Performance Management in Data Loading: an Overview

Applies to:
SAP BI 7.0 / SAP Net weaver 2004s (Support package 14 - 17). For more information, visit the EDW homepage.

Summary
This document talks about various checks for performance improvement in data extraction and data staging for timely information retrieval.

Author: Saurabh Choudhary
Company: Infosys Technologies Limited
Created on: 22nd June 2010

Author Bio
Author is currently working as Consultant in Infosys Technologies Limited.
# Table of Contents

Performance management in Data loading .................................................................................................................. 3

Data Extraction ............................................................................................................................................................ 4
- Custom code extraction performance ......................................................................................................................... 4
- Data transfer parameters ................................................................................................................................................. 4
- Application specific setup ................................................................................................................................................ 4
- Extractor performance ................................................................................................................................................... 4
- Analyze database performance ..................................................................................................................................... 4

Data Staging .................................................................................................................................................................... 5
- Data load parameters ..................................................................................................................................................... 5
- Preference to tRFC option instead of IDOCs .................................................................................................................. 5
- Need of PSA .................................................................................................................................................................... 5
- Preference to ASCII files instead of CSV ...................................................................................................................... 5
- Custom transformation rules performance .................................................................................................................. 5
- Availability of master data prior to transaction data ................................................................................................... 5
- Number range buffering check ....................................................................................................................................... 5
- Parallelize multiple data loads ......................................................................................................................................... 6
- Deletion of indexes prior to data load & re-creation .................................................................................................... 6
- Database statistics & performance ................................................................................................................................ 6

Related Content ............................................................................................................................................................. 7

Disclaimer and Liability Notice ........................................................................................................................................ 8
Performance management in Data loading

To provide information in time, Performance of data loads and data maintenance plays a vital role.

BI Administrator Cockpit is used for performance analysis in data loading and data management. Transaction code RSMO can help to retrieve detailed information.

Five major steps are taken in processing a request.

1. Request a Data Package
2. Extract data from the source
3. Transformation
4. Surrogate Key Generation
5. Subsequent steps

As per the analysis among above steps “Transformation” takes the maximum time followed by “Surrogate key generation”. The runtime depends on the complexity of the transformations and the information model.

There are two different staging processes available, the classic info Source staging process with transfer and update rules and the new Data Source staging process with transformations. In the classic staging process transfer, rules will typically take some 1/3rd of the time and update rules take remaining time.

Careless use of user exits, transformations, transfer rules, or update rules causing most of the performance problems in extraction and staging process. From performance point of view pricey database operations and calculations should either be avoided or carefully reviewed and optimized (for example, by replacing single record database reads by multiple record database reads in the start or end routine).

While much of the data transformation performance is under control of the user, consequent data maintenance steps such as Data Store Object data activation attribute change runs, or aggregated rollups are predefined by system and can only be influenced only by changing the information model or optimizing database performance.
Data Extraction

Data extraction performance largely depends on the data model and data volume of the source system application. Still a few things can be done to improve overall data extraction performance.

Custom code extraction performance

- We write custom code in user exits or in the implementation of generic function module extractors. We should ensure that those are implemented in the most efficient way.
- We may use transaction SE30 to further analyze the run time delivered or custom extraction programs.

Data transfer parameters

- Various parameters including the size of data packages, the frequency of sending status information, the maximum number of parallel processes to be used, can be defined in transaction SBIW.
- We may increase the number of processes and the size of the individual data packages and reduce the frequency of status IDOCs to improve the performance of the extraction process depending on the availability of system resources.
- Specific recommendations for parameter settings are available at the SAP Service Marketplace but there is limited potential for improving extraction performance by changing these parameters.

Application specific setup

- There are various application specific options and settings in transaction SBIW that influence data extraction performance for the corresponding applications.

Extractor performance

- In Transaction RSA3 detailed analysis of actual extractor performance can be done using debug mode.
- RSA3 transaction allows testing of extractor without having an SAP BW system requesting data.

Analyze database performance

- In many situation up to date database statistics and index addition to extracted database tables help to improve extraction performance.
Data Staging

Flaws in the information model or in custom transformations cause bad performance in data staging process more commonly. An organized investigation helps to get root cause of the performance issues occurred.

Data load parameters

- We should set data load and other parameter using transaction SPRO (customizing) according to SAP recommendations.
- We should remember that defaults settings can be overridden by settings made in the Info packages or data transfer process for data package size and other parameters.
- These parameters setting changes at info package or data transfer process level may help in various situations. For example, help to increase the overall data package size where certain memory-intensive extractors or data staging processes need to run with lower data package size.

Preference to tRFC option instead of IDOCs

- It is not normally used anymore now a day. We should be sure to toggle the data transfer mode to tRFC where it is set to IDOCs. It will help to reduce the overhead implied in IDOC processing.

Need of PSA

- We need to be sure if PSA is required as an intermediate storage (for example, for error handling purposes) or not, and if not then we should turn it off.
- Particularly in an Enterprise Data Warehousing setup with a separate data staging layer, PSA is not needed.

Preference to ASCII files instead of CSV

- We should consider changing file format from CSV to plain ASCII files with fixed field lengths for better performance.

Custom transformation rules performance

- Transformations are executed for every single record. Inefficient and complex implemented custom transformations cause performance issues. We should try to put custom code as part of start or end routine wherever possible.
- We should follow common ABAP performance guidelines when implementing the custom code.
- We may use transaction SE30 to perform a runtime analysis of update process.

Availability of master data prior to transaction data

- We should load master data prior to transaction data. It helps to reduce the need of new surrogate keys creation for master data and improves data-load performance considerably.

Number range buffering check

- Transaction NRIV can be used to switch on number range buffering.
- Competent number range buffering reduces the time required to create new dimension keys and SIDs.
Parallelize multiple data loads

- We should try to maximize use of existing memory resources and CPU.
- We may scale Data staging processes with number and speed of CPUs in SAP BW scale in very well manner.
- Introducing parallel loads do not speed up individual data staging processes but it helps in reducing overall time required for data load.

Deletion of indexes prior to data load & re-creation

- We should try to delete indexes before the data load and recreate indexes after all data loads have been completed.
- This helps to prevent the incident of degenerated indexes as well which causes degradation in reporting performance.

Database statistics & performance

- Inaccurate and incomplete master data indexes and statistics can cause striking performance drops. It is must to get master and transaction data as part of the update process read by SAP BW.
- We should add indexes to extracted database tables and should keep the database statistics up to date. It helps to enhance extraction performance in various instances.
Related Content

http://help.sap.com

Mastering the SAP Business Information Warehouse (Second Edition)

For more information, visit the EDW homepage.
Disclaimer and Liability Notice

This document may discuss sample coding or other information that does not include SAP official interfaces and therefore is not supported by SAP. Changes made based on this information are not supported and can be overwritten during an upgrade.

SAP will not be held liable for any damages caused by using or misusing the information, code or methods suggested in this document, and anyone using these methods does so at his/her own risk.

SAP offers no guarantees and assumes no responsibility or liability of any type with respect to the content of this technical article or code sample, including any liability resulting from incompatibility between the content within this document and the materials and services offered by SAP. You agree that you will not hold, or seek to hold, SAP responsible or liable with respect to the content of this document.