

Informatica PowerCenter Partitioning Option

Delivering High Performance for Processing Massive Data Volumes

Avram Kornberg

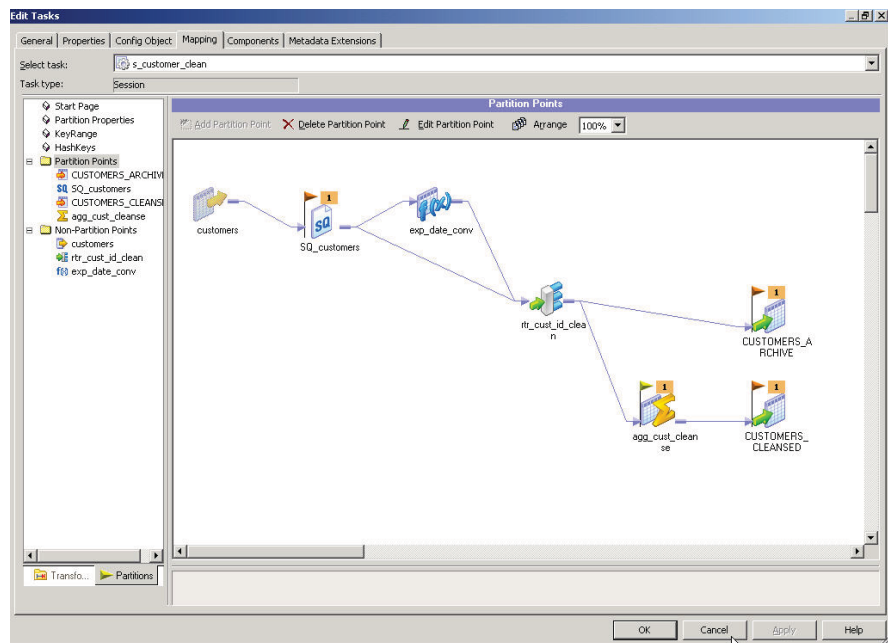
Senior Vice President and CIO
Natexis Banques Populaires USA

“PowerCenter helps us achieve greater productivity, drive down costs, minimize operational risks, and accelerate the time to market for new data integration-intensive enterprise applications.”

BENEFITS

- Scale cost-effectively to handle large data volumes
- Enhance developer productivity
- Optimize system performance in response to changing business requirements

The **PowerCenter® Partitioning Option** increases the performance of PowerCenter through parallel data processing, and it has been instrumental in establishing PowerCenter’s industry performance leadership. This option provides a thread-based architecture and automatic data partitioning that optimizes parallel processing on multiprocessor and grid-based hardware environments.



The Partitioning Option dynamically optimizes jobs for parallel processing at run time based on data-driven, key-driven, or database-supplied partitioning schemes, dramatically increasing PowerCenter’s performance.

Partitioning Option Key Features

Data Smart Parallelism

- Automatically aligns PowerCenter partitions with database table partitions to improve performance
- Automatically guarantees data integrity by leveraging the parallel engine of PowerCenter, which dynamically realigns data partitions for set-oriented transformations

Session Design Tools

- Create user-defined partitioning schemes quickly and easily
- Provide a graphical partitioning map for determining the best partitioning points
- Gather statistics on configurable session options, such as error handling, recovery strategy, memory allocation, and logging, to maximize performance

Integrated Monitoring Console

- Gathers session statistics, such as throughput, rows/second, error details, and performance optimizations, to identify potential bottlenecks and recognize trends
- Shows all session execution and dependency details

Multiple Partition Schemes

- Support parallelization through multiple mechanisms, including key range, hash algorithm-based, round robin, or file partitions
- Maximize data throughput via concurrent processing of specified partitions along the data transformation pipeline

Partitioning Option Benefits

Scale Cost-Effectively to Handle Large Data Volumes

With the Partitioning Option, you can execute optimal parallel sessions by dividing data processing into subsets that are run in parallel and spread among available CPUs in a multiprocessor system. When different processors share the computational load, large data volumes can be processed faster. When sourcing and targeting relational databases, the Partitioning Option enables PowerCenter to automatically align its partitions with database table partitions to improve performance. Unlike approaches that require manual data partitioning, data integrity is automatically guaranteed because the parallel engine of PowerCenter dynamically realigns data partitions for set-oriented transformations (e.g., aggregators or sorters).

Enhance Developer Productivity

The Partitioning Option provides intuitive, GUI-based, session design tools that reduce the time spent on initial and ongoing configuration and performance tuning tasks. You can easily create user-defined partitioning schemes. A graphical partitioning map helps you determine the best points of partitioning. Configurable session options, such as error handling, recovery strategy, memory allocation, and logging, make it easier to gather statistics used to maximize performance.

Optimize System Performance in Response to Changing Business Requirements

The Partitioning Option lets you easily gather in-depth session statistics such as throughput, rows/second, error details, and performance optimizations. These statistics help you identify potential bottlenecks and recognize trends. An integrated monitoring console lets you view all session execution and dependency details. With the metadata-driven architecture of PowerCenter, data transformation logic is abstracted from the physical execution plan. This feature enables rapid performance tuning without compromising the logic and design of the original data mappings. You can continually and easily optimize system performance in the face of increasing data loads and changing business requirements.

Learn More

Let Informatica help you determine how the PowerCenter Partitioning Option can maximize the business value of your data.

Visit us at www.informatica.com/powercenter or call 1.800.653.3871.

About Informatica

Informatica is a leading provider of enterprise data integration software and services. With Informatica, organizations can gain greater business value by integrating all their information assets from across the enterprise. Thousands of companies worldwide rely on Informatica to reduce the cost and expedite the time to address data integration needs of any complexity and scale.



Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 www.informatica.com

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Ireland · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

© 2008 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and PowerCenter are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

6683 (09/11/2008)