Based on Talend’s award winning enterprise data integration technology, **Talend Integration Suite MP** is a highly scalable, massively parallel data integration platform that scales to the highest volumes of data.

Geared toward enterprises that need to process extreme data volumes in ever tightening time windows, Talend Integration Suite MP exceeds the most demanding requirements and supersedes all existing performance benchmarks.

Based on Talend Integration Suite, Talend Integration Suite MP includes its core modules—Business Modeler, Job Designer, and Metadata Manager—providing teamwork, and development consolidation, as well as industrialization and monitoring-oriented features.

Designed to take full advantage of available parallelism, Talend Integration Suite MP is based on the execution server architecture (CPU, OS, etc.) which greatly improves processing performance. The Massively Parallel features address the common issues enterprises encounter when dealing with processing of very large data volumes.

**Talend FileScale Technology**

Talend Integration Suite MP features the unique FileScale technology which leverages the execution server hardware architecture and maximizes the performance of low-level sort algorithms.

The FileScale technology automatically configures the processes using an executable file optimized for each client architecture and using statistical analyses of a sample of the data to be processed.

The FileScale technology works in bulk mode on (very) large files. It takes full advantage of the execution architecture as it is not restricted by the JVM or execution engine limitations typical of traditional data integration architectures.

FileScale technology sorts and transforms data using innovative high-performance mathematical algorithms for data processing.

Talend FileScale’s main characteristics include:

- The most efficient sort algorithm: sorting data is usually the major issue for enterprises dealing with very large amounts of data. Talend’s FileScale technology addresses this by selecting the most appropriate algorithm to sort the incoming data in parallel with any other data set operations.
- Dedicated FileScale components—based on the low-level programming language, FSLang—help transform data and perform all standard dataset-type operations including sorting, aggregating, deduplicating, filtering and transforming data, as well as checking file structures. The FileScale components were specifically designed and fine-tuned for their own single-task processing, resulting in dramatic gains in performance and efficiency.
Using MapReduce architecture, FileScale automatically breaks down any data processing operation into a number of granular processes, and distributes them across multiple child processes, leveraging the entire hardware architecture for optimal and fastest processing. This architecture also allows processing the data in data set mode—not row-by-row—for seamless efficiency and consistency.

Massively parallel data integration processes

The challenges involved in processing large amounts of data are similar to those of any large-scale project. Typically, the best approach is to divide the task into as many subtasks as possible and distribute them over all available resources in order to process them in parallel.

Similarly, Talend Integration Suite MP³ benefits from multi-server, multi-CPU, and multi-core architectures where code and separate sub-processes can be executed in parallel to make the most of the architecture. This massively parallel feature maximizes enterprise server capabilities and the number of processors available, greatly improving processing time.

Talend Integration Suite MP³ also automates the break down of data sets into many parallel streams, for further acceleration of the processing, leveraging the massively parallel loaders of leading RDBMS engines.

Other enterprise-grade features

- **Teamwork and development consolidation:** The Shared Repository is designed to consolidate all project information and enterprise metadata in a centralized repository shared by all stakeholders in the integration processes—business users, job developers, and IT operations staff.
- **Industrialization:** Numerous wizards help automate connections to heterogeneous sources, including enterprise-scale platforms such as SAP. Joblets can be reused, such as core models, to facilitate the industrialization of data integration processes. The Change Data Capture (Publish & Subscribe) feature quickly identifies and captures data that has been added to, updated in, or removed from database tables and makes the changes (and only these changes) available to subscribers.
- **Complex deployments:** Based on Web 2.0, including the AJAX technology, the administration center web application offers friendly ergonomics and fast-refresh capability. Job Conductor coordinates and schedules the execution of data integration jobs including event-based scheduling for real-time integration. Grid Conductor optimizes the scalability and availability of the integration processes by ensuring an optimal use of the execution grid, automatically distributing jobs across the grid. Distant Run enables the remote execution of jobs on specified systems.
- **Data Quality:** The optional data quality features include an in-depth data mining and profiling tool to detect non-compliant and poor-quality, data as well as a data cleansing tool that helps improve data for full data governance.
- **Execution monitoring:** The Activity Monitoring Console and Dashboard provide real-time gauges and status indicators, as well as a business-model orientation, enabling business stakeholders to view both the current status and historical data of any job.

Supported platforms:

Development: any Java-compatible platform
Execution (FileScale technology):
- Windows 32 & 64 bit
- Solaris & OpenSolaris (SPARC & Intel x86)
- HP-UX
- IBM AIX
- Linux 32 & 64 bit

Technical Support

Staffed by technical experts who are directly involved with Talend’s Research & Development organization, Talend’s Technical Support Centers provide prompt, effective, high-quality support services to Talend Integration Suite MP³ subscribers.

More Information: www.talend.com | info@talend.com