

ELN Dashboard - 3 - Technical - Data Preparation

- [High Level Project Architecture](#)
- [Architecture Data Flow](#)
 - [DataPrep Flow](#)
- [blocked URL Data Mapping](#)
- [Data Model](#)
- [Steps descriptions](#)
 - [Retrieve source system data and push them into GCS](#)
 - [Description :](#)
 - [Push data from GCS to Staging tables](#)
 - [Description :](#)
 - [Insert data from Staging into ODS tables](#)
 - [Description :](#)

High Level Project Architecture

Provide link here or the architecture diagram?

GCP Project

GCP project name: prj-dashb-eln-dev / prj-dashb-eln-test

Google Group: gcp-sl-dashb-eln-non-prod@solway.com

Admins (as it is in Dec 2022):

antoine.rey@solway.com (PO)

sebastien.ladouce@solway.com (Data Architect Industrial / R&I)

matteo.menghetti@solway.com (Data Engineer)

Architecture Data Flow

Here is a suggested template for Data Model + Data Mapping :

<https://docs.google.com/spreadsheets/d/1bD8AIGsNUI2sgANoOEKTuHBkxhsNVTD8cmOPYEIoLw>

Data Model

=> Document: https://app.genmymodel.com/api/projects/_k07o4lBOEe29ie0vpi-P5A/diagrams/_k07o4oBOEe29ie0vpi-P5A/svg

Data Mapping

=> Document : "[DA&AI - Data Architecture & Engineering - Data Knowledge - 11839 ELN / Data Mapping to DM](#)"

DataPrep Flow

Schema showing the different STEPS of the application flow - with the data involved at each step

[blocked URL](#)

Data Mapping

Link to data mapping :

Data Model

Data model schema : https://lucidapp.eu/lucidchart/f69303ec-ceee-479a-a02e-2ee4d64067de/edit?page=0_0&invitationId=inv_63cb8899-1c14-4e41-a8cf-ecfb081be65c#

Steps descriptions

Retrieve source system data and push them into GCS

Description :

Retrieve ELN and Success Factor data and upload them into GCS. More information on this step can be found [here](#) .

INFORMATION	DESCRIPTION
Tool	Talend
Access rights	ELN db : info saved into Keepass (path to keepass: \\ACEW1DTLNDENG02\Keepass) API Management Platform : Application id and secret saved in keepass (path to keepass: \\ACEW1DTLNDENG02\Keepass)
Source	ELN Oracle database and Success Factor data (retrieved via SAP API management platform)
Location	Data are uploaded into the GCS bucket
Format	Data is saved as csv files inside GCS
Sizing	Each extraction should upload less then 50 MB of data in GCS (files are temporarily stored on the remote engine and removed at the end of the run)
Assessment	The following job (GCS to Staging) will validate that data is consistent
Scheduling	Data is extracting every day
Criticality	Medium
Logging	Logs are stored in the Talend remote engine

Push data from GCS to Staging tables

Description :

Data that has been uploaded to buckets is then moved to staging tables. More information on this step can be found [here](#) .

INFORMATION	DESCRIPTION
Tool	Talend
Access rights	GCP service account :
Source	GCS bucket
Location	GCS bucket
Format	Data is saved as csv files inside GCS and added as Big Query table
Sizing	Each extraction should upload less then 50 MB of data in GBQ. This step is supposed to take less then 5 minutes to process all tables.
Assessment	logs table allows to determine if the upload has been successful.
Scheduling	Data is extracting every day at it is executing right after having uploaded the csv files to GCS
Criticality	Medium
Logging	Logs are stored in the Talend remote engine

Some mappings/filters are configured using some mapping tables. This is managed in this [Google Sheet](#) .

Insert data from Staging into ODS tables

Description :

Information is then inserted into ODS tables. More information on this step can be found [here](#) .

INFORMATION	DESCRIPTION
Tool	Talend
Access rights	GCP service account :
Source	GBQ Staging tables
Location	GBQ ODS tables
Format	Big Query tables

Sizing	Each extraction should upload less than 50 MB of data in GBQ. This step is supposed to take less than 5 minutes to process all tables.
Assessment	logs table allows to determine if the upload has been successful.
Scheduling	Data is extracting every day at it is executing right after having recreated the Staging tables
Criticality	Medium
Logging	Logs are stored in the Talend remote engine