|  |
| --- |
| , RD Dep. |
| BI.Talend\_Data\_Integration |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| REVISION HISTORY | | | | | |
| Ver. | Description of Change | Author | Date | Approved | |
| Name | Effective Date |
| 1 | Initial version | Sergey Moukavoztchik | 12/01/2018 |  |  |
|  |  |  |  |  |  |

Contents

[1. Preparation 3](#_Toc503856554)

[2. Tasks 3](#_Toc503856555)

# Preparation

1. Download and unarchive Talend Open Studio for Data Integration.
2. Create project
3. Created jobs and other repository elements should be Exported and pushed to separate repo at github.
4. Link to the repo need to be added to google sheet https://docs.google.com/spreadsheets/d/1PaA-xGj1087e0Uzd5vstPeCj8e198EyUO2j-3s\_9DK4/edit?usp=sharing
5. Optionally you can add text document with a description of your jobs.

# Tasks

1. Given several files with monthly increments named Payments\_yyyymmdd.csv. You need to implement the process to load monthly increment. It should distinguish which files from input folder was not loaded and process them. ETL will be tested by adding new set of files before rerunning a process.
2. Create subjob for loading file to the DB.
3. Subjob should be parametrized by date of file or retrieve a name of file, which need to be loaded. This design decision is up to you.
4. Consider that that file could contain data issues. Gather them in separate file.
5. Perform appropriate error handling.
6. Input folder and DB connection params should be set via contexts.
7. Put your DDL statements in separate SQL file, than I can reproduce your initial DB setup.

\*Note:

- To control which files was already loaded you can store last loaded date or names of loaded files in DB table and check with it to define which was already loaded and what is not.

- To retrieve list of dates from some point of time to present moment you can e.g. use sql query:

--consider that start date is 2017-01-01  
select add\_months(to\_date('20170101','yyyymmdd'), rownum-1)

from dual

connect by level < months\_between(current\_date, to\_date('20170101','yyyymmdd'));