ALB Data Dev for Conductivity

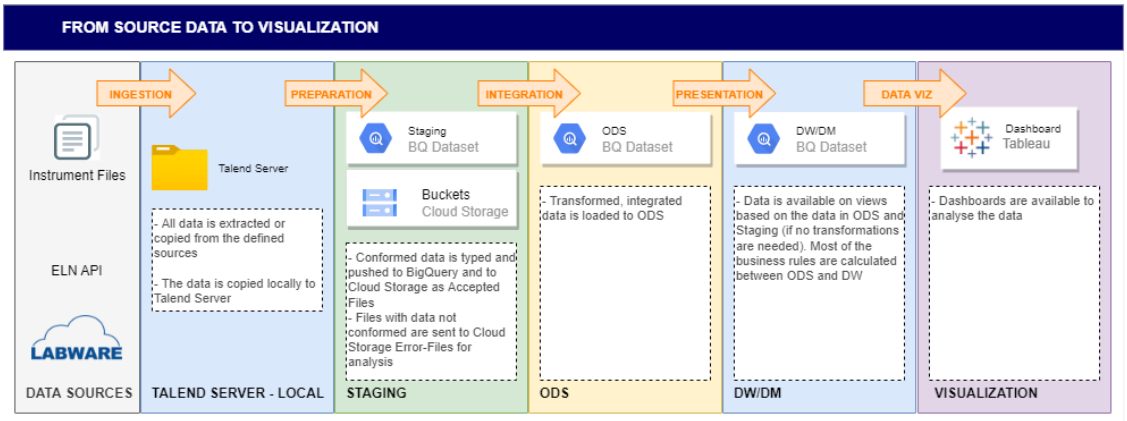
This page presents the data development documentation for Battery Conductivity for all sites.

This page was copied from Battery. Needs to be updated

The standard ETL areas and processes are described on ALB Data Dev Architecture - General.

Summary

Data flow diagram



*No Labware for Batteries Conductivity

All the schemas are available on google drive for further edition:

Data Ingestion

Data Sources

ELN

Simplified Flow

ELN API <Extraction> Local JSON File

Spreadsheets

Spreadsheet	Description	Data Status*
ELN Conductivity Measurement	This is a standard ELN spreadsheet for **all sites**	HOT
ELN Battery Experiment Properties	This is a standard ELN spreadsheet for **all sites**	HOT

*Data Status: HOT: Data currently being updated at the source. It should be loaded regularly. COLD: No data changes at the source. It should be loaded just once.

The list of the spreadsheets/tables extracted from JSON files coming from ELN can be found in the Data Mapping of the next section.

Related documents

Document Name	Link
Battery - ELN Data Model	
Battery - ELN Template Documentation	

Instruments

Simplified Flow

Lab server Google Share Drive <Copy> Local XML/TXT file

Files

The instrument files are, most of them, **manually** added to a Google Drive Shared folder

Site	Description	Status
Seoul		HOT
Brussels		HOT
Bollate		HOT
Aubervilliers		HOT

*Data Status: HOT: Data currently being updated at the source. It should be loaded regularly. COLD: No data changes at the source. It should be loaded just once.

Data Preparation or Parsing

Simplified Flow

Local XML/TXT file <Load> Cloud Storage/Staging BigQuery

Data Mapping Source => Staging (Talend)

The spreadsheet below presents all data transformation between the raw files (extracted files) and a BigQuery Staging table.

Data Integration or Computing

The data integration phase for batteries follows the standard approach describe on ALB Data Dev Architecture - General.

Simplified Flow

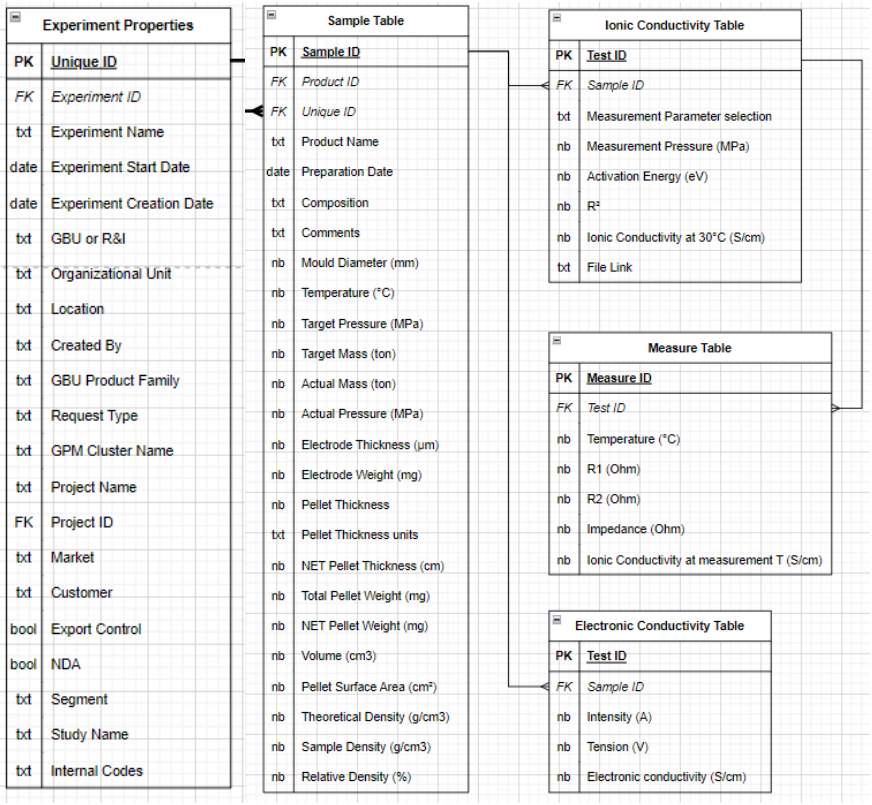
Staging BigQuery <Transform> **ODS** BigQuery

Data Mapping Staging => ODS (BigQuery SQL views)

The spreadsheet below presents all data transformations from Staging tables to ODS tables. This steps aims to structure the files in the target table format and checking if the column's type (schema) are conformed.

Data Model

The following data model presents the tables present on ODS dataset and the relation between them:



Data Presentation (DW/DM)

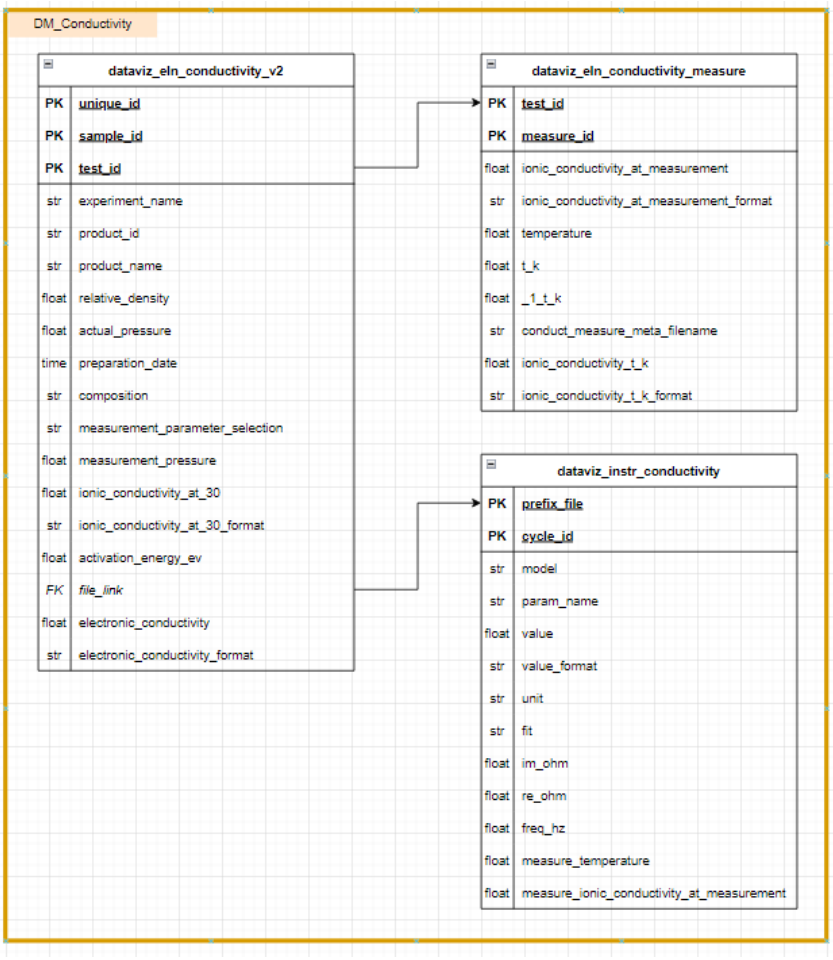
> **Simplified Flow**
>
> **ODS** BigQuery <Transform> **DW** BigQuery <Expose> **DM_Conductivity** BigQuery

Data Mapping ODS => DW/DM (BigQuery SQL views)

The spreadsheet below presents all data transformations between the tables on ODS and DM_Conductivity. This steps aims to create views for the Data Visualization.

Data Model

The following data model presents the tables present on DW/DM dataset and the relation between them:



For the DM dataset (DM_Conductivity), we have the views on the top of the tables before. As defined in the convention, there is no need of the abbreviation "conduct". No need either of Talend jobs.

Orchestrating Jobs

All the jobs are run in sequence under the follow job and project name on Talend Cloud:

Project	Job/Flow	Associated TMC Plan
RnI_ACN_Battery	F010_RnI_ACN_Battery_ELN_IDBS_Orch_Flow	PL_RNI_ACN_BATTERY_CONDUCTIVITY_ELN_DAILY
RnI_ACN_Battery	F020_RnI_ACN_Battery_ELN_Integration_Orch_Flow	PL_RNI_ACN_BATTERY_CONDUCTIVITY_ELN_DAILY
RnI_ACN_Battery	F011_RnI_ACN_Battery_Instruments_Orch_Flow	PL_RNI_ACN_BATTERY_CONDUCTIVITY_INSTRUMENT_DAILY
RnI_ACN_Battery	F021_RnI_ACN_Battery_Instr_Integration_Orch_Flow	PL_RNI_ACN_BATTERY_CONDUCTIVITY_INSTRUMENT_DAILY

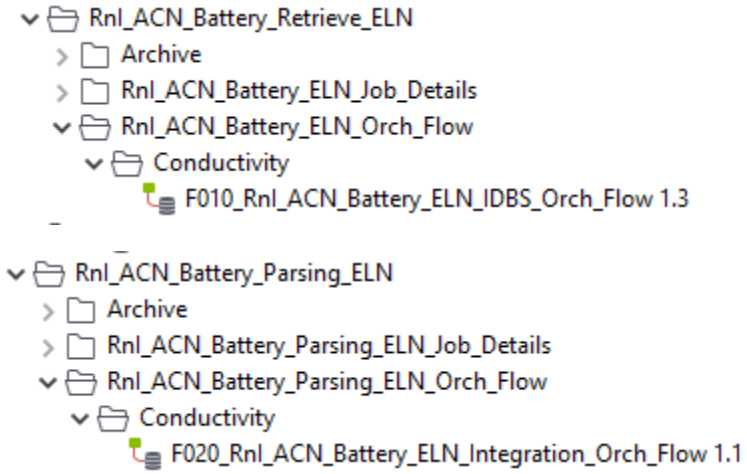
For scheduling details check the Operational documentation.

The following jobs should not be orchestrated and only run once during the deployment :

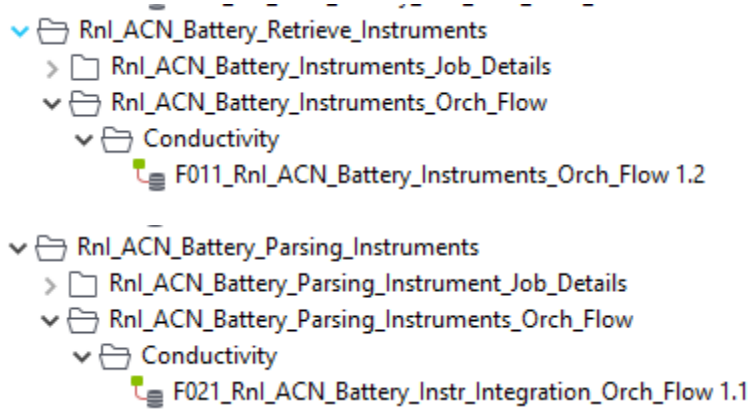
Project	Job/Flow	Associated TMC Plan
RnI_ACN_Battery	F010_RnI_ACN_Battery_Create_BQ_Views	PL_RNI_ACN_BATTERY_CONDUCTIVITY_CREATE_VIEWS

Talend

ELN

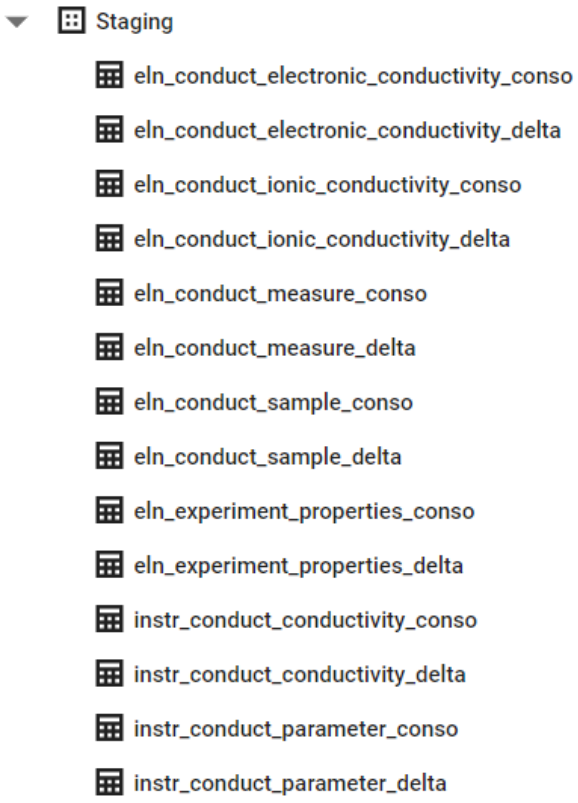


Instruments

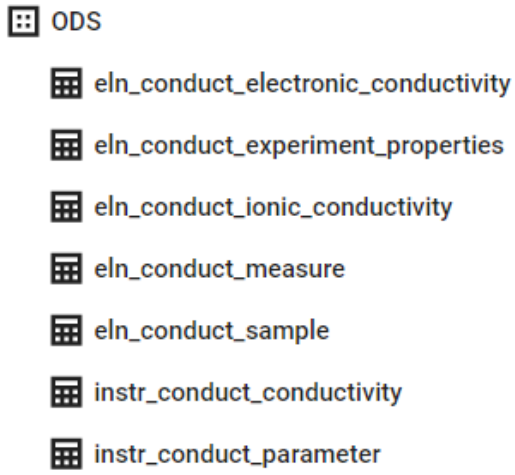


Big Query

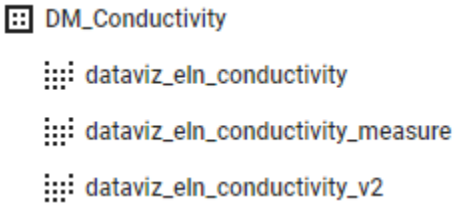
Tables (Staging)



Views(ODS)



Views(DM_Conductivity)



Data Visualization

Conductivity (GCP and Tableau documentation) :