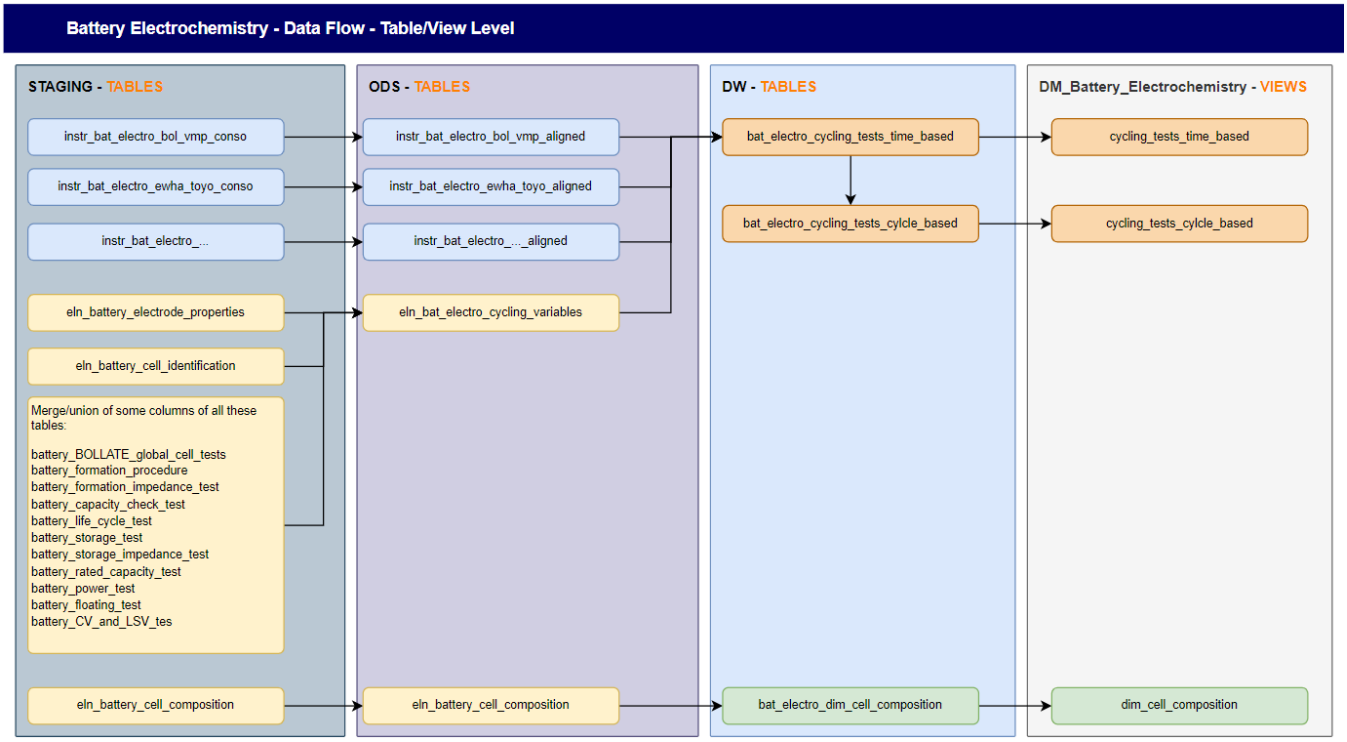
ALB Data Dev for Battery Electrochemistry

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

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

Summary

Data flow diagram



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

Source	Spreadsheet	Description	Data Status*
ELN EWHA BOLLATE	ELN current template	This is a standard ELN spreadsheet for **all sites**	HOT
ELN BOLLATE	ELN historic template	This is a ancient format for ELN spreadsheet for BOLLATE. The data is still necessary but new spreadsheet won't be created.	HOT
ELN Battery mapping	ELN historic template	This is a ancient format for ELN spreadsheet for EWHA. The data is still necessary but new spreadsheet won't be created.	HOT
ELN MuSigma	pré-ELN	This spreadsheet has EWHA data before starting using ELN. There is no new data being added to it. All data is already ingested.	COLD
ELN Separator data Mapping	?	Existing spreadsheet to be loaded to BQ	?

*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	https://app.diagrams.net/#G1sD6OqKnBzSR_SrGvzlhEl7s5F5vUGQqD
Battery - ELN Template Documentation	

Instruments/Cyclers

Simplified Flow

Lab server <Copy> Local CSV/XML/TXT file

Files

Cyclers	Site	Description	Status
Maccor	EWHA		HOT
Maccor	Bollate		HOT
BioLogics	EWHA		HOT
BioLogics BCS	Bollate		HOT
BioLogics VMP3	Bollate		HOT
Arbin	EWHA		HOT
Arbin OLD	Bollate		COLD
Arbin (LPT)	Bollate		HOT
PNE	EWHA		HOT
Toyo	EWHA		HOT
Biologic SP150 & SP300 (VSP)	Bollate		HOT
BioLogics BCS	AUB		HOT

BioLogics BCS	BRU		HOT
BioLogics VMP	AUB		HOT
BioLogics VMP	BRU		HOT
BioLogics MPG	AUB		HOT
BioLogics MPG	BRU		HOT
BioLogics SP	AUB		HOT
BioLogics MTZ	AUB		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.

SFTP/File Location - Aubervilliers and NOH data

The instrument files are, most of them, **manually** added to a Google Driver's folder or a folder in lab servers.

The spreadsheet **Where to find instrument files?** has the full list of instrument folders and location:

SFTP/File Location - EWHA

https://docs.google.com/document/d/12FtHrijv22d1IM8FPXrYxx8nQVM4UeZlW8-ges3lYUQ/edit#

SFTP/File Location - Bollate

Data Mapping

No data mapping needed as the data is copied exactly as the source.

Data Preparation or Parsing

Simplified Flow

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

Data Mapping

The spreadsheet below presents all data transformation between the raw files (extracted files) and a BigQuery delta table. Some files are unstructured and semi-structured. This steps aims to structure the files in the target table format and checking if the column's type (schema) are conformed expected.

(old one)

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

The spreadsheet below presents all data transformations between the tables on Staging and ODS. This steps aims to add some calculations and intelligence to the data. Not all tables will need passing through this step.

Raw Data Pairing (all sites)

This spreadsheet shows how to align all cyclers in just one commun table schema:

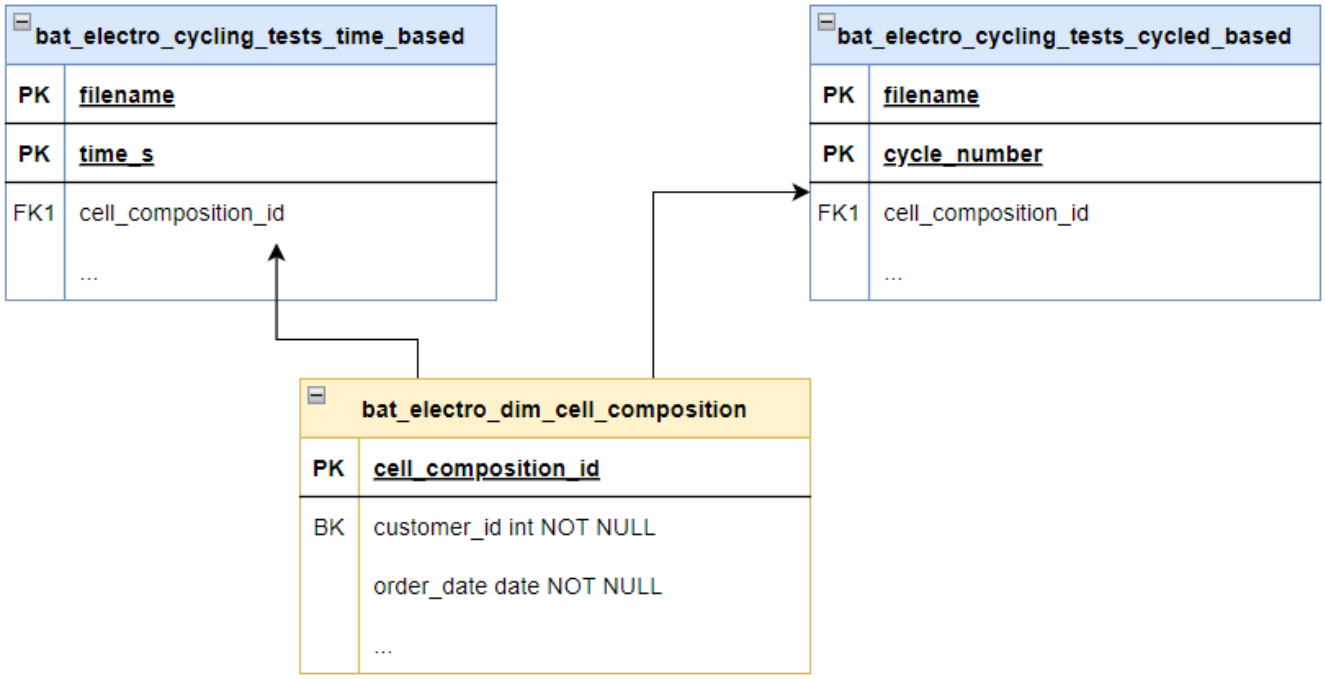
Data Presentation (DW/DM)

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

Data Mapping

Data Model

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



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

Orchestrating Jobs

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

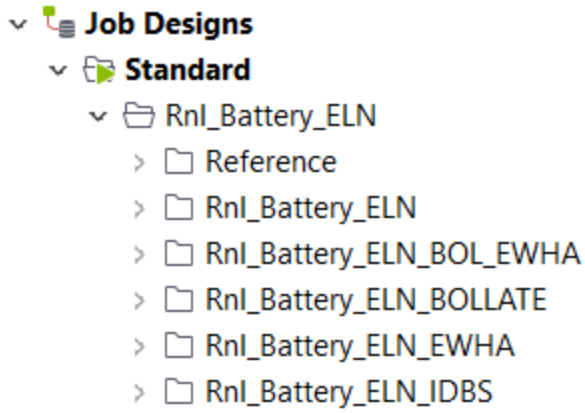
Project	Job/Flow
RnI_ACN_Battery	F000_RnI_ACN_Battery_
RnI_Battery_ELN	
RnI_Battery	
RnI_Battery_BOLLATE	

For scheduling details check the Operational documentation.

Talend

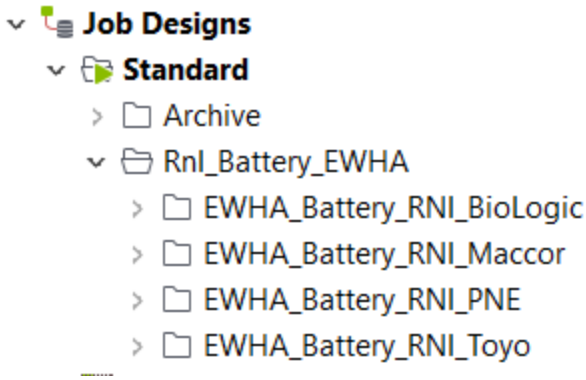
ELN

Project: RnI_Battery_ELN



Instruments

Project: RnI_Battery (EWHA)



Project: RnI_Battery_BOLLATE