

DIO

- Description
- Tools: Talend
 - From source to ODS
 - From ODS to DM
 - From DM to Operational Dashboard
 - Access rights
- Source
- Destination
 - DataOcean
 - Product
 - Format
 - Sizing
- Loading
 - 1.1 Full load
 - 1.2. Reloading data
 - 1.3 Plan to schedule
 - 1.4 Timing
 - 1.5 Pre Requisites
- Criticality
- Logging

Description

The data source getting from prj-data-dm-industrial, which get from prj-data-dm-industrial-test.EDC.
V_EDC_monthly_inventory_performance_data_finance_dio

tables in order to get the information of DIO.

Tools: Talend

Talend project DATA_OCEAN_DOMAIN_INDUSTRIAL

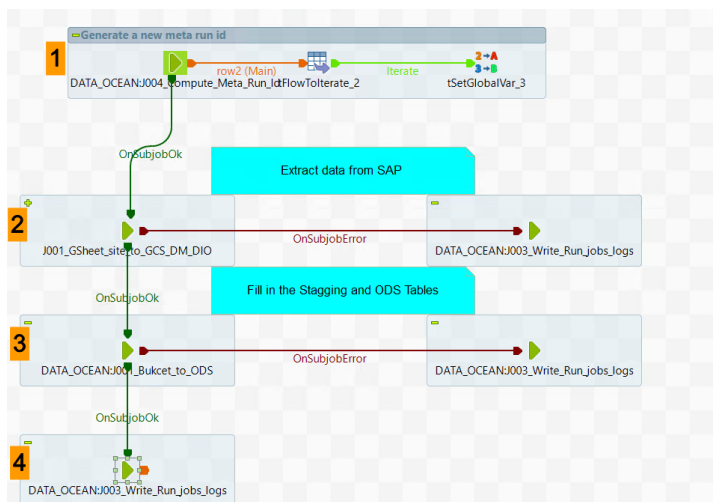
- F001_GSheet_site_to_BQ_DM_DIO

From source to ODS

Job:

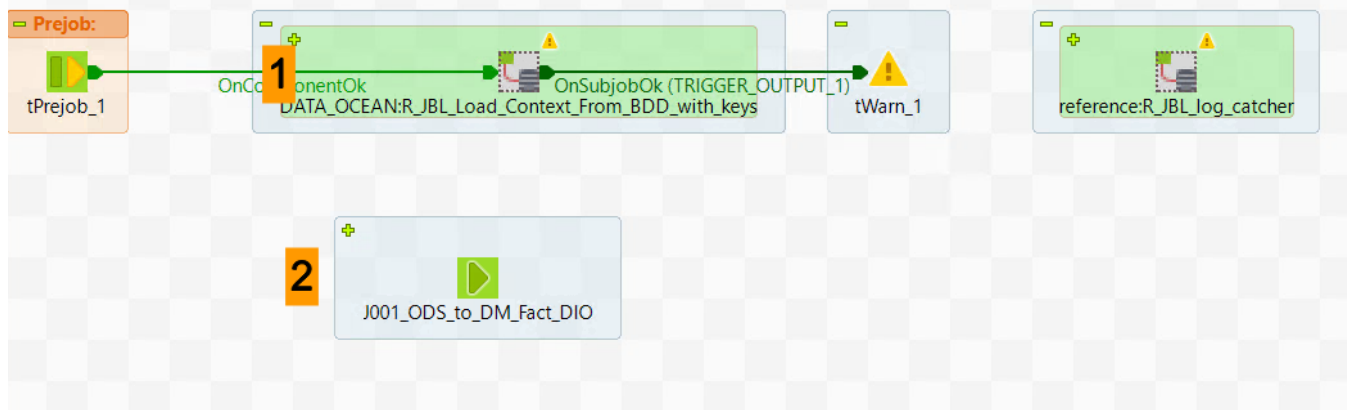
- F001_GSheet_site_to_BQ_DM_DIO

(DIO)



1. Define variable and Generate meta_run_id
2. Call reference job to load from EDC.V_EDC_monthly_inventory_performance_data_finance_dio` and save the output file to bucket cs-ew1-prj-data-dm-industrial-test-staging/dio
3. Load from Bucket to STG and ODS
4. Update log

From ODS to DM



1. Load context from the RDS to be used in talend
2. Load data from ODS to DM

From DM to Operational Dashboard

Dataflow

prj-data-dm-industrial-test.EDC.V_EDC_monthly_inventory_performance_data_finance_dio prj-data-dm-industrial-test.STG.
STG_FIL_0000_0000_F001_F_M_dio prj-data-dm-industrial-test.ODS.ODS_FIL_0000_F001_F_M_dio prj-data-industrial-dash-test.ODS_DataOcean.
V_ODS_dio prj-data-industrial-dash-test.DM.FACT_dio prj-data-industrial-dash-test.DPL.V_FACT_dio

Access rights

It is required to access prj-data-industrial-dash-test project and prj-data-dm-industrial-test project.

Source

BigQuery

Project = prj-data-dm-industrial-test

BQ Dataset = EDC

BQ View = V_EDC_monthly_inventory_performance_data_finance_dio

Destination

DataOcean

- Bucket = cs-ew1-prj-data-dm-industrial-test-staging/
 - dio
 - FIL_IND_0000_0000_F001_20250314122333_0000_F_M_dio.csv
- STG Table names = STG_FIL_0000_0000_F001_F_M_dio
- ODS Table names = ODS_FIL_0000_F001_F_M_dio

Product

- GCP = prj-data-industrial-dash-[env]
- DataOcean
 - V_FACT_dio

Format

columnar format

Sizing

- STG_FIL_0000_0000_F001_F_M_fc around 17,392 records

Loading

1.1 Full load

- **Dio** Source to ODS : [PL_DM_IND_Dio](#)

1.2. Reloading data

Dio

- Just run the [PL_DM_IND_Dio](#) .

1.3 Plan to schedule

It is scheduled by plans below on WS_DATA_OCEAN_DOMAIN_INDUSTRIAL

- [PL_DM_IND_Dio](#) - Every 30 minutes and start at 06:00am UTC until 06:00pm

1.4 Timing

- Case 5 minutes from source to ODS (full)

1.5 Pre Requisites

3 tables are created from a Gsheet.

A Scheduled Query is then executed in BigQuery, which basically pivots various information from the tables and transforms them into a tabular view. From there, the input source is created to be consumed in Talend.

For more details, access the documentation below:

- [M2P Indus Dash - Documentos Google](#)

Scheduled Query Explanation

Overview

This SQL script automates the validation and materialization of multiple datasets in **BigQuery** . It follows a structured approach to:

1. **Validate the source data** – Ensuring the dataset is accessible and contains valid data.
2. **Drop and recreate materialized tables** – Removing outdated tables and creating new ones with additional metadata.
3. **Enhance traceability** – Adding metadata fields such as row identifiers, timestamps, and Google Sheets references.

The script processes three datasets separately:

- **Wave Project Star** (wave_project_star)
- **Wave Activity Star** (wave_activity_star)
- **Monthly Inventory Performance - Data Finance** (monthly_inventory_performance_data_finance)

Each dataset follows the same process, ensuring reliability and error handling.

Criticality

- Low?

Logging

Check the loading records

```
select job.job_name , job.meta_start_date , job.meta_execution_id , logs.meta_run_id , logs.meta_source_system , logs.meta_step , logs.meta_status , logs.meta_num_lines , logs.meta_error_lines from STG.log_tables logs join STG.run_jobs job on logs.meta_run_id = job.meta_run_id
where job_name like '%DIO%'
order by job.meta_start_date desc
```