>
<td>waiting</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Extraction Time is too high?

Further Analysis in case of PERFORMANCE problems extracting data...

Transaction: RSA3

For specific application areas specific notes exist; refer to relevant SAP notes and apply when necessary.
Extraction Time is too high?

Analyze high ABAP Runtime:

Particularly Useful for User Exits

Use SE30 Trace option “in parallel session”. Select corresponding Work process with extraction job.
Extraction Time is too high?

Identify expensive SQL Statements

Use ST05 Trace with Filter on the extraction user (e.g. ALEREMOTE). Make sure that no concurrent extracting jobs run at the same time with this execution.
Extraction Tuning: Load Balancing

Parallel processes:
- distribute to different servers
  - avoid bottlenecks on one server
  - Config in table ROIDOCPRMS

RFC destinations (SM59)

- Example: RFC connection from BW to R/3 and R/3 to BW
- InfoPackages, event chains and Process Chains: all can be processed on specified server groups.
- XML Data loads: HTTP/HTTPS processes can be allocated to specific server groups

Expected Results:
- Avoid CPU/Memory bottlenecks on one server
- Greater Throughput: Faster time to completion per request
Extraction Tuning: Configuring DataPackage Size

Size of DataPackages: Influencing Factors
- Specific to application datasource, the contents and structure of records in the extracted datasets.
- Package size: impacts frequency of COMMITs in DB.
- SAP OSS note 417307: Extractor Packet Size Collective Note for SAP Applications
- Consider both the source system and the BW system (RSADMINC).
- Package size specified in table ROIDOCPRMS and/or InfoPackages

Scenario:
- Set up the parameters according to the recommendations; if upload performance is not improved, try to find other values that fit exactly your requirements.

Expected results:
- In a resource constrained systems, reduce DataPackage size
- In larger systems, increasing the package size to speed collection;
  - but take care not to impact communication process and unnecessarily hold work processes in SAP source system.
- Greater throughput = Faster time to completion per request

© 2003 SAP Labs, LLC, Know-How Netowke, Ron Silberstein14
Further Analysis in case of Resource problems when extracting data...

OSS note 409641 for details

Double Check ROIDOCPRMS Settings

If no entry was maintained, the data is transferred with a standard setting of 10000 kbyte and 1 “Info IDOC” for each data packet.
Further Analysis in case of Resource problems when extracting data...

Consider Decreased Data Package Size for a specific InfoPackage
Use Selection Criteria

Consider building indices on DataSource Tables based on selection criteria.
Extractor Tuning: SAP and Generic DataSources

SAP Content extraction
- Convert old LIS extractors to Logistics Extraction Cockpit
- V3 Collection jobs for different DataSources can be executed in parallel
- Tune customer exit coding

Generic extractors:
- Collector jobs can be executed in parallel
- InfoPackages executed in parallel to extract data
  - Not possible for delta extracts from one generic data source
- Investigate Secondary indexes on fields used for selection
- Optimize custom ‘collector’ ABAP coding
Extraction / Load Tuning: Flat Files

- Use a **predefined record length** (ASCII file)
- File should **reside on the application server** not on the client PC
- Avoid large loads **across a network**.
- Avoid reading load files **from tape** (copy to disk first)
- Avoid placing input load files **on high I/O disks**
  Example: same disk drives or controllers as the DB tables being loaded.
Agenda

Data Load Overview

Extraction Performance

PSA, Transfer Rules, Update Rules

Loading InfoCubes and ODS Objects

Parallel Data Load

Aggregate Maintenance
Analyze high PSA Upload Times

Data load Monitor: Transaction RSMO
 PSA Partitioning

Data Load → PSA Table Partitions

**BW: Threshold Value for Data Load**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Status-IDOC</th>
<th>Packet size</th>
<th>Partition size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>20000</td>
<td>1,000,000</td>
</tr>
</tbody>
</table>

Size of each PSA partition, value in number of records

**Transaction SPRO or RSCUSTV6:**

- From SPRO Business Information Warehouse > Links to Other Systems > Maintain Control Parameters for the Data Transfer

  - Note: If you start more than one load process at a time expecting to have each request in a separate partition, it probably will not work as expected; the PSA threshold is not yet reached when the second process starts writing into PSA
Data Processing - Transfer Rules

Data load Monitor: Transaction RSMO

Transfer Rules
Data Processing - Update Rules

- Source System
- Business Information Warehouse
- Extractor
- S-API
- ALE
- IDOC
- Update rules
- Transfer rules
- PSA
- tRFC
- Data load Monitor: Transaction RSMO

**Update Rules**

Select a node and choose the context menu for more information.
Debugging and tuning Update and Transfer rules:
- Simple tool for debugging of transfer or update rules
- Improves error search and analysis – together with the enhanced error messages
Routines: Potential Performance Bottlenecks

Identify the expensive update/transfer rules:

Debug one update/transfer rule to the next update rule for each InfoObject.

Also use ST05 or SM30

Recommendations:

- SINGLE SELECTs are one of the performance “killers” within these codings; use buffers (such as internal tables) and array operations instead.

- Avoid too many library transformations, as they are interpreted at runtime (not compiled like routines)
  - The transformation engine or library is new in BW 3.0
Agenda

- Data Load Overview
- Extraction Performance
- PSA, Transfer Rules, Update Rules
- Loading InfoCubes and ODS Objects
- Parallel Data Load
- Aggregate Maintenance
Data Load: Data Targets

Data Load Monitor - RSMO

Load Into Data Targets
Initial and Large Data Load Volume Tuning

Buffering Number Range (InfoCube):

Activate the number range buffer for the dimension ID’s

Reduces application server access to Database.
- i.e. set the number range buffer for one dimension to 500, the system will keep 500 sequential numbers in memory
- SAP OSS note 130253: Notes on upload of transaction data into BW

Scenario:
- High volumes of transaction data: significant DB access (NRIV table) to fulfill number range requests.

Expected Results:
- Accelerates data load performance per load request.

Note:
- After the load, reset the number ranges buffer to its original state: minimize unnecessary memory allocation.
Transactional Data Load Performance Tuning

Load Master data **before** transaction data
- Creates all SIDs and populates the master data tables (attributes and/or texts).
- SAP OSS note 130253: Notes on upload of transaction data into BW

**Scenario:**
- Always load master data before transaction data (ODS and InfoCube).
- When completely replacing existing data, delete before load!

**Expected Results:**
- Accelerates transaction data load performance: all master data SIDs are created prior to transaction load, and need not be determined during transactional data load (large overhead).
“Snapshot” Reporting: Data Deletion

- Some reporting scenarios require no historical data

Scenario:

- When completely replacing existing data, delete before load!

Expected Results:

- Data deleted from PSA can reduce PSA read times
- Data deleted from InfoCube reduces deletion and compression time.
  - “Drop partition...“ DDL statement instead of “delete from table...“ DML statement only takes seconds
- Deleting Data also speeds data availability (aggregates, etc)
ODS Activation in BW 2.x

Data Packets / Requests can not be loaded into an ODS object in parallel

- overwriting functionality
- Locking on Activation table

Sequential load

Staging Engine

active data

change log

New/modified data

Activation

Req1

Req2

Req3

Req3, 1, 2
ODS Activation in BW 3.x

New queuing mechanism replacing previous Maintenance (M)table

- **active data**
- **change log**
- **Activation queue**

- **Doc-No.**
- **Req.1**
- **Req.2**
- **Req.3**
- **Parallel load**

**Staging Engine**
Upload to Activation queue
- Data from different requests are uploaded in parallel to the activation queue.

Activation
- During activation the data is sorted by the logical key of active data plus change log key.
- This guarantees the correct sequence of the records and allows inserts instead of table locks.

Before- and After Image
- Request ID in activation queue and change log differ from each other.
- After update, the data in the activation queue is deleted.
Control of ODS Data load Packaging and Activation

Transaction RSCUSTA2

Max. no. of parallel Dialog work processes
Min. no. of recs per package
Max. Wait Time in secs. for ODS Activation
Server Group for RFC Call when Activating Data in ODS

Controls data packet size utilized during parallel update/activation and number and allocation of work processes.
Non-Reporting ODS Objects:

- **Data Archiving ODS**: ARCHODS1
- **Status Info**
  - Version
  - Save
  - Status
- **Settings**
  - **BEx Reporting**: Standard
  - **ODS Object Type**
  - **Unique Data Records**
  - **Check table for InfoObject**
  - **Set quality status to 'OK' automatically**
  - **Activate ODS object data automatically**
  - **Update data targets from ODS**

**BEx Flag**: Computation of SIDs for the ODS can be switched off.

**BEx-flag must be switched on if BEx-reporting on the ODS is executed**

---

Loads are faster as Master Data SID tables do not have to be read and linked to the ODS data.
Further ODS Loading/Activation enhancements (3.x)

Update of ODS object with unique records
- Significantly simplifies activation process
- No lookup of existing key values
- No updates in active table, only inserts
- Note: Data owner is responsible for uniqueness!

Index maintenance
- Indexes speed querying
- Slow down activation

Parallel SID creation
- SIDs are created per package
- Multiple packages are handled in parallel by separate dialog processes
Admin WB > Modeling > InfoCube Manage > Performance Tab

Recommendation: Drop secondary Indexes for large InfoCube data loads

Create Index button: set automatic index drop / rebuild

Statistics Structure button: set automatic DB statistics run after a data load
Agenda

- Data Load Overview
- Extraction Performance
- PSA, Transfer Rules, Update Rules
- Loading InfoCubes and ODS Objects
- Parallel Data Load
- Aggregate Maintenance
Parallel Data Load

System-controlled parallelism

- Parallel upload by packaging the source data
  - Packets are created during the extraction and sent simultaneously to BW
  - Packet size for flat files definable in table RSADMINC (IDOCPACKSIZE)
  - Packet size for mySAP source systems definable in table ROIDOCPRMS (MAXLINES)

Administrator controlled parallel data load

- InfoPackages
  - Loading from the same or different data source(s) with different selection criteria simultaneously
  - Enables parallel PSA → Data Target process as PSA partitions can be posted in parallel
- Initial fill of aggregates
  - Use aggregate hierarchy to schedule parallel filling jobs
During extraction, packets are generated and sent simultaneously.

Packets are updated to data targets in parallel.
Multiple processes can be started in parallel manually, for example:

- InfoPackages (loading from multiple data sources or from the same data source with different selection criteria simultaneously)
- Initial filling of multiple aggregates
... and their influence on parallelization and usage of dialog or batch process for loading from PSA to data targets
New in BW 3.X

Process Chains
- Replacement for Event chains

Transaction RSPC

Process type:
- Delete Index
- Generate Index
- Auto suggestion depending on InfoPackage selected.

Process Chain TechEd session:

BW 210 BW Automation with Process Chains
Process Chains Parallel Data Load

Start
- Load Sales Order Data from Various Sources

Load Data
- Delta XML: Sales order data
- ZPAK_2HKH2UQY02H77175LV0NRUIB
- Flat File: Sales Order Details

AND
- Load OK: Trigger CDS Activation
- Load Failed completely

ODSO Data
- Activate Data in ODS object 'Sales Order Details'

Further Processing
- Update InfoCube from ODS Object 'Sales Order Details'
Agenda

- Data Load Overview
- Extraction Performance
- PSA, Transfer Rules, Update Rules
- Loading InfoCubes and ODS Objects
- Parallel Data Load

Aggregate Maintenance
Roll-up

- The roll-up process populates all aggregates of an InfoCube with the newly loaded delta load
- Basic Rule: InfoCube and all depending aggregates must be in-sync, i.e. you must see the same data no matter if it has been derived from the InfoCube or the aggregate
- Newly loaded data is not available for reporting until it has been rolled up into the aggregates

Aggregate Hierarchy

- Aggregates can be built out of other aggregates to reduce the amount of data to be read and, hence, to improve the roll-up performance
- Aggregate hierarchy is determined automatically
- General Guideline: define few basis (large) aggregates and many small aggregates that can be built from the hierarchy level before
Example: Optimized Aggregate Hierarchy

Basic InfoCube

Few large basis aggregates

Many small aggregates

Example

{Material, Material Group, Customer, Day}

{Material Group, Customer, Day}

{Material Group, Month}
Change Run

Aggregates can contain
- Dimension Characteristics
- Navigational Attributes
- Hierarchy Levels

When master data changes, the changes of the navigational attributes/hierarchies must be applied to the depending aggregates; this process is called change run.

Newly loaded master data is not active until the change run has been applied the changes to all aggregates.

Threshold for delta and new build-up in customizing.

The Change Run can be parallelized across InfoCubes; see SAP note 534630 for more details.

Check aggregate hierarchy (see Roll-Up for more details).

Try to build basis aggregates that are not affected by the change run, i.e. no navigational attributes nor hierarchy levels.

The following slides show details on the process itself...