

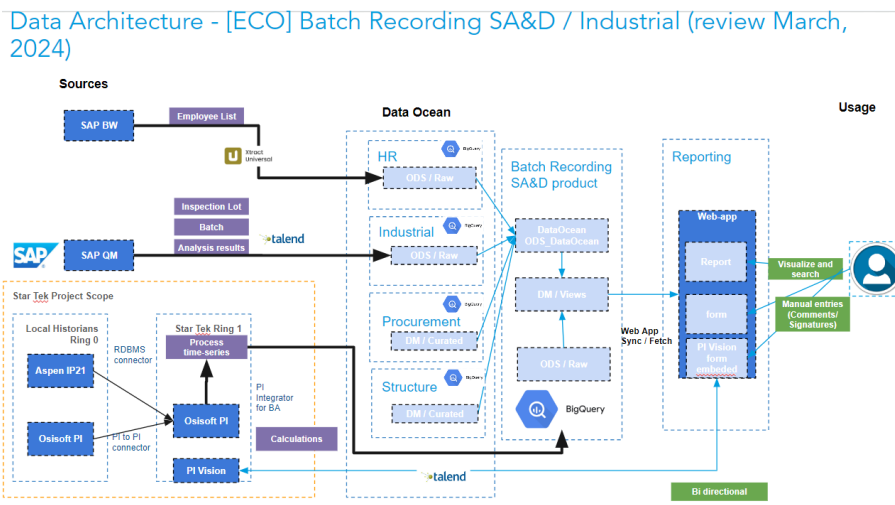
Electronic Batch - 4 - Analytics Pipeline

- Architecture Data Flow
 - Overall Architecture
- Steps descriptions
 - Electronic Batch
 - Description
 - Tools
 - Access rights
 - Validation process
 - Configuration
 - Source locations
 - Source formats
 - Destination locations
 - Destination formats
 - Sizing
 - Scheduling
 - Timing
 - Criticality
 - Logging

Architecture Data Flow

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

Overall Architecture



Steps descriptions

The scopes included this project are the extractions from SAP BW, SAP QM, and Star Tek.

The logic of the data exposure to the web application is done through Views within Google BigQuery.

Electronic Batch

Description

Collect the Industrial data into a centralized data repository / Single Source of Truth that enhances data quality to expose to Web Application for End Users

Tools

Talend is used for the orchestration of all extractions and loading of the data into Google BigQuery.

Access rights

Talend Jobs require valid GCP service account credentials to load data into Google Cloud Storage and Google Big Query.

The talend jobs must have valid credentials for accessing SAP and Star Tek. Refer to the technical documentation.

Validation process

How to validate that this output match business expectations

Configuration

which kind of information or configuration this step uses ?

Source locations

*There are two principle data sources for Electronic Batch Data
SAP QM (Tables related to Batch & Quality)
Star Tek (Technical data related to batches)*

*SAP BW (HR Data related to users and roles at GBU)
Data Ocean Domains
- Structure : Materials and Material Types
- Procurement : Commercial Products*

Source formats

The format of the source datasets

Destination locations

Extracted data is physically stored in Google BigQuery in the EU multi region for all projects.

For new data being extracted from SAP QM, it is stored in the GCP projects for the industrial domains (prj-data-dm-industrial-(dev/test/ppd/prod) and exposed to the GCP ebatch projects through Views.

For new data being extracted from Star Tek for GBU sites, it is stored in the GCP projects dedicated for ebatch (prj-data-sad-ebatch-(dev/test/ppd/prod)

Destination formats

While the extractions of tables from SAP QM is to tables within BigQuery, the data exposure is managed through Views in the project to be consumed by the Ebatch Web Application.

Sizing

Expected data volume

Scheduling

Is there an automatic schedule ?

Extractions are programmed through Talend

At what frequency ? What is the trigger ?

Hourly extractions for all tables.

Timing

The average time expected for :

- *Approximately 10 minutes for all jobs*

Criticality

High / Medium / Low

Logging

Logs are available through Talend, and through the respective GCP projects used in the project.