talend | Connecting the Data-Driven Enterprise

Talend Project Audit

User Guide

6.4.1

Adapted for v6.4.1. Supersedes previous releases.

Publication date: June 29, 2017

Notices

End User License Agreement

Table of Contents

Preface

1. General information

1.1. Purpose

This User Guide helps to understand the reports resulted from auditing a project based on *Talend Project Audit* in a normal operational context.

Information presented in this document applies to *Talend Project Audit* **6.4.1**.

1.2. Audience

This guide is for business users in charge of checking the quality of the processes used to create data integration Jobs realized in *Talend studios*.

1.3. Content

Chapter 1: gives *Talend Project Audit* main concepts.

Chapter 2: describes *Talend Project Audit* report.

1.4. Typographical conventions

This guide uses the following typographical conventions:

- text in **bold:** window and dialog box buttons and fields, keyboard keys, menus, and menu options,
- text in **[bold]:** window, wizard, and dialog box titles,
- text in `courier`: system parameters typed in by the user,
- text in *italics*: file, schema, column, row, and variable names,
- The  icon indicates an item that provides additional information about an important point. It is also used to add comments related to a table or a figure,
- The  icon indicates a message that gives information about the execution requirements or recommendation type. It is also used to refer to situations or information the end-user needs to be aware of or pay special attention to.
- `Any command is highlighted with a grey background or code typeface.`

2. Feedback and Support

Your feedback is valuable. Do not hesitate to give your input, make suggestions or requests regarding this documentation or product and find support from the **Talend** team, on **Talend Community** at:

https://community.talend.com/

Chapter 1. Overview of Talend Project Audit

This document introduces the audit approach used by *Talend Project Audit* and lists the key areas that are audited in a project.

1.1. What is Talend Project Audit

Talend Project Audit transforms project data flows to valuable business information. It introduces an auditing approach for evaluating various aspects of Jobs implemented in *Talend studios*. The aim is to provide solid quantitative and qualitative evaluation for process-oriented decision support.

A project audit is the process of collecting and evaluating information about job operations designed in a Studio. *Talend Administration Center* is required to invoke an audit. The evaluation of obtained information determines if the technical processes and data flows you are using in your audited project are operating effectively and efficiently to achieve the project goals or objectives.

Audit results are presented in printable reports generated to provide extensive data regarding different areas in the audited project.

For more information about the areas audited in a project, see *Key areas audited in a project*.

You can later use the results provided by *Talend Project Audit* to get a clear picture regarding the status and performance of different elements in the audited project. Using this significant amount of information, you can analyze and monitor the performance of these elements in order to improve efficiency.

For detailed information about the data presented in *Talend Project Audit* reports, see *Talend Project Audit reports*.

Before auditing, ensure that you have the required rights to access a project or the project items you want to audit. For further information about the rights required to access a project and its items, see Talend Administration Center *User Guide.*

1.2. Key areas audited in a project

Talend Project Audit is an inquiry into the key data processing aspects of a project that is designed in one of *Talend* studios.

Talend Project Audit provides several functions for auditing a project through investigating different elements in Jobs designed in a Studio. *Talend Project Audit* reviews:

- degree of difficulty of Jobs and components used in Jobs,
- Job error management,
- documentation and Job versioning,
- usage of metadata items in Job designs,
- layout-related issues in Job and subjob graphical designs,
- Job analysis.

These aspects depend on each other and are correlated accordingly during project investigation in *Talend Project Audit*.

For more information about the audit results of the above key areas, see *Talend Project Audit reports*.

1.3. Benefits of auditing a project

Talend Project Audit helps to enhance a high level of awareness of the quality and effectiveness of the standards and techniques used throughout your projects realized in one of *Talend* studios. It provides information that enhances key areas in your project and thus leads to better realization of future data integration projects.

The evaluation of obtained information determines if the projects designed in a studio operates effectively and efficiently to achieve the organization's objectives.

This opportunity will lead to quick adoption of best practices to help your organization to apply data integration subsequent projects effectively.

To summarize what has been discussed earlier, possible benefits of auditing a project are:

- providing statistical information about key concepts in Jobs included in a project,
- identifying errors and weaknesses, if any, in a project,
- evaluating the performance of the techniques used in the Jobs included in the investigated project,
- improving future project performance by avoiding misused processes and techniques.

1.4. Moving forward

Strong project auditing practices are critical to the successful execution of all data-related projects realized in *Talend studios* and, consequently, to the overall accomplishment of your organization.

Periodic audits of the processes of different projects ensure that standards and techniques related to job realization are identified, investigated, and fixed and the process is improved. As a result, every aspect of the project is equally important and must be reviewed thoroughly.

Chapter 2. Talend Project Audit reports

Auditing specific concepts in a data integration project will result in generating a project analysis report. *Talend Administration Center* is required for this generation. This report will provide you with information on almost all aspects of the audited project designed in one of *Talend* studios in order to help you make process-oriented decisions based on the results of the audit.

This chapter aims at helping the interpretation of a *Talend Project Audit* report. It discusses report content related to each and every audited aspect in a data integration project.

If you need further information about how to carry out an audit and then generate a report accordingly, see *Talend Administration Center* **User Guide**.

Talend Project Audit report is divided into one introductory section listing the properties of the audited project along with six other sections, each presenting quantitative and qualitative data about the investigated key area in the audited project.

The below sections discuss in details the audit results for every key area in the investigated project.

2.1. Project properties

The table in this section lists identifying information about the audited project:

Item	Description
Project file	Path that specifies the location of the audited project
Label	The technical name (used by the system) of the actual project file
Description	User-defined description of the audited project
Author	Login of the user who initially created the project in *Talend Studio*
Product version	the version of the Studio used to build the project to be audited

2.2. Job/Component complexity

Talend Project Audit provides a method to assess the complexity of Jobs used in the audited project. The analysis incorporates numeric values for specific complexity elements that are weighted in their contribution to job overall complexity.

A full description of findings is then presented in tables and charts.

2.2.1. Project summary

The table in this section lists quantitative indicators of all elements used in the audited project.

Item	Description
Job count	Total number of Jobs used in the project.
Subjob count	Total number of subjobs used in the project
Component count	Total number of components used in the project.
Note count	Total number of notes used in the project
Context variable count	Total number of context variable used in the project
Schema count	Total number of schemas used in the project
Schema column count	Total number of schema columns used in the project
Complexity rating	Complexity rating for the project that is equal to the sum of all individual job ratings

Click the highlighted text (clickable text) in your Audit report to display the detailed referenced content without having to scroll down the pdf pages to find it.

2.2.2. Job rating chart and details

This section provides visual representation of the investigated data related to the complexity rating of all Jobs in the audited project.

The chart represents quantitative and qualitative structures that show the number of Jobs used in the project along with their complexity rating, ranging from `very simple` to `very complex`.

Job complexity rating is calculated using numeric values specific to the elements complexity in the investigated Job. Examples of those elements are number of components used in the Job, number of context variables used in the Job, number of the **tMap** components used in the Job and so on.

The below list gives the basis for calculating job complexity:

- n*number of components per Job,
- n*number of context variables per Job,
- n*number of the **tMap** components used in the Job,
- n*number of input columns in the **tMap**(s) used in the Job,
- n*number of output columns in the **tMap**(s) used in the Job,
- n*number of inputs in the **tMap**(s) used in the Job,
- n*number of outputs in the **tMap**(s) used in the Job.

where "n" is a coefficient that helps balancing the actual "weight" of each criteria in the audit results.

And then the resulting figures are defined as the following to give the different job complexity ratings used:

- `0 - 50`: `very simple`,
- `51 - 100`: `simple`,
- `101 - 300`: `moderate`,
- `301 - 500`: `complex`,
- `> 500`: `very complex`.

The job rating chart is accompanied (in the **Job rating - details** section in *Talend Project Audit*) by job rating details providing numeric values for each complexity element shown.

The below table lists the numeric values for different complexity elements of the investigated Job.

Item	Description
Identifier	Name of the Job
Auth.	Email of the author of the Job
Creation	Creation date of the Job
Update	Date of the last modification done on the Job
Status	Status of the Job
Version	Version of the Job
Components	Number of components used in the Job
Context	Number of context variables used in the Job
Notes	Number of notes used in the Job
Rating	Complexity rating for the Job based on the defined criteria

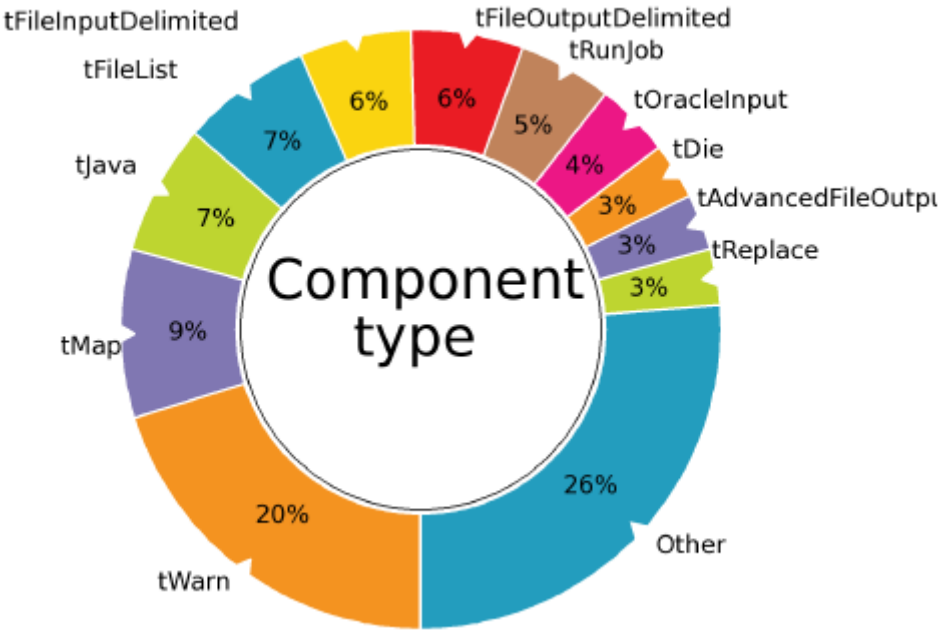
Click the highlighted text (clickable text) to display the detailed referenced content without having to scroll down the pdf pages to find it.

2.2.3. Component usage pie chart

The components pie chart (pie-shaped chart) is used to give a quick overall visual impression of the audit results regarding proportions of each of the component types used in the audited project.

It is split into segments, illustrating percentages of components based on their types. It uses those percentages or fractions to compare component types used in the project. The whole is equal to 100%.

Below is an example of a pie chart for an audited project.



All component types with less than 2.5% proportion will be grouped under **Other**.

2.2.4. Components usage details

The component pie chart is accompanied by component usage details providing numeric values for each component type shown on the pie chart.

The below table lists numeric values for components types.

Item	Description
Component type	Name of components used in the audited project
Count	Number of occurrences of the same component per project
Percentage	The fraction of 100 representing the frequency of each of the components used in the audited project

Click the highlighted text (clickable text) to display the detailed referenced content without having to scroll down the pdf pages to find it.

2.2.5. Jobs grouped by components

The tables in this section provide lists of Jobs using the same components and indicate, at the bottom of each table, the total number of a component's occurrences per Job.

Item	Description
Job	Name of the Job that uses a specific component
Count	Number of occurrences of the same component per Job

Click the highlighted text (clickable text) to display the detailed referenced content without having to scroll down the pdf pages to find it.

2.2.6. Schema column chart and details

This section provides visual representation of the investigated data related to the complexity rating of the schema and schema column count used in the audited project.

The chart represents quantitative and qualitative structures to show in one instance the number of columns used in each schema along with complexity rating, ranging from `very low` to `very high`.

Schema column complexity rating is calculated using numeric values for the number of columns used in each schema in the audited project.

Schema column complexity ratings are defined as:

- `0 - 10`: `very low`,
- `11 - 30`: `low`,
- `31 - 60`: `moderate`,
- `61 - 100`: `high`,
- `101 - 999999`: `very high`.

The schema column rating chart is accompanied (in the **Schema columns-details** section in *Talend Project Audit* report) by schema columns details providing numeric values for each element shown in the chart.

The below table lists the numeric values for column count per schema and schema count per project.

Item	Description
Schema column count	Number of columns per schema
Schema count	Number of schema per project, grouped by number of columns used in schema
Percentage	The fraction of 100 representing the frequency of each group of schemas (grouped by column count) used in the audited project

2.2.7. Jobs with 10 most schema column count

Another method used in *Talend Project Audit* for assessing the complexity of the audited project is counting schema columns in all Jobs in the investigated project and then grouping Jobs in groups relative to the 10 highest schema column figures.

Consequently, this section in *Talend Project Audit* report will provide 10 tables presented in ascending order. Each table groups all Jobs that hold the same figure among the 10 highest schema column figures.

Each of the 10 tables is preceded by the schema column figure used as the grouping factor for all Jobs listed in the table. The number of schemas used in each of the listed Jobs is given in a separate column.

Item	Description
Job	Name of the Job in the audited project
Schema count	Number of schema per Job

This report is parameterized. It is possible to define the number used as grouping factors.

This type of investigation will help diagnosing system performance problems, for example through identifying the columns that are present in the job design but not really used during job execution.

An example of this is when you use the **Lookup** flow in your job design. Usually, a lookup schema has numerous columns and you only use limited number in your Job.

2.2.8. Jobs with 5 least schema column count

The method used here for assessing the complexity of the audited project is calculating schema columns in all the Jobs used in the project and then grouping Jobs in groups relative to the 5 lowest column figures.

Consequently, this section in *Talend Project Audit* report will provide 5 tables presented in ascending order. Each table groups all Jobs that hold the same figure among the 5 lowest schema column figures.

This type of investigation will help identifying Jobs where schema is not defined, for example.

Each of the 5 tables is preceded by the schema column figure used as a grouping factor for all Jobs listed in the table. The number of schemas used in each of the listed Jobs is given in a separate column.

Item	Description
Job	Name of the Job used in the audited project
Schema count	Number of schema per Job

This report is parameterized. It is possible to define the number of figures used as the grouping factor.

2.3. Job error management

This section provides detailed information about trigger types and trigger numbers used in the investigated project. It also details what trigger types are used in what Jobs.

2.3.1. Trigger usage in a project

The table in this section lists numeric values for trigger types used in the audited project.

Item	Description
Trigger type	Triggers can be any of the following types: **-On Subjob Ok** **-On Subjob Error** **-Run if** **-On Component Ok** **-On Component Error**
Trigger count	Number of triggers per project
Percentage	The fraction of 100 representing the frequency of each trigger type in the audited project

2.3.2. Trigger usage in Jobs

The table in this section lists numeric values for trigger types used in Jobs part of the audited project.

Item	Description
Trigger type	Triggers can be any of the following types: **-On Subjob Ok** **-On Subjob Error** **-Run if** **-On Component Ok** **-On Component Error**
Job count	Number of Jobs using at least one trigger type
Percentage	The fraction of 100 representing the frequency of each trigger type in a job fraction relative to job total number in the project. For example, if **On Subjob Ok** is used in 25 Jobs where global job count in the project is 154 and the number of jobs using at least one trigger type is 48, the percentage can be figured by the proportion 25*100 / 154 which works out to be about 16.23

2.4. Job documentation/versioning

This section provides detailed information about folder level, item type, item status, and item versioning used in the investigated project. It also details the usage of job folders and job versioning in the investigated project.

The audit results will show if you profit from the documentation and versioning functions in the Studio to correctly classify different types and versions of documents and Jobs in the investigated project.

2.4.1. Documentation item type

The table in this section lists numeric values for item types used in the audited project.

Item	Description
Document item type	Type of the document item
Item count	Number of items of the same type per project
Percentage	The fraction of 100 representing the frequency of each item type in the audited project

2.4.2. Documentation folder level

The table in this section lists numeric values for folder hierarchy levels used in the audited project.

Item	Description
Document folder level	Numeric value that represents folder level: **0**: the item is not attached to a folder **1**: the item is attached to a one-level-deep folder **2**: the item is attached to a two-level-deep folder ...
Folder count	Number of folders (of the same level) per project
Percentage	The fraction of 100 representing the frequency of each folder level in the audited project

2.4.3. Documentation item status

The table in this section lists numeric values for different item statuses used in the audited project.

Item	Description
Document item status	Status of the item used: **checked**, **unchecked**, **validated**
Item count	Number of items of the same status per project
Percentage	The fraction of 100 representing the frequency of each item status in the audited project

2.4.4. Documentation item versioning

The table in this section lists numeric values for item versions used in the audited project.

If you create many versions of the same item, only the last item version is taken into account.

Item	Description
Document item versioning	Last version of the added item
Item count	Number of items of the same version per project
Percentage	The fraction of 100 representing the frequency of each item version in the audited project

2.4.5. Job folder usage

The table in this section lists numeric values for job folder hierarchy levels used in the audited project.

Item	Description
Folder level	Numeric value that represents job folder level: **0**: the Job is not attached to a folder **1**: the Job is attached to a one-level-deep folder **2**: the Job is attached to a two-level-deep folder ...
Job count	Number of Jobs of the same level per project
Percentage	The fraction of 100 representing the frequency of each job level in the audited project

2.4.6. Job versioning usage

The table in this section lists numeric values for job versions used in the audited project.

If you create many versions of the same Job, only the last job version is taken into account.

Item	Description
Job version	Last version of each of the Jobs in the audited project
Job count	Number of Jobs of the same version in the audited project

Percentage	The fraction of 100 representing the frequency of each job version in the audited project

2.5. Metadata

Talend Project Audit calculates the percentage of the usage of repository, property, and schema metadata in the investigated project.

The audit results will show if you use the **Metadata** repository integrated in *Talend Studio* to store your predefined metadata and thus be able to reuse it in different jobs.

2.5.1. Repository metadata

The table in this section lists numeric values for metadata types used in the audited project.

Item	Description
Metadata type	Metadata can be of one of the following types: **-DB connections** **-File delimited** **-SAP connections** **-File Excel** **-File regex** **-File positional** **-File Idif** **-File xml** **-LDAP schema** **-WSDL schema** **-Generic schema** **-Salesforce schema**
Count	Number of each of the above metadata types used in the audited project
Percentage	The fraction of 100 representing the frequency of each metadata type in the audited project

2.5.2. Property metadata

The table in this section lists numeric values for property metadata types used in the audited project.

Item	Description
Metadata type	Property metadata can be of one of the following two types: **-Built in** **-Repository**
Count	Number of each of the above property metadata types used in the audited project
Percentage	The fraction of 100 representing the frequency of each property metadata type in the audited project

2.5.3. Schema metadata

The table in this section lists numeric values for schema metadata types used in the audited project.

Item	Description
Metadata type	Schema metadata can be of one of the following two types: **-Built in** **-Repository**
Count	Number of each of the above schema metadata types used in the audited project
Percentage	The fraction of 100 representing the frequency of each schema metadata type in the audited project

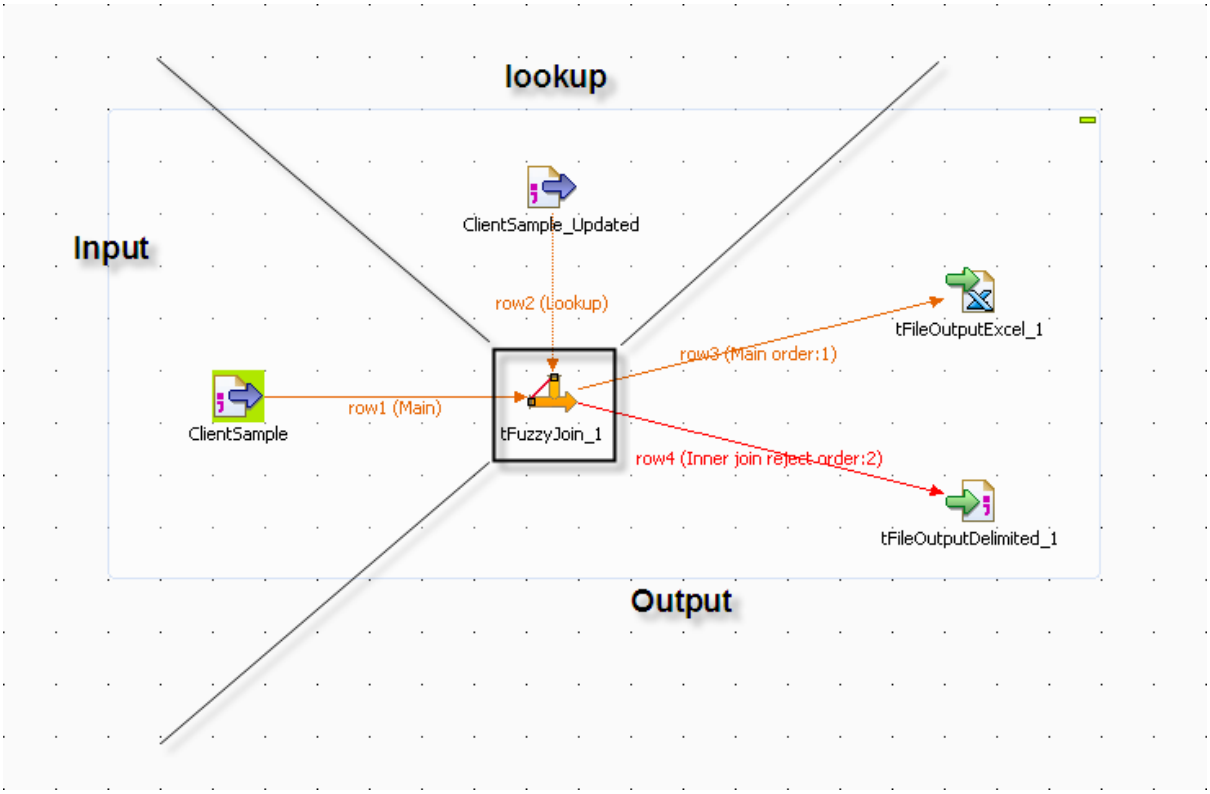
2.6. Layout

The placement of different types of components in a job design is very important to simplify understanding the process in the Job.

In a Job design, all types of components are placed in relation to the main flow component. Input components should be placed to the left, output components should be placed to the right, and a subflow component with a **Lookup** link should come higher above.

To assess the quality of the placement of different components in Jobs part of the audited project, *Talend Project Audit* analyzes misplaced input and output components and components with a **Lookup** links in subjobs.

The below figure illustrates an example of the placement of different components in a specific Job.



A full description of findings is then presented in a table.

2.6.1. Misplaced components

The table in this section lists numeric values for misplaced components used in subjobs:

Item	Description
Job	Name of Job
Subjob	Name of subjob used in the specified Job
Input component	Number of input components in the specified subjob that are not placed to the left
Lookup component	Number of subflow components with a **lookup** link that are not placed above the main flow component in the specified subjob
Output component	Number of output components in the specified subjob that are not placed to the right

Click the highlighted text (clickable text) in your Audit report to display the detailed referenced content without having to scroll down the pdf pages to find it.

2.7. Job analysis

To assess the quality and the importance of all Jobs used in the audited project, *Talend Project Audit* conducts a job analysis depending on certain synthetic elements, like number of components, schemas and triggers used in every Job of the audited project. The results of the investigation of these synthetic elements are weighted in their contribution to job overall importance.

Full description of findings is then presented in separate tables, one table per Job. These tables are listed in descending order: more important to less important.

2.7.1. Job synthetic view

Each table in this section lists numeric values on different synthetic elements of the investigated Job:

Item	Description
Indicator	Investigated synthetic elements can be one of the following types: **-Job rating** **-Component count** **-Schema column count** **-Context variable count** **-Job notes count** **-Misplaced input component** **-Misplaced lookup component** **-Misplaced output component** **-Trigger usage**
Value	Number of each of the above investigated elements in the investigated Job of the audited project

Click the highlighted text (clickable text) in your Audit report to display the detailed referenced content without having to scroll down the pdf pages to find it.