

Paper 292-2008

Utilizing SAS® SPD Server Dynamic Cluster to Manage Very Large Data

Dan Sargent, SAS Institute Inc.

ABSTRACT

Getting data into and out of a data warehouse environment by extracting, transforming, and loading tables is important. It is just one consideration for managing a large warehouse. This paper provides an overview to using the SAS® SPD Server Dynamic Cluster feature that makes the ETL process possible with large data. Dynamic Cluster tables enable the partitioning of data based on criteria in the data allowing parallel loading of tables and the removal of data from a table during updates and refreshes easy. This feature makes managing a multibillion-row table possible and efficient. However it is very important to understand the types of queries that tables of a warehouse will satisfy. The next step in managing data is to optimize the tables to enhance the performance of queries that join, merge, subset, and summarize tables in a data warehouse. By understanding the query patterns, an administrator can optimize the table in a warehouse for both loading query performance. The paper will show system resource consumption such as wall time and CPU utilization for query types when different organization of the rows within tables are applied.

No paper was submitted for publication in the *Proceedings*. Check <http://support.sas.com/rnd/papers/> or contact the author.

CONTACT INFORMATION

Dan Sargent
SAS Institute Inc.
dan.sargent@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.