

# Informatica PowerCenter Partitioning

## Optimizing Parallel Processing on Multiprocessor Hardware

The Informatica® PowerCenter® Partitioning option optimizes parallel processing on multiprocessor hardware by providing a thread-based architecture and built-in data partitioning. GUI-based tools reduce the development effort necessary to create data partitions and streamline ongoing troubleshooting and performance tuning tasks, while ensuring data integrity throughout the execution process. As the amount of data within an organization expands and real-time demand for information grows, the PowerCenter Partitioning option enables hardware and applications to provide outstanding performance and jointly scale to handle large volumes of data and users.

### THE BENEFITS OF POWERCENTER PARTITIONING:

- Enable faster time to data delivery by optimizing parallel session execution while preserving data integrity
- Enhance developer productivity with intuitive, integrated GUI-based session design tools
- Improve ongoing performance with comprehensive management and monitoring capabilities

## Expanding Data Volumes and Shrinking Processing Windows Require Increased Efficiencies

As organizations gather increasing amounts of data and the need for real-time information grows, global businesses must process larger amounts of data in shorter periods of time if they are to continue to meet service-level agreements and business needs. To keep up with ever-increasing data processing demands, organizations are finding it necessary to purchase additional hardware and software licenses. This strategy, however, increases operational costs. A more efficient and cost-effective alternative is to scale existing hardware to meet these demands.

Improving scalability means using existing multiprocessor hardware and operating systems more efficiently and effectively by performing data partitioning across processors to enable parallel processing. Today, however, developers must manually allocate workflows across different CPUs to process workloads in the allocated timeframe. Optimizing sessions is a painstaking iterative process that is both costly and prone to error. Developers must also take care to manually ensure data integrity. As a result, development efforts and costs are increasing disproportionately to the data volumes.

Organizations wishing to both optimize their existing hardware and make the most efficient use of their developer resources need a solution that reduces the development effort necessary to achieve high performance. Such a solution should provide tools that simplify the task of partitioning sessions for more efficient processing across multiple CPUs, optimize parallel processing across multiprocessor systems, and enable ongoing session refinement to keep up with increasing data loads over time—all while ensuring high levels of data integrity.

## PowerCenter Partitioning—Simplifying Data Partitioning Across Multiprocessor Systems

The Informatica PowerCenter Partitioning option provides an intuitive GUI-based design tool that helps developers partition and optimize data flows across multiprocessor systems. The option enables efficient processing through parallel sessions and thread pooling to databases. Built-in Data Smart parallelism ensures accurate results without the need for manual scripting, thereby providing data integrity. The PowerCenter Partitioning option also provides comprehensive management and monitoring during execution to enhance troubleshooting and enable ongoing performance optimization. In the face of increasing data loads and the need to meet service-level agreements, these capabilities allow organizations to achieve high performance, quickly and easily, by maximizing the efficiency of their multiprocessing systems and maintaining that performance over time as data volumes continue to grow.

### Enable Faster Time to Data Delivery by Optimizing Parallel Session Execution While Preserving Data Integrity

The PowerCenter Partitioning option allows developers to take a set of data and divide it into smaller subsets for parallel (concurrent) processing on different processors within a multiprocessor system. By allowing different processors to share computational load, parallel processing provides a faster way to process large amounts of information.

The PowerCenter Partitioning option provides a wide range of choices for data partitioning. Pipeline parallelism breaks up the extract, transform, and load process into separate threads that work independently and in parallel to decrease execution time.

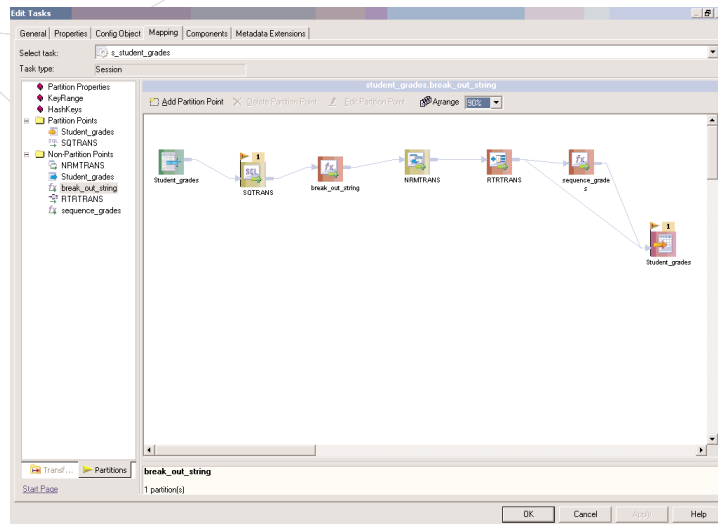


Figure 1: Graphical partitioning map

Transformation parallelism applies the same principles of Pipeline parallelism to the transformation process. Partition parallelism divides data processing across available CPUs to ensure scalability while managing the movement of data blocks between partitions. Data Smart parallelism preserves the quality and integrity of the data through the most complex partitioning and transformations without the need for manual scripting. The thread-based architecture exploits multiplexing or time slicing to

leverage the availability of additional processors while making the most of the operating system resources through efficient memory management. In addition, data is processed in blocks rather than a row at a time, which increases performance and speeds up processing. Together, these data partitioning capabilities improve throughput and scalability to enable faster time to data delivery and enhance the organization's ability to meet service-level agreements in the face of increasing data loads.

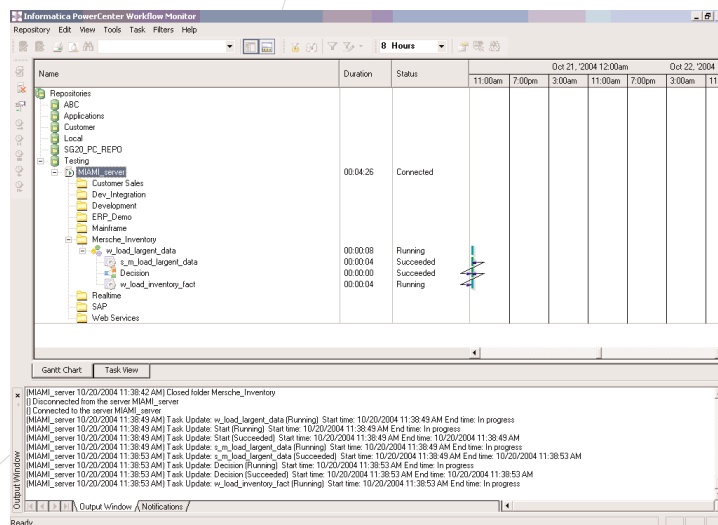


Figure 2: Gantt chart view

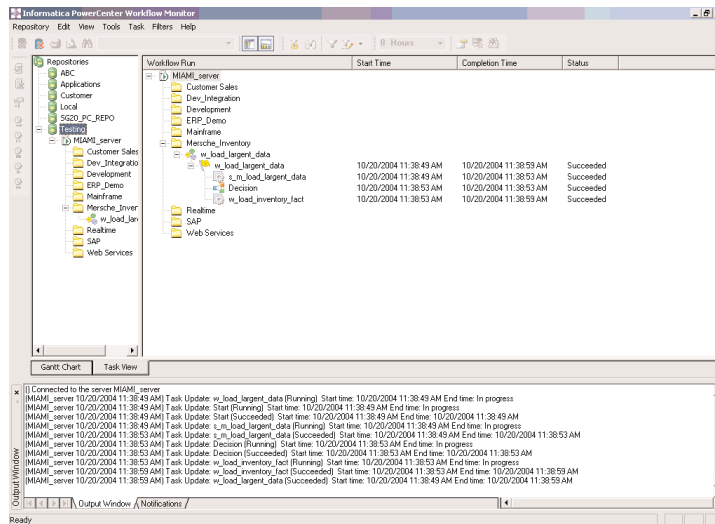


Figure 3: Task view

### Enhance Developer Productivity with Intuitive, Integrated GUI-Based Session Design Tools

The PowerCenter Partitioning option provides a GUI session design tool that simplifies the creation of user-defined partitioning schemes, including round robin, hash, and key range. A graphical partitioning map provides a picture of the mapping to help developers determine the best points of partitioning. Configurable session options such as error handling, recovery strategy, memory allocation, and logging make it easier for the user to gather as much execution metadata as possible. Whenever developers partition data across multiple processors, the process is an iterative one; using the PowerCenter GUI session design capabilities speeds this partitioning process and enhances developer productivity. Furthermore, the PowerCenter Partitioning option eliminates the need for time-consuming hand coding, which can result in wasted development efforts and added costs.

### Improve Ongoing Performance with Comprehensive Management and Monitoring Capabilities

Using the GUI tool, developers can easily gather in-depth session statistics such as throughput, rows/second, error details, and performance optimizations. These statistics enable developers to perform quick and in-depth execution analysis to identify potential bottlenecks and recognize trends. In addition, an integrated monitoring console enables users to view all session execution and dependency details in a variety of ways. For example, Gantt Charts (see figure 2) show chronological order, while task views illustrate the workflow order. The monitoring console also provides a single view of all data integration processes occurring on all processors within the system. By leveraging comprehensive metadata in an easy-to-use manner, the PowerCenter Partitioning option makes rapid session refinement possible, enabling developers to continually optimize system performance over time and maintain high performance while data loads continue to increase.

“One of the main drivers behind the success of our very high-performance, highly scalable enterprise data warehouse has been the performance and scalability of PowerCenter. . . . PowerCenter’s performance gives us the confidence to scale our data warehouse into the 10- to 20-terabyte range in the years ahead.”

—Mark Cothron, Data Warehouse Architect,  
Ace Hardware

Visit [www.informatica.com/products/powercenter](http://www.informatica.com/products/powercenter), or call 1.800.653.3871 to learn more.



Worldwide Headquarters, 2100 Seaport Boulevard, Redwood City, CA 94063, USA  
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.970.1179 [www.informatica.com](http://www.informatica.com)

**Informatica Offices Around The Globe:** Australia • Belgium • Canada • France • Germany • Japan • the Netherlands • Singapore • Switzerland • United Kingdom • USA

© 2004 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, Turning integration into insight, 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 tradenames or trademarks of their respective owners.

J50266 6528 (11/17/04)