



# Talend Open Studio for Data Quality

Release Notes

## 5.2.1

## **Copyright**

This documentation is provided under the terms of the Creative Commons Public License (CCPL).

For more information about what you can and cannot do with this documentation in accordance with the CCPL, please read: <http://creativecommons.org/licenses/by-nc-sa/2.0/>

## **Notices**

All brands, product names, company names, trademarks and service marks are the properties of their respective owners.

---

## Table of Contents

|  |   |
|--|---|
| System Requirements .....                              | 1 |
| Data Quality: New Features .....                       | 2 |
| 1. Profiling perspective .....                         | 2 |
| 2. Indicators .....                                    | 2 |
| Data Quality: Bug Fixes / Change Log .....             | 3 |
| 1. Bug Fixes .....                                     | 3 |
| Data Quality: Known Issues and Known Limitations ..... | 4 |
| Documentation .....                                    | 5 |
| 1. New documents .....                                 | 5 |
| 2. Revised documents .....                             | 5 |
| 3. Known issues .....                                  | 5 |

# System Requirements

Talend Open Studio for Data Quality users should refer to <http://www.talend.com/docs/community/prerequisites.html> and [http://talendforge.org/wiki/doku.php?id=doc:installation\\_guide](http://talendforge.org/wiki/doku.php?id=doc:installation_guide) for more information on Installation and System Requirements.

# Data Quality: New Features

## 1. Profiling perspective

- Hive profiling: support of profiling data in a Hive database from the Profiling perspective. All types of data quality analyses are supported in Hive profiling except Time Correlation Analysis.
- Mass edit for JDBC connections: after migrating JDBC connections from different projects or different versions of the studio, some of these connections, basically to SQL Server 2005 and 2008, will have their driver path empty.

Users can set the driver path for several JDBC connections simultaneously and thus avoid the tedious work of setting the missing path manually for each of the migrated connections.

- Analyzing duplicates in the results of a business rule analysis: a ready-to-use analysis can be generated on the join results of an SQL business rule that is used to analyze database tables. The results of this analysis helps to better understand why there are more records in the join results than in the table.
- Defining the number of database connections in the analysis editor: this number can be set according to the database available resources, that is the number of concurrent connections each database can support.

## 2. Indicators

- Benford's law frequency indicator: this first-digit law system indicator is usually used to indicate accounting and expenses fraud in lists or tables through examining the actual frequency of the digits 1 through 9 in numerical data.
- Email Java user-defined indicator: this Java user-defined indicator can be imported from Talend Exchange via the Profiling perspective. It connects to the mail server and checks if the email exists.
- Indicator Selection window: numerous visual and usability improvements have been done to the Indicator Selection window. This window now allows for automatic scrolling and easier navigating of columns in large tables when selecting indicators. Users can sort columns in the window and view partially selected rows and columns.
- Merging indicators: system indicators that have been modified in a studio instance and then migrated into the current version will not overwrite the system indicators that are already in the studio. All modifications from older versions will be integrated in the system indicators.

# Data Quality: Bug Fixes / Change Log

## 1. Bug Fixes

In addition to the above new features a number of improvements within the entire product and significant bug fixes have been done — About 200 bug fix or improvements during the 5.2.1 release .

Please see the corresponding Change Log on our bug tracking system for more details on the individual issues:

- Talend Data Quality 5.2.1:

<https://jira.talendforge.org/secure/IssueNavigator.jspx?requestId=13941>

# Data Quality: Known Issues and Known Limitations

We encourage you to consult the JIRA bug tracking tool for a full list of open issues:

<https://jira.talendforge.org/secure/IssueNavigator.jspa?requestId=13940>

# Documentation

## 1. New documents

The following new documentation is available for this release.

- *Talend Big Data Studio User Guide*: this guide addresses only big data related features and capabilities offered in the big data studios.

## 2. Revised documents

In addition to updates to the content across the documentation set, the following specific documentation changes have been made.

- *Talend ESB Getting Started User Guide* has been renamed *Talend ESB Getting Started Guide*, and *Talend ESB Service Factory User Guide* has been renamed *Talend ESB Service Developer Guide*.
- Updates have been made to several *Talend ESB* guides to reflect changes in certain Apache products.
  - *Talend ESB Development Guide*: Apache CXF 2.7
  - *Talend ESB Mediation Developer Guide*: Apache Camel 2.10
  - *Talend ESB Service Developer Guide*: Apache CXF 2.7
  - *Talend ESB STS User Guide*: Apache CXF 2.7 - STS

## 3. Known issues

The documentation of certain technical fixes that arrived too late to be included in the current release will be dealt with in the next release.

<https://jira.talendforge.org/secure/IssueNavigator.jspa?requestId=13952>