Informatica PowerCenter Partitioning Option

Process Massive Data Volumes with High Performance

The Informatica® PowerCenter® Partitioning Option helps your IT organization take advantage of parallel data processing in multiprocessor and grid-based hardware environments. This option increases PowerCenter’s performance with a thread-based architecture and automatic data partitioning capabilities. The option executes optimal parallel sessions by dividing data processing into subsets, which run in parallel and are spread among available CPUs in a multiprocessor system.

Unlike approaches that require manual data partitioning, the Partitioning Option automatically guarantees data integrity because PowerCenter’s parallel engine dynamically realigns data partitions for set-oriented transformations. Configurable session options, such as error handling, recovery strategy, memory allocation, and logging, make it easier to gather statistics used to maximize performance.

By enabling hardware and software to jointly scale to handle large volumes of data and users, the Partitioning Option helps your IT organization maximize its technology investments, boost productivity, and optimize system performance.

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

Data Smart Parallelism
This feature improves system performance by automatically aligning PowerCenter partitions with database table partitions.

Guaranteed Data Integrity
Leveraging PowerCenter’s parallel engine to dynamically realign data partitions for set-oriented transformations, such as aggregators or sorters, this option guarantees data integrity.

Session Design Tools
Intuitive, GUI-based session design tools are available for creating partitioning schemes, determining the best partitioning points, and gathering statistics on error handling, recovery strategy, memory allocation, and logging.

Integrated Monitoring Console
All session execution and dependency details can be viewed from an integrated monitoring console, where your IT team can gather session statistics—such as throughput, rows/second, error details, and performance—to identify potential bottlenecks and recognize trends.

Multiple Partition Schemes
This option provides multiple partition schemes to support parallelization through key range, hash algorithm-based, round robin, or file partitions and through concurrent processing of partitions along the data transformation pipeline to maximize data throughput.

Key Benefits

Maximize Return on IT Investments
With this option, your IT team can divide data processing into subsets that are run in parallel and spread among available CPUs in a multiprocessor system. Large data volumes can be processed faster. By scaling hardware and software to handle large volumes of data and users, this option maximizes return on your IT investments.

Boost IT Productivity
This option’s session design tools simplify and streamline repetitive configuration and performance tuning tasks to boost your IT team’s productivity. You can easily create user-defined partitioning schemes. A graphical partitioning map helps you determine the best points of partitioning. Configurable session options make it easier to gather statistics used to maximize performance.

Optimize System Performance in Response to Changing Business Requirements
This option optimizes PowerCenter performance by empowering your IT team to gather in-depth session statistics to recognize trends and identify potential bottlenecks. Data transformation logic is abstracted from the physical execution plan, which enables IT to rapidly tune performance without compromising the logic and design of the original data mappings. With this option, you can easily optimize system performance in the face of increasing data loads and changing business requirements.