

Talend Open Studio

The Open Source Data Integration platform

<http://talend.ow2.org>



- ETL/data Integration
- Business Intelligence
- GPL License
- Easy to Use
- Java SE

Talend democratizes data integration. Through an open and innovative platform, Talend provides a simple, powerful solution.

Many organizations have been forced to address data integration needs by implementing custom programs or routines. More often than not, these integration routines are designed to solve a specific requirement - rather than architected to address global issues - and with no attempt to leverage existing processes. This approach poses robustness challenges, represents a real risk to the integrity and consistency of the information system, and is difficult and expensive to maintain.

A versatile data integration solution

Talend Open Studio is a powerful and versatile open source solution that addresses all of an organization's data integration needs:

- Synchronization or replication of databases
- Right-time or batch exchanges of data
- ETL (Extract Transform Load) for BI or analytics
- Data migration
- Complex data transformation and loading
- Basic data quality

Talend Open Studio comprises three major applications (Business Modeler, Job Designer, and Metadata Manager) within a single graphical development environment based on Eclipse, which is easily customized to specific corporate needs. Users get support from the active Talend Community and leverage contributions natively in their environment.

Business-oriented process modeling

Talend Open Studio's Business Modeler leverages a top-down approach, allowing line-of-business stakeholders to get involved in the design of the integration processes and to monitor development progress. The Business Models are non-technical, business-oriented views built using a convenient library of symbols and links.

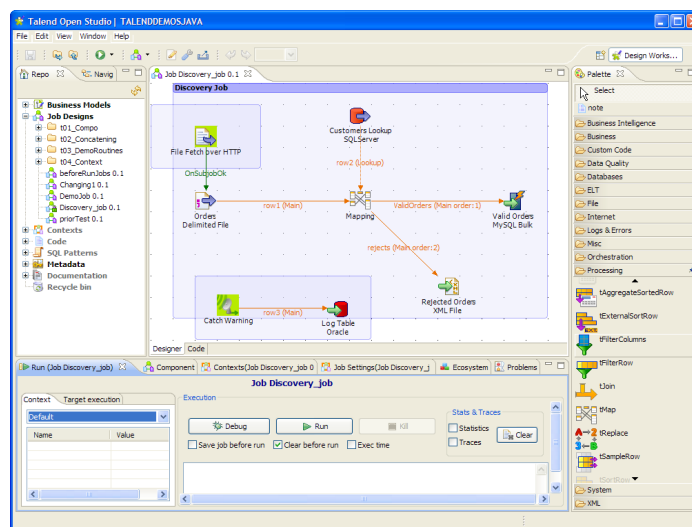
The Business Modeler also regroups all relevant documentation supporting the data integration process in a business-friendly diagram. This is a very efficient way of monitoring the integration processes and analyzing potential problems.

Graphical process development

Talend Open Studio's 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 components and connectors onto the workspace, drawing

connections and relationships between them, and setting their properties.



Components and connectors cover all types of tasks and operations on the data itself or on the sequencing of the workflows. Connectors help access and read/write all data source and target systems. Properties are configured centrally in one view when selecting each component involved in the Job or can be inherited from the Metadata Manager.

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. Other orchestration components as well as various types of relationships help users sequencing their process execution. A built-in console view lets users quickly monitor execution and track performance.

Metadata-driven design and execution

Talend Open Studio is a metadata-driven solution. All metadata is stored and managed in a Metadata Manager - the Repository - shared by all the modules. The Repository centralizes all project information and ensures consistency across all integration processes.

Meta information related to source and target systems of the integration processes is easily loaded in the Metadata Manager through advanced database or file introspection,

facilitated by a number of wizards. The Metadata Manager is based on an open relational model, where Job dependencies can be easily identified, facilitating the maintenance of the data integration, data migration and data synchronization Jobs. Contextual data, such as database connection details or file paths can also be centralized in the Metadata Manager, facilitating their use and update. Additional pieces of codes, routines or methods can also be unified in the Repository, facilitating the reuse and refactoring of process parts.

Advanced and versatile connectivity

Talend Open Studio offers native technical and business connectors to 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 solution.

Some examples of connectors:

Files: CSV, Excel, positional, XML, LDIF, EBCDIC, multi-schema, unstructured

Databases: MySQL, MS SQL, DB2, Oracle, Ingres, PostgreSQL, Sybase, Vertica, Infobright, ParAccel, Teradata, JDBC, AS/400, eXist, Hive, LDAP, Informix

Applications: SAP, SugarCRM, MS Dynamics, Salesforce, Openbravo, HL7, Talend MDM

Log & error: info, warning, statistics, filtering

Data quality: deduplication, fuzzy logic matching, CRC

Transformation: filtering, mapping, reference, aggregation, lookup, XSLT, (de)normalizing

Internet & Protocols: Web Services, FTP, HTTP, SSH, POP

Orchestration: iteration, loop, replication, unification

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

Talend Open Studio leverages industry-standard languages that include Java, Perl and SQL. This allows users to easily enrich existing components or to create their own. A dedicated community application - Talend Exchange - helps users plug these newly created components natively into the environment. Users can also write routines and other pieces of code and store this information centrally in the Repository for reuse.

Real-time debugging

Talend Open Studio includes powerful testing, debugging and tuning features that allow the real-time tracking of data flowing through the whole transformation processes, including execution statistics and an advanced trace mode.

And, of course, all code generated by Talend Open Studio, regardless of the target language, is always visible and accessible from the design environment.

Deployment and maintenance

Advanced execution context management (test, staging, production, etc.) facilitates integration processes deployment. Implicit loading of context parameters directly in the job design helps develop the various execution environments and manage them easily. Deployment of processes across

enterprise's systems can easily be performed as data services or as data integration services via the convenient export tool.

Automated documentation generation provides complete and up-to-date technical reference documentation (in XML and HTML) that helps various users and stakeholders to maintain and update inherited processes.

The dependency detection feature helps users identify dependencies among integration processes developed within Talend Open Studio and simplifies the global update of the large number of processes stored centrally in the Repository.

Robust and scalable execution

Unlike many integration solutions which are based on a centralized integration server, or can only use RDBMS engines to process data, Talend Open Studio allows users to export processes into executable files that can be distributed across a grid of systems or exposed as Web Services. These systems do not need to be dedicated to executing integration processes. Instead, Talend Open Studio leverages available resources, regardless of their nature.

Talend Open Studio leverages both the traditional ETL (Extract-Transform-Load) approach as well as the ELT (Extract-Load-Transform) approach. ELT leverages the power of the RDBMS engines to execute the data transformations inside the database, achieving unmatched performance for high-volume batches. For each subset of a process, it is possible to choose the most suitable approach, and hence to obtain the highest level of performance and scalability.

This architecture design, which is especially suited to leverage grids of inexpensive servers, as well as high-range systems, enables data to be processed at a location closest to its source (thus decreasing data transfers), and maximizes the use rate of computing resources.

Active Community of users

Users of Talend Open Studio can rely on a large and active community of users via the TalendForge online community tools:

- Forum: <http://www.talendforge.org/forum/> is the place where all questions can find an answer.
- Wiki: <http://www.talendforge.org/wiki/> provides useful installation and general use information.
- Bugtracker: <http://www.talendforge.org/bugs/> is the tool to report any issue with Talend Open Studio.
- Exchange: <http://talendforge.org/exchange/> is where external contributions can be shared.
- Babili: <http://www.talendforge.org/babili/> is the localization portal where all Talend products can be translated.

Thanks to community contributions, Talend Open Studio is now available in more than 15 languages.

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