IDBS Access-Rights

ELN API • Oracle Views • Datalab Talend GCP BigQuery

Project Name	Each ELN product follows a distinct IDBS access rights workflow
ETL	Talend
Status	Deployed in PROD

1. Project Overview

Objectives

The ELN system hosts multiple independent research projects. The Talend ETL flow enforces a project-level access policy: each collaborator is assigned to one or more specific ELN projects, and the pipeline only extracts and loads the data belonging to that collaborator's assigned projects.

Scope

- All active Talend jobs connecting to ELN API, Oracle Views, and DataLab.
- Access management rules governing ELN products assignment per collaborator.
- GCP BigQuery datasets and tables that serve as the final landing zone.

2. Architecture & Data Flow

The ETL pipeline follows a classic ETL pattern orchestrated by Talend. The high-level data flow is described below:

[ELN API]

[Oracle Views] [Talend ETL Jobs] [GCP BigQuery Tables]

[Datalab Platform]

Flow Steps

- Extract: Talend connects to each source (ELN REST API, Oracle JDBC, Datalab (GCP)) and extracts raw data according to the job configuration.
- Filter / Route: For ELN sources, the flow applies a project-based access filter, ensuring each collaborator's data is scoped to their assigned ELN project(s) only.
- Transform: Data is cleaned, typed, and mapped to the target GCP schema within the Talend job components.
- Load: Transformed data is written to the corresponding BigQuery dataset and table using the GCP BigQuery connector.

3. Source Systems

Source System	Type	Description
ELN API	REST API	Electronic Lab Notebook – exposes project data per collaborator access level
Oracle Views	Relational DB Views	Pre-aggregated relational data extracted via JDBC connections
Datalab Platform	Internal Data Platform	Analytical datasets and processed outputs from the Datalab environment

3.1 ELN API

Authentication:

(TEST)

- **REST / HTTP:** Protocol
- **Authentication Type :** Basic HTTP
- **URL:** https://mylab-test.syensqo.com:8443
- **Login:** (Contact Admin)
- **Password:** (Contact Admin)

(PROD)

- **REST / HTTP:** Protocol
- **Authentication Type :** Basic HTTP
- **URL:** https://mylab.syensqo.com:8443
- **Login:** (Contact Admin)
- **Password:** (Contact Admin)

Note :

- Get the list of container IDs from the Oracle view, then iterate over it as an input variable in the URL.
- Container IDs are should be available in this view :IDBS_EWB_SEC.HISTORY_ENTRIES.ENTITY_PATH

Exemple :

- https://mylab-test.syensqo.com:8443/ewb/services/1.0/feeds/entities/ (ContainerID_List)

3.2 Oracle DB

Authentication:

(TEST)

- IDBS_DB_Database = ELNTEST.eua.solvay.com
- IDBS_DB_Host = ACEW1TELNBDB02.eua.solvay.com
- IDBS_DB_Port = 1521
- IDBS_DB_UserName = (Contact Admin)
- IDBS_DB_Password = (Contact Admin)

(PROD)

- IDBS_DB_Database = ELNPROD.eua.solvay.com
- IDBS_DB_Host = ACEW1TELNBDB02.eua.solvay.com
- IDBS_DB_Port = 1521
- IDBS_DB_UserName = (Contact Admin)
- IDBS_DB_Password = (Contact Admin)

View:

- IDBS_EWB_SEC.HISTORY_ENTRIES.TARGET_ENTITY_ID AS Experiment_ID
- IDBS_EWB_SEC.HISTORY_ENTRIES.ENTITY_PATH AS FOLDERPATH
- IDBS_EWB_SOLVAY_CUSTOM.DM_UserAccessPath_Full AS PERMISSIONS
- DBS_EWB_SOLVAY_CUSTOM.DM_EntityRolePermissions_Full AS ROLES

Note:

- The Oracle database connection is enabled for Talend Remote Engines. However, when connecting to the database via a VDI, the connection may be disabled.

3.3 Datalab

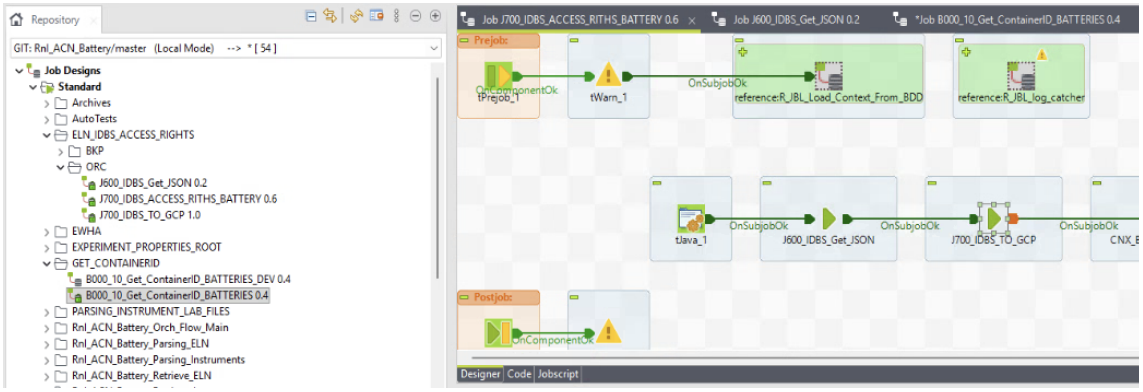
- **Source:** GCP Bigquery table
- **Project:** gcp-sqo-datalab-(*)
- **Dataset:** bq_ds_datagrow_dev_ads_static
- **Table:** application_use
- **filter:** user_status='active' AND ROLE IN ('admin', 'lab_manager')

3.4 Target table:

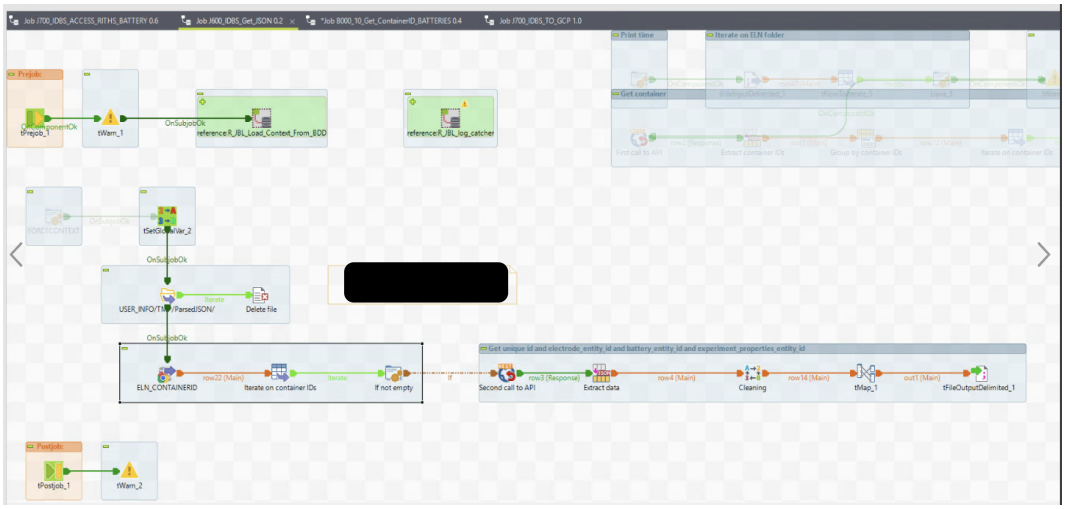
Bigquery: DM.ELN_IDBS_AccessRights

Talend tasks:

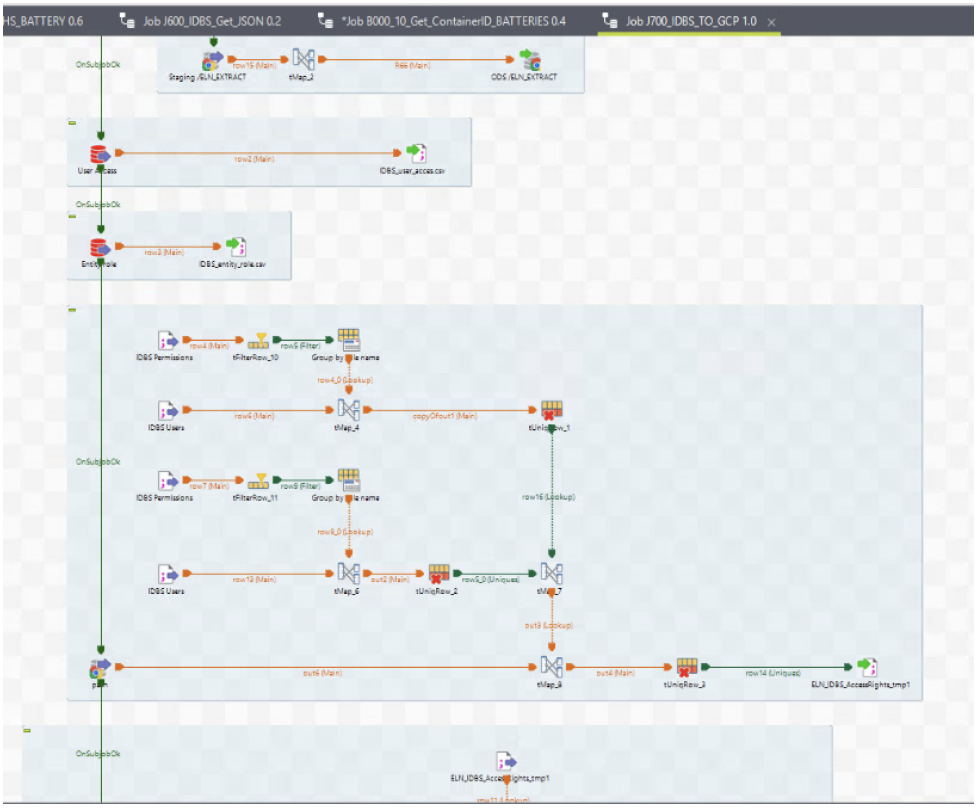
(Master)



(Get Experiments)



(Cross experiments with collaborators and roles)



Error Handling Strategy

- All jobs implement a tDie component on the error branch to terminate gracefully and log a meaningful message.
- API connectivity failures trigger a tWarn followed by a retry loop.
- Job execution logs are visible in TMC.

Scheduling

- ELN GCP job: Daily at 02:00 AM (server local time)
- Oracle GCP job: Daily at 02:00 AM
- Datalab GCP job: Daily at 02:00 AM

4. Contacts & responsibilities:

- antoine.rey@syensqo.com / ELN Administrator
- prasanth.gnanasekar@syensqo.com / Data Engineering - Flow Maintenance

5. Annexe document:

This annex document could be helpful if you are looking for more details on access rights related to ELN. (link)