|  |
| --- |
| , RD Dep. |
| BI.Talend Data Integration Exit Task |

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

Contents

[1. Preparation 3](#_Toc503489527)

[2. Tasks 3](#_Toc503489528)

# 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

* Several files provided customers.csv, products.csv, purchase\_yyyymmdd.csv (foreseen ability to load new incremental data from purchase\_yyyymmdd.csv). Your goal is to create dimensional model based on this data. Create dimensions at your choice. Add SK where needed.
* Design ETL process to load the data.
  + Load data files to the landing area at first. After that, ELT is preferred for further transformations and processing.
  + To populate the surrogate keys you can use DB sequences. Just add invocation of sequence’s .nextval function to your SQL query or ELTMap.
* Additionally calculate and add to a fact table KPI isInterState = 1 in case byer and warehouse State differ, or 0 if State is the same.