

## Data Integration Features Comparison Matrix

Talend Open Studio

vs

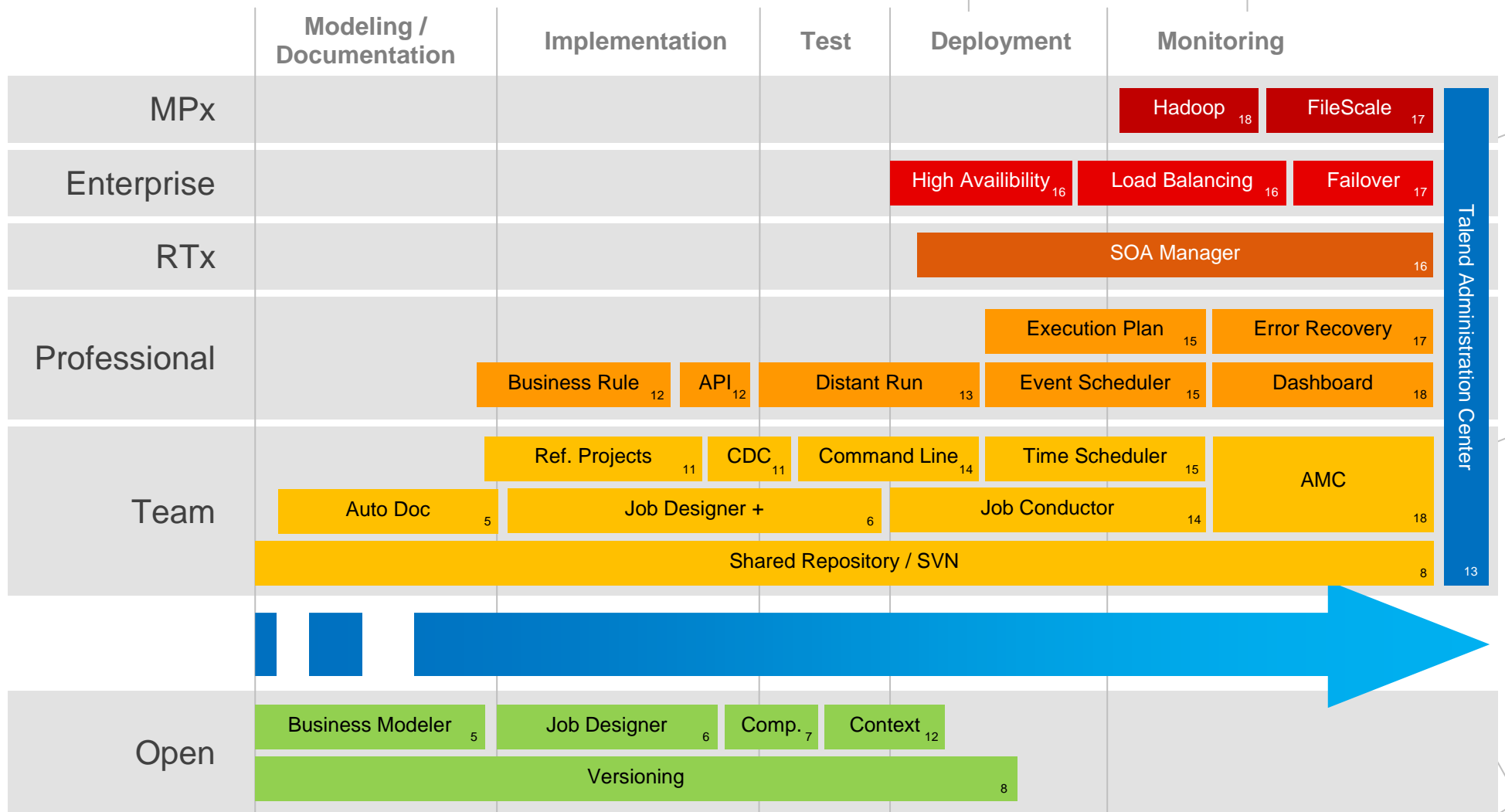
Talend Integration Suite

Team, Professional, RTx, Enterprise & MPx editions

<http://www.talend.com/products-data-integration/matrix.php>

Q1 2011

# Feature details



## License Type & Indemnification

### Open Source GPL license *this feature is included in:*

Open ●

Team

Professional

RTx

Enterprise

MPx

The GNU General Public License is a license that establishes the legal conditions for the distribution of free software of a GNU project.

The purpose of the GNU GPL license is to guarantee the following rights to the user:

- the right to execute the software for any use and without limitation;
- the right to analyze the functioning of the software and adapt it to their needs.

If the author of modifications to the software decides to distribute this software, he or she must do so under the GPL license.

The entire text of the GPL license can be viewed at: <http://www.opensource.org/licenses/gpl-2.0.php>

### Source code access *this feature is included in:*

Open ●

Team ●

Professional ●

RTx ●

Enterprise ●

MPx ●

Talend Open Studio source code is available at: <http://www.talendforge.org/trac/tos/>

For complete transparency and consistency, Talend also provides the clients who request it access to the source code of all of the tools available in the commercial edition.

### Subscription license *this feature is included in:*

Open

Team ●

Professional ●

RTx ●

Enterprise ●

MPx ●

The "enterprise" versions include value-added features (see below) and services that enhance the open source products; these versions are distributed under a commercial license.

Talend's pricing model guarantees transparency and predictability: the price is not based on the volumes of data or potential additional needs for connectors or CPUs, rather it corresponds to the number of developers (Studio), the level of features (edition selected) and the subscription term.

This subscription approach guarantees your return on investment: the number of licenses can be increased or decreased every year to adapt to the evolution of a project's range and its staff.

The Talend solutions are cheaper to deploy, maintain and support; they are 50 to 80% less expensive than the equivalent proprietary solutions.

## Indemnification

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx

Because open source software results from collaborative development efforts, the final code combines contributions from diverse resources. If the integration of the various contributions to the code is not carefully managed and controlled, the final software use might infringe upon the original contributors' rights.

The end user might then be subject to legal and financial prosecution for infringement, even though such infringement was not intentional.

Talend offers an Indemnification clause to its subscription customers. This is a guarantee for the user that Talend will provide legal and financial protection, in the event that the Talend code infringes the rights of a third party.

## Support & Documentation

### Community-based: forums, Bugtracker...

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx

The Talend user community, composed of tens of thousands of professionals, is extremely active. The main contributions of the community include:

- testing and the quality of new versions,
- requests for new features,
- product translation and localization,
- support and exchanges via the forums,
- development and sharing of new components, connectors, jobs, models and other plug-ins.

Talend Exchange enables community members to publish their own plug-ins in order to share them with other users. Most of these contributions are ultimately integrated into the product, after in-depth testing and improvements are completed by our in-house R&D team.

Additionally, Talend contributes to numerous key open source projects and is a member of the Eclipse and Apache Foundations.

### Enterprise grade support with SLAs

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx

By subscribing to Talend Support Services, you benefit from the experience of our in-house technical experts, who are daily in touch with our R&D team. These services were established to insure effectiveness, security, and peace of mind of our subscription customers. They are available in three levels: Silver, Gold and Platinum. Each of these levels is associated with guarantees related to the initial time spent to respond to a declared bug, the response time spent to provide a patch, etc.

<http://www.talend.com/professional-support/support.php>

## User Guide, Reference Guide

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx

The documentation of Talend Open Studio is available as free download in PDF format, in English and French. Two guides, the User Guide (276 pages) and the Components Reference Guide are available at: <http://www.talend.com/resources/documentation.php>

You can also buy a printed version of those guides on Amazon:

[http://www.amazon.com/Talend-Open-Studio-Guide-Version/dp/1453632077/ref=sr\\_1\\_1?ie=UTF8&s=books&qid=1296040035&sr=1-1](http://www.amazon.com/Talend-Open-Studio-Guide-Version/dp/1453632077/ref=sr_1_1?ie=UTF8&s=books&qid=1296040035&sr=1-1)

Furthermore, you can have a look at our Tutorials which are a valuable source of information!

You can access them at: <http://www.talendforge.org/tutorials/menu.php>

## Modeling / Documentation

### Business Modeler

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx



The **Business Modeler** is a non-technical tool (like Microsoft Visio). It helps you to structure all relevant documentation and technical elements supporting the data integration process in a business-friendly diagram allowing different teams (Design, Dev, Test, Prod...) to work on a common model, using a common tool.

For example, Business Users use business models to express their data integration needs. The IT development and operation staff can thus better understand these business needs and translate them into technical processes (Jobs). After each technical implementation stage (Jobs) is completed, the business model can easily be updated, showing the progress of development for other stakeholders to follow up.

DBAs can use business models to share the required DB connection metadata and system architect can thus have the big picture of the required needs in terms of data integration.

Designing business models is part of enterprises' best practices that organizations should adopt at a very early stage of a data management or integration project in order to ensure its success. Because Business Models usually help detect and resolve quickly project bottlenecks and weak points, they help limit the budget overspendings and/or reduce the upfront investment.

### Auto Doc

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx

This functionality permits generating, on request, a detailed technical documentation for all your jobs. This documentation gathers job metadata (author, version, status, update date, etc.), a graphical view of the job and all the parameters of all the components used in this job in an interactive format easy to use (HTML / XML).

This documentation can be easily enriched with personalized comments.

## Auto Doc+

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx



With AutoDoc+, the technical documentation (see previous paragraph) is automatically generated for each version of each job: when you save a job, its documentation is updated and stored in the Repository; therefore, it is automatically shared and available for all users.

AutoDoc+ also permits customizing the graphical display of this documentation by adding your own logo and the name of your company, or by changing the colors through a customized CSS.

## Implementation

## Job Designer

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx



The Job Designer provides both a graphical and a functional view of the actual integration processes using a graphical palette of components and connectors.

Integration processes are built by simply dragging and dropping the components and connectors onto a graphical workspace, drawing connections and relationships between them, and setting their properties.

The Job Designer capabilities give access, via an exhaustive library of components, to all types of source and target needed for data integration, data migration or synchronization processes.

Components and connectors cover all types of tasks and operations on the data itself, on the data management as well as on the data flow sequencing. Connectors help access and read/write all data source and target systems for data integration, data migration and data synchronization. Parameters are configured centrally in one view when selecting each component involved

in the Job or can be inherited from the Metadata Manager (Repository).

Complex components are equipped with dedicated and intuitive graphical interfaces or built-in wizards helping users to build their Jobs.

To maintain the readability of a Job design, the Job diagram can be divided into Subjobs, and then can be set out as child and parent Jobs to sequence their execution. Orchestration components as well as various types of relationships help user sequencing their process execution. A built-in console view lets users quickly monitor execution, check and track performance directly from the Studio.

## Components

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx



Talend offers native technical and business open source connectors to access all IT environments. This wide array of connectors is the key to the successful interoperability of applications and databases; it allows bridging diverse and heterogeneous data structures at unmatched performance rates. It is also continually expanding, enriching the features of the Talend data integration, data migration and data synchronization open source solutions.

More than 550 components are available, free of charge, 60% of which are designed and developed by the Talend community.

Connectors and components developed externally can be shared via the Talend Exchange (<http://talendforge.org/exchange/>). A number of submitted components go through validation and optimization by Talend, before they get integrated natively and supported.

Refer to <http://www.talendforge.org/components> for an exhaustive list of supported connectors.

## ETL support

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx

ETL (Extract, Transform & Load) is the default mode used by Talend's data integration solutions. It consists in processing data rows one right after the other in a flow mode. This mode is specifically adapted to be used in heterogeneous environments and it enables the integration of any technology in the source and target systems (web service, files, databases, MOM, business applications, etc.).

ETL mode can also be used in both batch and real time processing. The ETL processes can be run in parallel to further accelerate their execution.

Talend's unique architecture is not restricted to any execution engine since it generates autonomous processes that can be deployed on any server (internal or external to the company). Also, the ETL processes can be executed as close to the data as possible minimizing access time and bandwidth consumption in addition to eliminating bottlenecks.

In the same Job, this approach can be combined with the ELT approach (see following paragraph) to obtain the highest level of performance without any architectural constraints.

## ELT support

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx

Talend's data integration solutions also support ELT mode (Extract, Load & Transform), which consists in processing data in a set operation (using the Union, Except and Intersect operators) directly on the DBMS of the target database.

This mode is reserved for use in a homogeneous environment (one database). It has the advantage of benefiting from the material resources available and is particularly recommended when processing very large volumes of data in "Data Warehouse Appliance" environments like Teradata, Netezza, etc.

In the same Job, this approach can be combined with the ETL approach (see previous paragraph) to obtain the highest level of performance without architectural constraints.

## Versioning

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx



The versioning of items in the Talend Studio can be easily managed through the native manual versioning functionality.

A major and minor version number is automatically set at the Job creation, and can then be easily incremented over time and updates, via the dedicated Version control panel available directly in the Designer perspective of the Talend Studio.

All items created in the Studio can thus be versioned: Business Models, Jobs, Routines, Metadata, and Documentation...

The versioning is generally part of a best practice scheme that aims to facilitate the item reusability as well as the reverting to a previous development stage when needed.

## Shared Repository

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx



**The Shared Repository** (or Metadata Manager) is designed to consolidate all project information and enterprise metadata in a centralized repository shared by all stakeholders in the integration processes.

On the Studio side, users are granted access to projects according to their roles and permissions defined in Talend Administration Center.

This shared repository thus enables team work and collaboration between all people involved in an integration project. It helps to store and share all their Talend items: Business Models, Jobs (processes), Joblets, Routines, Metadata definitions (such as connections to source/target systems)...

Behind the shared repository is an industry-standard source manager (Subversion) that allows storing and managing all versions of all items.

An automatic locking system guarantees that the Job that is being designed is effectively locked and that no other user could change the same job at the same time.

From version 4.0, we leverage the full versioning power of subversion, allowing you to deal with different branches, check-in/check-out, manual or automatic commit, comments...

## Data Viewer

this feature is included in:

Open

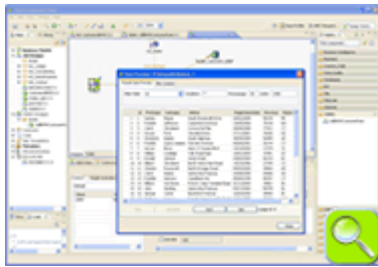
Team

Professional

RTx

Enterprise

MPx



While developing Jobs with Talend you may need to view the content of various source or target systems (files, DB, etc). The Data Viewer helps you drill down the data source/target systems regardless the application usually needed to open it: Notepad for txt & csv files, a SQL query browser for database tables, MS Excel for .XLS files, html browser, etc... No need to have multiple different tools and no need to browse systems to drill down to where the data lies, the Data Viewer uses the defined source/target path settings to go straight to the actual data.

The Data Viewer can save you a lot of time as this tool is directly accessible within the Studio, through a simple right-click on any component. It is a convenient way to view data contained in your source/target systems regardless of their format (Excel, DB table, CSV...) while you are developing your integration processes.

## Wizards

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx

## Dynamic Schema

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx

Dynamic schemas allow the designing of jobs with an unknown column structure and number. Depending on the choice of the developer, dynamic columns can be mapped directly to the target using Pass-through mode.

The main application of such functionality can be a replication scenario or simple one-to-one mapping of many columns. This feature makes designing these types of jobs easy. For example, a developer that needs to migrate a whole database with hundreds of tables can do so without knowing all of the table structures using a single job!

## Impact Analysis

this feature is included in:

Open

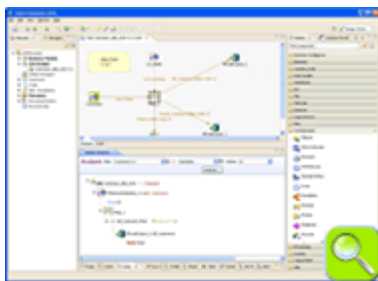
Team

Professional

RTx

Enterprise

MPx



The Impact Analysis feature helps you understand what could be the consequences of a change. This feature is available from the Metadata Manager. You can perform an Impact Analysis on any column of any metadata (database, file...). The result of the Impact Analysis shows in a graphical & interactive report where you can track down the column and see all operations applied to it, from source to target throughout the Job.

You can export this report as an HTML file.

## Data Lineage

this feature is included in:

Open

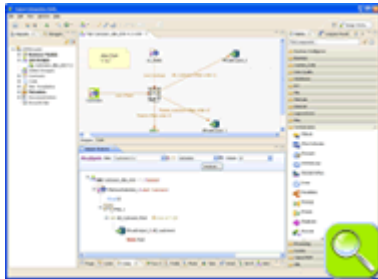
Team

Professional

RTx

Enterprise

MPx



The Data Lineage feature helps you understand where a change occurred.

This feature is available from the Metadata Manager and can be carried out on any column of any metadata (DB, file). The result of the data lineage shows in a report which traces a change from the target end component of a Job up to the source end.

You can export this report as an HTML file.

## Job Compare

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx



The Job Compare feature helps identify differences between two job versions or different jobs.

Job Compare is fully integrated in Talend Integration Suite Studio. The result of Job Compare is a visual and interactive report in html or xml where differences are highlighted.

In this example, the comparison report shows that the delimiter field in the tFileInputDelimited component properties is not defined the same way for both jobs being compared: in version 3.2 delimiter is "\t" while in version 4.2 it is "\n".

## Joblets

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx



Joblets help you factorize a job part (or Subjob) into a Joblet component. Simply select the components forming the Job part you need to reuse or want to factorize and click on the "Refactor to Joblet" menu item.

Automatically, the job design gets simplified, as the selected components are collapsed into a single Joblet component. This Joblet component can be shared through the dedicated Joblets folder in the Palette of components and is thus easily reusable in any other Job.

Joblets drastically simplify the maintenance of redundant and complex jobs.

Additionally, an "Impact Analysis" mechanism helps you find out which jobs use a defined Joblet.

## Reference projects

this feature is included in:

Open

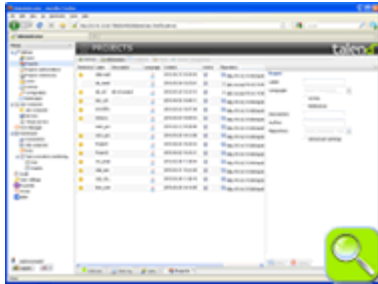
Team

Professional

RTx

Enterprise

MPx



The **Reference Projects** help avoid duplication (copy-paste) of items (Jobs, Routines, Documentation, Metadata...) between projects.

"Slave" projects are linked to one (or more) "Master" project(s) by reference and thus inherit items from the parent(s) project(s). The resources coming from the Master project appears in the Slave project in read-only mode: they are available for reuse and execution but cannot be modified.

Because a strong link is thus established between Slave and Master projects, then, as soon as someone modifies an item in the Master project, all slave projects get updated accordingly.

The Reference Projects share all redundant items of a project (Jobs, templates, metadata) in order to make them available to other projects. This feature helps to leverage and reuse the 30% of items that are usually common to all Data Integration projects, reducing drastically the associated maintenance.

## Change Data Capture

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx



Data warehousing involves the extraction and transfer of data from one or more databases into one or more target systems for analysis. However, this means the extraction and transfer of huge volumes of data which can be very consuming in both resource and time.

The ability to capture only the changed data in real time is known as **Change Data Capture** (CDC). Capturing changes reduces the traffic of data between systems and helps reduce ETL time.

Talend CDC architecture is based on a publisher/subscriber model. The publisher captures the data changes and makes them available to the subscribers (Talend Jobs). Subscribers utilize the data changes obtained from the publisher.

This feature detects changed records in real time, allowing the changed data to be sent immediately to Subscriber Jobs consequently cutting the time needed to load and update data during ETL or operational data integration.

Talend's Change Data Capture features the most commonly used modes: Trigger and Redo logs. The available mode depends on the type of databases involved.

## Business Rules

this feature is included in:

Open

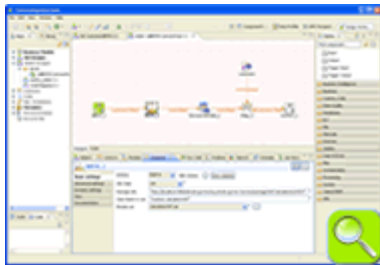
Team

Professional

RTx

Enterprise

MPx



**Business rules** are generally defined by business users through specification documents which are then interpreted and implemented by technical staff.

Talend Integration Suite embeds a business rule engine that helps users configure their own business rules. Users can thus define market segmentation criteria (by age, region...) and set their business rules via an Excel spreadsheet or through the Drools Guvnor interface directly the web-based Talend Administration Center.

The Drools Guvnor interface enables business experts to use a graphical editor to create and edit rules quickly and directly, control access to rules and other features, manage rule versions and modification over time. Rules can be tested and called from the

developed jobs. (see next slide)

## API

this feature is included in:

Open

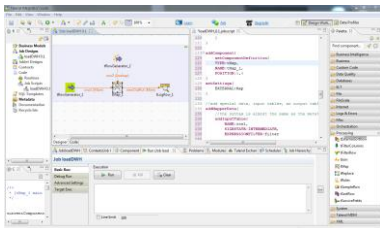
Team

Professional

RTx

Enterprise

MPx



Sophisticated but simple to grasp scripting language which enables programmatic job generation

Fully featured dedicated editor including syntax highlight, code completion aid & warning system

Automatic reverse engineering from an existing Job design to produce the corresponding script

## Test

### Context Management

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx



Contexts enable nearly any parameter of components / jobs to be externalized. This helps for example users to define parameters on the fly at run time or to use different settings for testing/production.

Contexts can be defined as needed for all types of environments (Development, Test, Production...) with no limitation in terms of number of context created.

Users can switch context at any time, design time or run time to use the defined setting.

Parameter values can also be changed via a dialog box at design and testing time. Additionally, a dedicated parameter-loading component can be used to override any value dynamically.

## Distant Run

this feature is included in:

Open

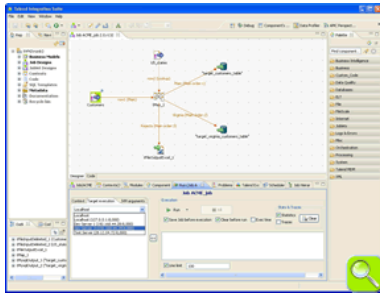
Team

Professional

RTx

Enterprise

MPx



The **Distant Run** feature enables the remote execution of jobs on any server directly from the studio.

This can be extremely useful when you need to test jobs, for example:

- in a configuration similar to the production environment,
- on various operating systems,
- upon request on specific systems,

as it avoids going through complex deployment procedures.

Target system can be selected dynamically at run time directly from the Studio. All regular debug, trace and real-time statistics options remain available in this remote execution mode.

## Deployment

### Talend Administration Center

this feature is included in:

Open

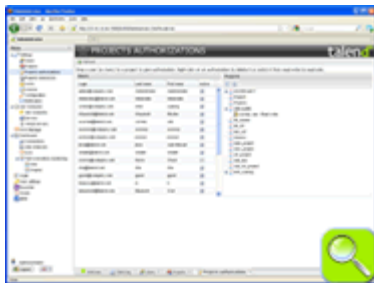
Team

Professional

RTx

Enterprise

MPx



All subscription offers come with one Studio (or more depending on the user number) and a software part which can be installed on a server and administrated through a web-based interface, the **Talend Administration Center**.

All Studios are thus no more in *local* mode but *remotely* connected to the projects defined in the Talend Administration Center.

Talend Administration Center is a lightweight application (in a browser, no deployment needed) that helps integration project managers to administrate users, projects, user privilege, license...

Project authorizations are assigned easily on a per user basis (supporting LDAP directory). And users are thus granted rights to access projects based on their role: No permission, Read Only, Read & Write...  
Users can then share repository items (Jobs, Business Models, DB connection metadata...) with other users, directly in their Studio, for the projects they are authorized on. More information on the shared repository in slides hereafter.

Depending on the Talend Integration Suite Edition you subscribed to, numerous additional plug-ins are available on the left navigation panel (Dashboard, SOA manager, Server manager...).

## Job Conductor

*this feature is included in:*

Open

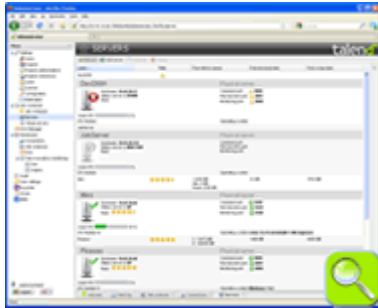
Team

Professional

RTx

Enterprise

MPx



The **Job conductor** coordinates the execution of data integration jobs. It provides a centralized execution interface from which all jobs can be started upon request or according to time-based (from Team Edition) or event-based (from Professional Edition) schedules.

The Job Conductor module relies on “JobServers” or agents which are small applications that are installed on each server where Jobs will be executed on.

After your agents are set up, the Job conductor allows you to monitor, in real time, all your hardware resources (available CPU, RAM, HD...) helping you distributing job executions over the grid, based on the best available server. The native JMX support allows you to monitor over 40 indicators. Any job can thus be deployed onto any server in just one click!

## Command Line

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx



Integration processes developed with the Job designer can be deployed, updated and executed outside the Talend Studio GUI, using the **Command Line** module.

Talend **Command Line** module provides a set of command line options that allow developers and administrators to easily perform batch operations.

Nearly all Job management functions offered through the Talend Studio and the Talend Administration Center are also available through the Command Line. This includes for example functions like: updating Job properties, promoting projects to production, exporting/importing Jobs or sets of Jobs, etc.

The Command Line feature makes it easy and quick to roll out numerous and complex Job deployments and executions including their dependencies and execution metadata.

The native command line Help provides an exhaustive list of all available commands with a short function description.

## Time Scheduler

this feature is included in:

Open

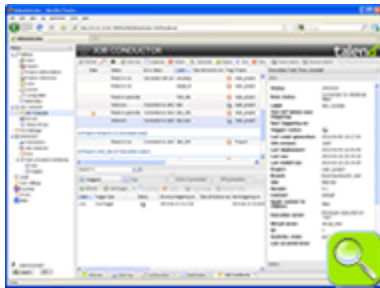
Team

Professional

RTx

Enterprise

MPx



The **Time-based scheduler** helps you roll-out a job execution at a defined time and date (first Monday of the month, every Tuesday...) or on a regular basis, over a period of time. A Task is used to centralize all information necessary for the job execution (project name, job name, job version, server...)

The task is then triggered upon schedule and the job is thus deployed & executed automatically onto the defined server at the defined time. A convenient status system helps you monitor the triggering state and the execution roll-out success/failure directly from the Job Conductor.

From the Professional Edition, an additional event/file based scheduling feature is available. (See Event scheduler slide).

## Event Scheduler

this feature is included in:

Open

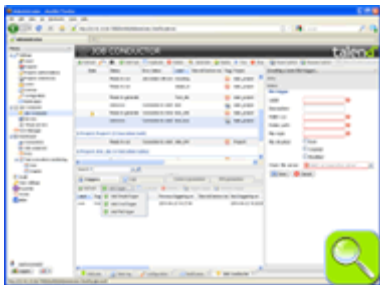
Team

Professional

RTx

Enterprise

MPx



The **Event Scheduler** extends time-based scheduling capabilities for real-time integration.

The event listener allows the process executions to trigger an execution on an on-demand basis, or based on an event.

Events can be file-based such as file appearing, disappearing or file modification or SQL-based using "wait for" condition. Once the expected event is identified, the execution task is triggered and the job deployment and roll-out are carried out.

You can easily add new event triggers to any task, extending the industrialization of automatic executions.

## Execution Plan

this feature is included in:

Open

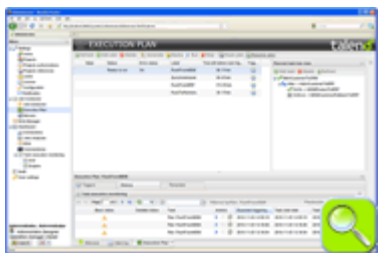
Team

Professional

RTx

Enterprise

MPx



The **Execution Plan** feature helps you sequence and orchestrate the various Job executions and ease the error recovery, directly from the Job Conductor. The execution plan is a task-based feature that outlines dependencies among different tasks orchestrating the execution sequence.

The task dependencies are defined through a hierarchical view of main and child tasks where each task can have a subordinate task.

Execution plans can be scheduled, triggered and can use environment-defined execution parameters from this single view of Job Conductor.

## SOA Manager / Webservices

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx



The SOA Manager offers a web-based and fully graphical environment to expose one or more data integration Jobs as services (Web Services), enabling their automatic deployment in and across heterogeneous applications and systems using SOAP binding (RPC or document-based). A dedicated WSDL wizard helps generate WSDL descriptors to expose Jobs as Web Services and find matching UDDI entries when consuming Web Services.

The SOA Manager features an advanced capability of incoming request management based on an optimized pooling and queueing system. The user-defined pool of active services handles a number of requests in real-time, while a queue manager handles the additional requests, buffering the throughput, for an asynchronous processing.

The SOA manager helps expose Jobs designed in Talend Integration Suite Studio in a Web service-centric environment and thus establishes interoperability across multiple applications and systems.

## Load Balancing

this feature is included in:

Open

Team

Professional

RTx

Enterprise

MPx



The **Grid Conductor** module (accessible through the Job conductor) optimizes the scalability and availability of the integration processes by ensuring an optimal use of the execution grid.

The grid conductor relies on the definition of virtual servers, which group available resources, regardless of the system type (CPU, OS...).

Tasks are assigned to virtual servers of the Grid Conductor rather than to a single execution server.

Via a constant monitoring of the resources available on the execution servers, Grid Conductor guarantees that all jobs execute smoothly at triggering time and fully leverage available resources, removing bottlenecks created by the traditional single-server approach.

This alleviates any concerns related to resource preemption when a large number of jobs run concurrently, or when non-dedicated servers are used. Grid Conductor also provides automatic fail-over in the event an execution resource becomes unavailable.

## High Availability

this feature is included in:

Open

Team

Professional

RTx

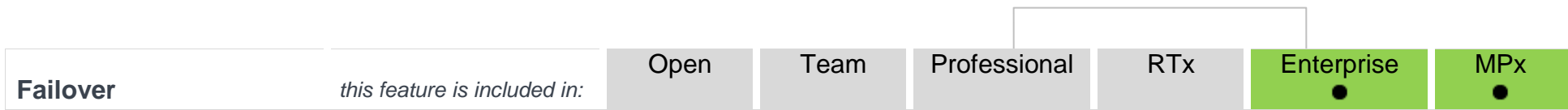
Enterprise

MPx



High Availability is achieved through the ability of deploying multiple Job conductors and job execution servers.

On the other end, clustering the databases guarantees failover and prevents any execution disruption.



**Benchmark Results**

**Configuration**

- 2 CPU Xeon 5520 Quad Core / 2.26GHz / 24 GB RAM
- Sun StorageTek 2540 Fiber Channel: 12 HD SAS 136Gb

**Results**

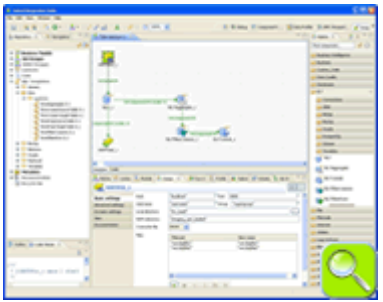
	1	10	100	300	550
#rows	6 million	60 million	600 million	1.8 billion	3.3 billion
data volume	740 MB	7.4 GB	74 GB	225 GB	415 GB
sorting time	6 sec	1 min 02 sec	21 min 30 sec	2 hr 01 min	5 hr 17 min

Independent benchmark made by Sun Micro

Talend Integration Suite MPx 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 MPx features the unique **FileScale** technology which leverages the execution server hardware architecture and maximizes the performance of low-level sort algorithms. 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. It leverages the MapReduce architecture to automatically break down any data processing operation into a number of granular processes, achieving great performances. See Sun Microsystems workbench: [http://blogs.sun.com/aja/entry/talend\\_s\\_new\\_data\\_processing](http://blogs.sun.com/aja/entry/talend_s_new_data_processing)



Apache Hadoop is an open source Java software framework that supports data-intensive distributed applications. It leverage Map Reduce architecture and enables applications to work with thousands of nodes and petabytes of data using large grid of inexpensive servers. Talend Integration Suite MPx includes a native support for Hadoop making it possible to scale to any level and support any complex data type, so companies can leverage their Hadoop clusters for peak data volumes and complex transformations.

A dedicated set of components available from the component Palette help read and write HDFS as well as Hive systems and include ELT and SQL template features.

# Monitoring

## AMC

*this feature is included in:*

Open

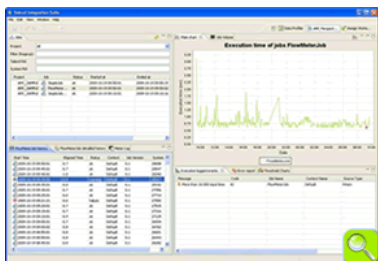
Team

Professional

RTx

Enterprise

MPx



**Talend Activity Monitoring Console** is a convenient graphical interface and a centralized supervising tool.

It provides detailed monitoring capabilities that can be used to consolidate the collected log information, understand the underlying Job interaction, prevent faults that could be unexpectedly generated and support system management decisions.

The Activity Monitoring Console monitors job events (successes, failures, warnings, etc.), execution times and data volumes through a single console from a centralized point.

This tool is available as a stand-alone tool or may be fully integrated in the Studio.

## Dashboard

*this feature is included in:*

Open

Team

Professional

RTx

Enterprise

MPx



The **Dashboard** is a Web-based version of the Activity Monitoring Console that can be accessed easily through a Web browser.

The Dashboard provides execution performance diagrams and status indicators, enabling any stakeholder to view both the current and historical status of any integration process execution.

It also provides detailed monitoring capabilities that can be used to consolidate log information collected, understand the underlying component and job interaction, provide task execution information in a timely manner, prevent faults that could be unexpectedly generated, support the system management decisions.

## Error Recovery

this feature is included in:

Open

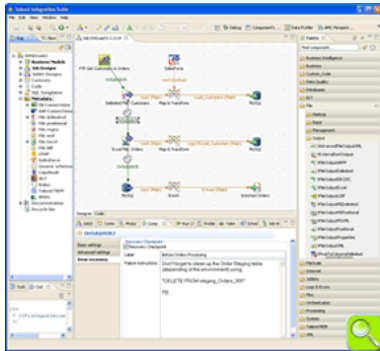
Team

Professional

RTx

Enterprise

MPx



Job execution processes can be time-consuming, as are backup and restore operations. Talend Integration Suite Studio includes a recovery checkpoint capability that is set up at Job design time.

In case of failure, processes can be resumed from one of the checkpoints. Job developers can also design and integrate specific error management in response to specific error conditions using the checkpoint "on-failure" instruction function.

Recovery checkpoints can be appropriately initiated at specified intervals of the data flow (on trigger connections). The purpose of it is to minimize the amount of time and effort necessary when a Job execution process needs to be restarted due to a failure.

With the help of the error recovery checkpoint feature, the process can be restarted from the latest checkpoint prior to the failure (or any other checkpoint before the failure occurred), rather than from the beginning of the Job execution process.