

Materials Data Flow

Talend Data Flow Documentation

Project: Specialty Polymers

Source: Labwer (Shared Folders on NAT) Target: GCP BigQuery Tables

Data Engineering Team

1. Project Overview

This document provides the technical documentation for the Talend data pipeline implemented under the Specialty Polymers project. It covers the end-to-end data flow from the source system (Labwer shared folders deployed on a NAT environment) through to the target destination GCP BigQuery tables. It also describes the compression strategy applied by the Product Owner to manage large data volumes efficiently.

2. Data Flow Architecture

2.1 High-Level Flow

The pipeline follows a standard Extract Transform Load (ETL) pattern, enriched with a compression/decompression layer to handle the large data volumes characteristic of this project:



Source :

Category	Item / Source	Details / Target
Oracle Sources	labw-p-oracle-01.syensqo.com	
Oracle Sources	labw-q-oracle-01.syensqo.com	
Bollate File Paths	Source Directory	\\TBOLVRS06T\Lab Booster
Bollate File Paths	Example File Note that .txt files may contain complex and irregular structures, which can make parsing challenging.	\\10.53.6.10\labo\W-524600\TGA\DA CANC\23-11194-6715351-tga-residuo da acque 965pi plx485- aria - sciarrillo.txt
Alpharetta File Paths	Test Files Directory	\\USALPACDv02\Test Files\LabBooster

Talend Extraction & Tmp shared folder:

Category	Item / Source	Details / Target
Talend / Jobs	Orchestration Job	F730_Thermal_Data_Compression_Orch
Talend / Jobs	Sub-Job 1	J125_instrument_Raw_Data_Compressed_To_Bigquery
Talend / Jobs	Sub-Job 2	J125_Thermal_Raw_Data_Compressed_To_Bigquery

Talend / Jobs	SQL Queries Path	V:\PROD\Rn\ACN_Materials\SQLQueries
Local Storage	Temp Compression Folder	Z:\(ENV)\Rn\ACN_Materials\tmp\Working\data_compression

GCP (Bigquery & GCS)

Category	Item / Source	Details / Target
GCP Targets	Staging (Raw) - Delta	gcp-sqo-labbooster-materials-d.Staging.compressed_thermal_raw_data_delta
GCP Targets	Staging (Raw) - Conso	gcp-sqo-labbooster-materials-d.Staging.compressed_thermal_raw_data_conso
GCP Targets	ODS (compressed) - Delta	`gcp-sqo-labbooster-materials-p.ODS.compressed_raw_data_delta`
GCP Targets	ODS (compressed) - Conso	`gcp-sqo-labbooster-materials-p.ODS.compressed_raw_data_delta`
GCP Targets	ODS (compressed) - Delta	`gcp-sqo-labbooster-materials-p.ODS.compressed_thermal_raw_data_conso`
GCP Targets	ODS (compressed) - Conso	`gcp-sqo-labbooster-materials-p.ODS.compressed_thermal_raw_data_delta`
GCP Targets	ODS (Compressed + Filtred data)	`gcp-sqo-labbooster-materials-p.ODS.summary_results_conso`
GCP Targets	ODS Raw delta Table	`gcp-sqo-labbooster-materials-p.ODS.raw_data_delta`
GCP Targets	DM- Vue - Used as source for Tableau Software	`gcp-sqo-labbooster-materials-p.DM.summary_results`
GCP Targets	EMAIL - ALERT TABLE	`gcp-sqo-labbooster-materials-p.DM.ALERT_ERRORS_INSTRUMENTS_FILES`

Reporting:

Category	Item / Source	Details / Target
Reporting	Tableau Link	https://eu-west-1a.online.tableau.com/#/site/syensqo/views/MaterialsThermalv1_4DSC-TGA/MaterialThermalDSCdev

3. Alerts:

A dedicated Talend job is responsible for validating input source files before processing.

If a file is detected as corrupted or fails validation checks, the job automatically triggers an alert notification. This alert is sent to the relevant stakeholders to ensure prompt awareness and intervention.

Notifications are delivered via the SMTP server, enabling email-based alerts to communicate issues in near real time.

4. Contacts & responsibilities:

- prasanth.gnanasekar@syensqo.com / Data Engineering - Flow Maintenance