

# Wave

- Description
- Wave Connection
  - How to create connection
  - How to change the access time
  - How to reset password
- Talend
  - Detail job
  - Flow job
  - Access rights
- Source
  - Format
- Destination
  - Data Ocean
  - Product
  - Format
  - Sizing
  - Assessment
- Loading
  - 1.1 Incremental Load
    - Source to ODS
    - ODS to DM
  - 1.2 Full load
  - 1.3. Reloading data
    - Source to ODS
    - ODS to DM
  - 1.4 Plan to schedule
  - 1.5 Timing
- Criticality
- Logging

## Description

Wave is an external tool that calculates the cost of saving on each project. Talend can get the data via API of this server and the output of the API is an Excel file.

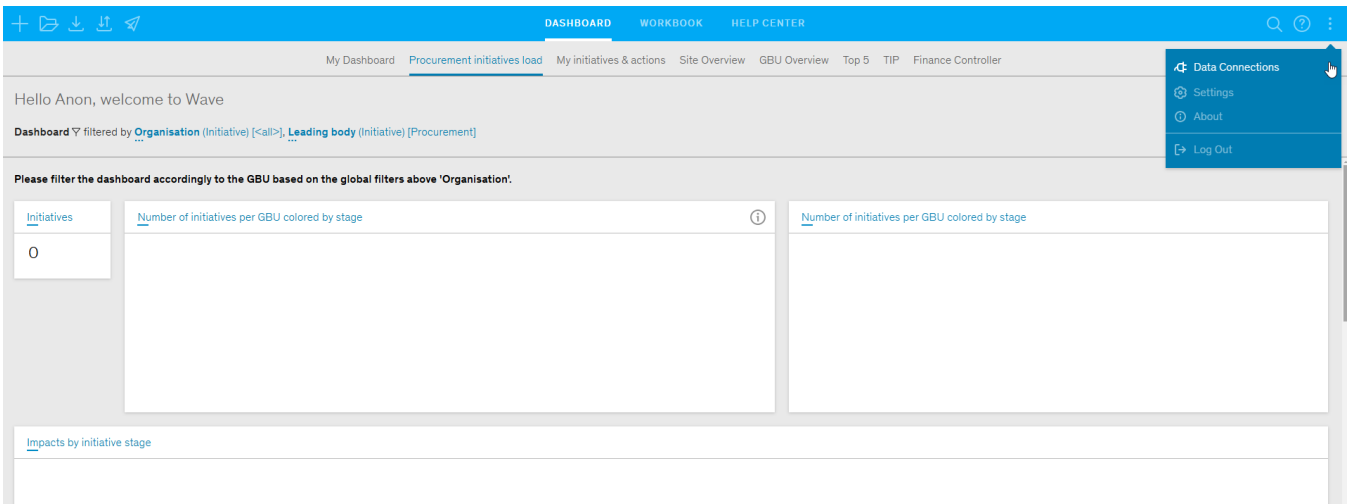
It is required to create a connection and set up the access time to allow Talend to read the data. If Talend accesses outside of the allowed time, it will not get the file.

The password is also required to be reset every 6 months since the password will expire.

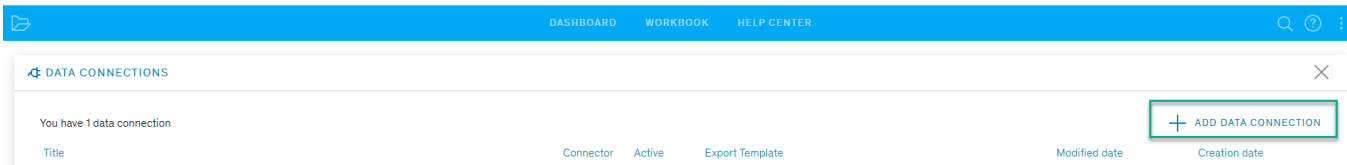
## Wave Connection

### How to create connection

1. Login <https://solvay.mckinseywave.com/login/auth>
2. Go to Data Connection



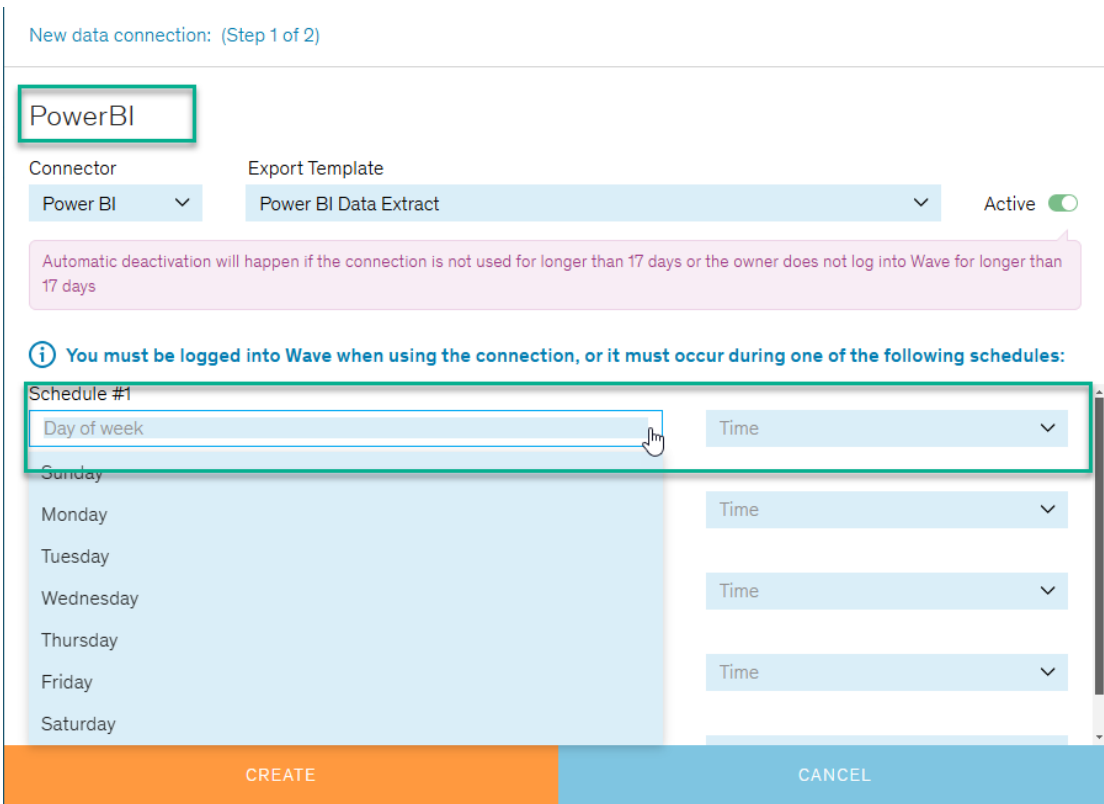
3. Click on ADD DATA CONNECTION



4. Enter the name of the connection and data/time that allow to access.

If not set schedule, it is required to login on the web in order to allow the access.

The time is GMT+1 time zone



5. It will generate URL endpoints, username and password. Click on generate password and save it.

Please note that each password will be expired every 6 months and the interface is required to access at least every 17 days, otherwise, the connection will be deactivated.

## Connect to Wave: PowerBI\_Template

Follow these steps to allow Power BI to pull data from Wave

1. Select "Get Data" and choose "Web" in Power BI Desktop
2. Copy the URL below

```
https://5cc98d328b7f.cdn.mckinseywave.com/api/data-connection/dff52a23-e5f1-43e1-8cab-d361777b497a/export
```



3. Paste the above URL into the URL field and click OK
4. Choose "Basic" in the "Access Web content" popup
5. Copy the email address below

```
talend_admin_dev@solvay.com
```



6. Paste the above email address into the user name field
7. Enter the previously generated password into the password field
8. If you need to reset the password then click the button below

GENERATE PASSWORD

9. Click "Connect". It may take several minutes to retrieve a large data export
10. You must be logged into Wave or be within the defined schedule for data connection calls to succeed

CLOSE (YOU CAN RETURN HERE LATER)

## How to change the access time

1. Login <https://solvay.mckinseywave.com/login/auth>
2. Go to Data Connections same as create connection
3. Right click on the connection that you want to change and select Update

The screenshot shows the 'DATA CONNECTIONS' interface. At the top, it says 'You have 1 data connection' and there is a '+ ADD DATA CONNECTION' button. Below this is a table with the following columns: Title, Connector, Active, Export Template, Modified date, and Creation date. The table contains one row for 'PowerBI' with the following details: Connector: Power BI, Active: On, Export Template: Power BI Data Extract, Modified date: 18-Nov-2024 5:06 AM, and Creation date: 16-Oct-2024 11:18 AM. A context menu is open over the 'PowerBI' row, showing options: Connect, Update (highlighted with a mouse cursor), View activity, Clear cache, and Delete.

Title	Connector	Active	Export Template	Modified date	Creation date
PowerBI	Power BI	On	Power BI Data Extract	18-Nov-2024 5:06 AM	16-Oct-2024 11:18 AM

4. Change date/time and click UPDATE

Update data connection: PowerBI

### PowerBI

Connector: Power BI | Export Template: Power BI Data Extract | Active:

**ⓘ You must be logged into Wave when using the connection, or it must occur during one of the following schedules:**

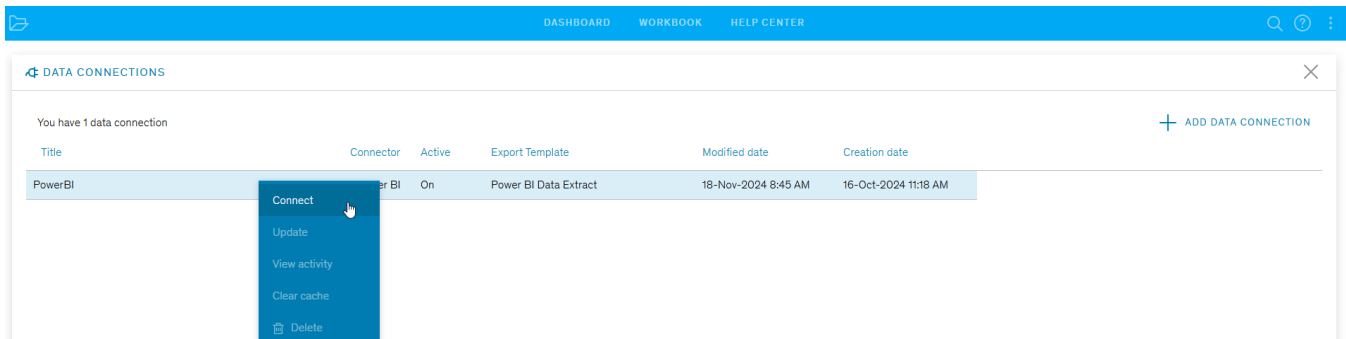
Schedule #1	Friday	12:00 PM - 01:00 PM Europe/Rome
Schedule #2	Monday	05:00 AM - 06:00 AM Europe/Rome
Schedule #3	Day of week	Time
Schedule #4	Day of week	Time
Schedule #5	Day of week	Time

**UPDATE** **CANCEL**

5. In case of the connector is deactivated, just mark it active here and keep login to the Wave in order to avoid deactivate.

## How to reset password

1. Login <https://solvay.mckinseywave.com/login/auth>
2. Go to Data Connections same as create connection
3. Right click on the connection that you want to change and select Connect



4. Click on reset password

## Connect to Wave: PowerBI\_Template

Follow these steps to allow Power BI to pull data from Wave

1. Select "Get Data" and choose "Web" in Power BI Desktop
2. Copy the URL below

```
https://5cc98d328b7f.cdn.mckinseywave.com/api/data-connection/dff52a23-e5f1-43e1-8cab-d361777b497a/export
```



3. Paste the above URL into the URL field and click OK
4. Choose "Basic" in the "Access Web content" popup
5. Copy the email address below

```
talend_admin_dev@solvay.com
```



6. Paste the above email address into the user name field
7. Enter the previously generated password into the password field
8. If you need to reset the password then click the button below

GENERATE PASSWORD

9. Click "Connect". It may take several minutes to retrieve a large data export
10. You must be logged into Wave or be within the defined schedule for data connection calls to succeed

CLOSE (YOU CAN RETURN HERE LATER)

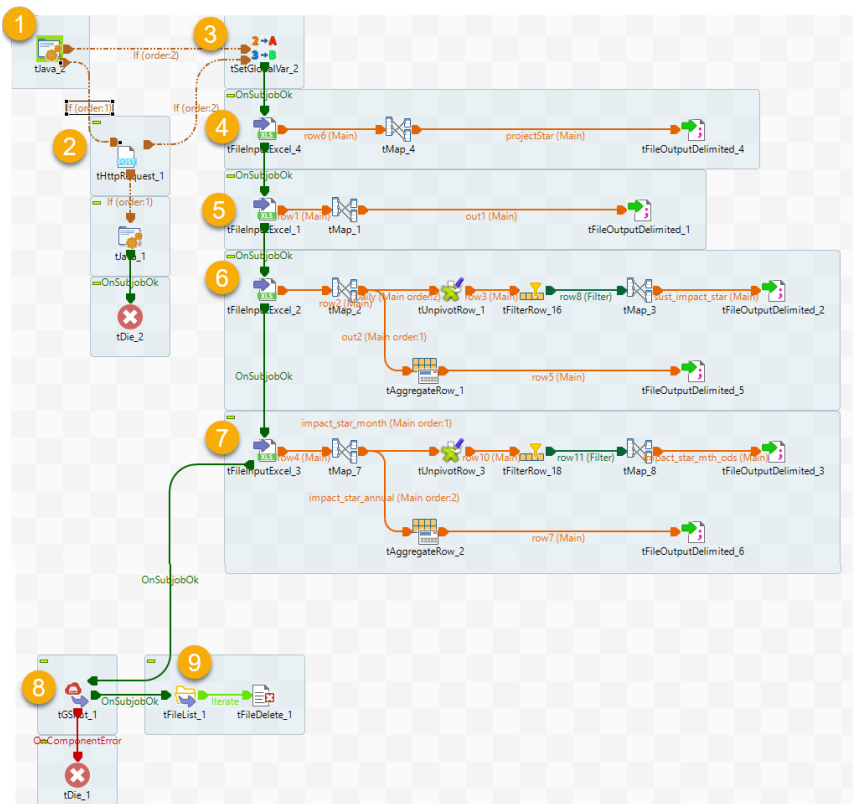
## Talend

The Talend project is on DATA\_OCEAN\_DOMAIN\_INDUSTRIAL

The project will required 4 sheets from the file output from the API. Then, in order to improve the performance of the loading and the size of data, the impactStar and SustainabilityImpactStar sheet are normalize into 2 tables

## Detail job

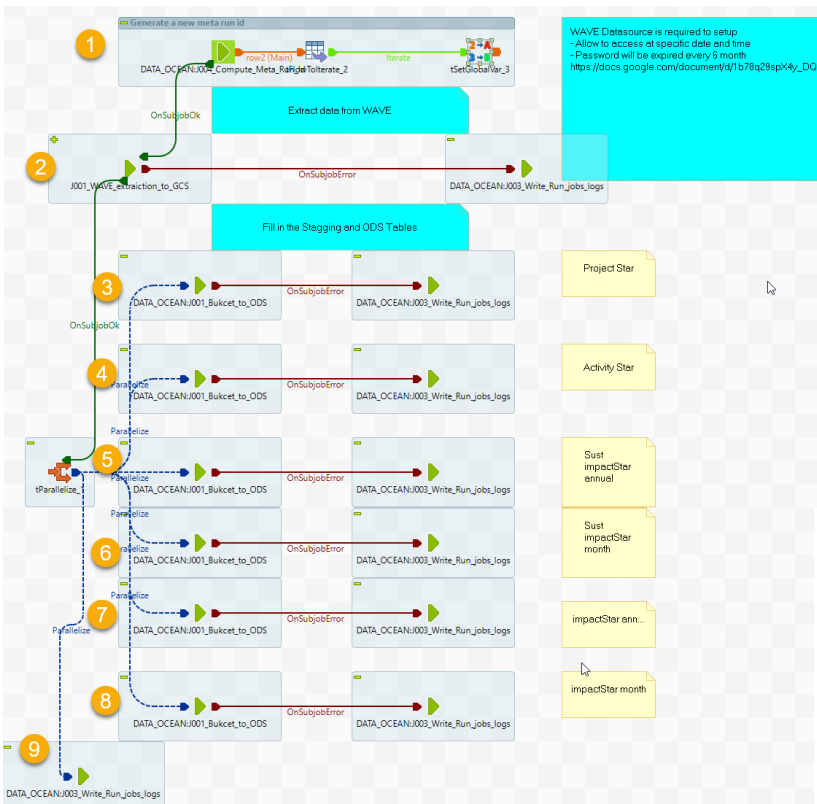
- J001\_WAVE\_extraction\_to\_GCS



1. tJava to check I\_VAR\_WAVE\_reload, true will get the file wave.xlsx, false will get from Wave API
2. tHttpRequest to call Wave api in case I\_VAR\_WAVE\_reload = false
  - a. if return code = 200, the api will generate the file wave.xlsx
  - b. if return code != 200, the the email will be sent to inform DataOps
3. tSetGlobalVar to set current date for the snapshot value of this loading
4. tFileInputExcel4 to read wave.xlsx on projectStar Sheet, then transform with tMap and save file to "WAV\_IND\_0000\_0000\_F001\_[yyyyMMddHHmmss]\_0000\_F\_D\_project\_star.csv"
5. tFileInputExcel1 to read wave.xlsx on activityStar Sheet, then transform with tMap and save file to "WAV\_IND\_0000\_0000\_F002-[yyyyMMddHHmmss]\_0000\_F\_D\_activity\_star.csv"
6. tFileInputExcel2 to read wave.xlsx on sustainabilityImpactStar Sheet, then spit into 2 files
  - a. month : unpivot by changing column (period) to row and remove the line which does not have data blank or 0 and save to "WAV\_IND\_0000\_0000\_F003\_[yyyyMMddHHmmss]\_0000\_F\_D\_sust\_impact\_star\_month.csv"
  - b. annual : keep all columns except the Purpose and period and aggregate the row to reduce duplicate rows and save file to "WAV\_IND\_0000\_0000\_F003\_[yyyyMMddHHmmss]\_0000\_F\_D\_sust\_impact\_star\_annaul.csv"
7. tFileInputExcel3 to read wave.xlsx on impactStar Sheet, then spit into 2 files
  - a. month : unpivot by changing column (period) to row and remove the line which does not have data blank or 0 and save to "WAV\_IND\_0000\_0000\_F004\_[yyyyMMddHHmmss]\_0000\_F\_D\_impact\_star\_month.csv"
  - b. annual : keep all columns except the Purpose and period and aggregate the row to reduce duplicate rows and save file to "WAV\_IND\_0000\_0000\_F004\_[yyyyMMddHHmmss]\_0000\_F\_D\_impact\_star\_annaul.csv"
8. Upload all 6 files to GCS
9. Delete the 6 files from Talend server

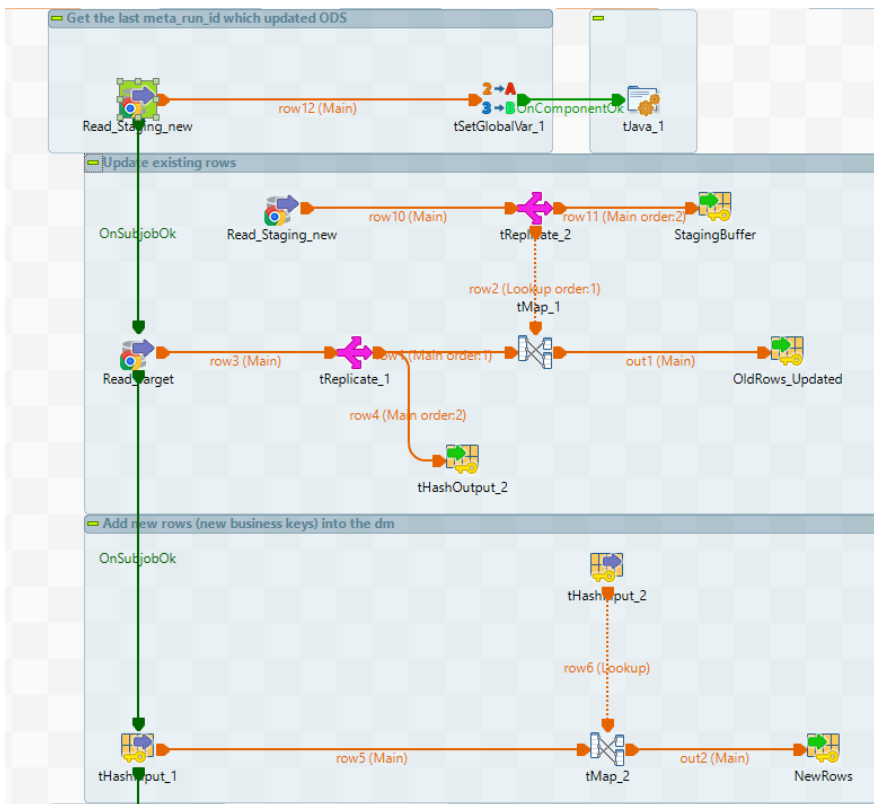
## Flow job

- F001\_WAVE\_TO\_BQ



1. Setup meta\_run\_id and filename of the output file
2. Call the detail job and pass parameters
3. Call the standard job to upload the projectStar files from GCS to ODS
4. Call the standard job to upload the activityStar files from GCS to ODS
5. Call the standard job to upload the sustainabilityImpactStar annual files from GCS to ODS
6. Call the standard job to upload the sustainabilityImpactStar month files from GCS to ODS
7. Call the standard job to upload the impactStar annual files from GCS to ODS
8. Call the standard job to upload the impactStar month files from GCS to ODS
9. Update the log.

- F011\_DIM\_wave\_project\_star\_SCD1



Select only last loading in ODS update to DIM\_wave\_project\_star

## Access rights

It is required to get access from WAVE admin [ednamaria.coan@solvay.com](mailto:ednamaria.coan@solvay.com)

## Source

- <https://solvay.mckinseywave.com/login/auth> prj-data-dm-industrial-[env]
- The API URL will be configured on Wave connection and the URL will keep in parameter I\_VAR\_WAVE\_url\_api

## Format

- columnar format (Excel - .xlsx)

## Destination

### Data Ocean

- Bucket = cs-ew1-prj-data-dm-industrial-[env]-staging
- DataOcean GCP = prj-data-dm-industrial-[env]
- STG Table name = prj-data-dm-industrial-[env].STG.
  - STG\_WAV\_0000\_0000\_F001\_F\_D\_project\_star
  - STG\_WAV\_0000\_0000\_F002\_F\_D\_activity\_star
  - STG\_WAV\_0000\_0000\_F003\_F\_D\_sust\_impact\_star\_annual
  - STG\_WAV\_0000\_0000\_F003\_F\_D\_sust\_impact\_star\_month
  - STG\_WAV\_0000\_0000\_F004\_F\_D\_impact\_star\_annual
  - STG\_WAV\_0000\_0000\_F004\_F\_D\_impact\_star\_month
- ODS Table name = prj-data-dm-industrial-[env].ODS
  - ODS\_WAV\_0000\_F001\_F\_D\_project\_star
  - ODS\_WAV\_0000\_F002\_F\_D\_activity\_star
  - ODS\_WAV\_0000\_F003\_F\_D\_sust\_impact\_star\_annual
  - ODS\_WAV\_0000\_F003\_F\_D\_sust\_impact\_star\_month
  - ODS\_WAV\_0000\_F004\_F\_D\_impact\_star\_annual
  - ODS\_WAV\_0000\_F004\_F\_D\_impact\_star\_month

- DM Tables name = prj-data-dm-industrial-[env].DM
  - DIM\_wave\_project\_star
  - DIM\_wave\_activity\_star

## Product

- Bucket = cs-ew1-prj-data-industrial-dash-[env]-staging
- Product GCP = prj-data-industrial-dash-[dev]
- Dataset = prj-data-industrial-dash-dev.DataOcean DPL
  - V\_ODS\_wave\_activity\_star\_fl
  - V\_ODS\_wave\_impact\_star\_annual\_fl
  - V\_ODS\_wave\_impact\_star\_month\_fl
  - V\_ODS\_wave\_project\_star\_fl
  - V\_wave\_activity\_star
  - V\_wave\_impact\_star\_annual
  - V\_wave\_impact\_star\_month
  - V\_wave\_project\_star
  - V\_wave\_sust\_impact\_star\_annual
  - V\_wave\_sust\_impact\_star\_month

## Format

- columnar format

## Sizing

ProjectStar around 7000 records

Sust ImpactStart annual around 9000 records

Sust ImpactStart month around 54000 records

ImpactStart annual around 85,000 records

ImpactStart month around 1,226,000 records

## Assessment

Data should be the same as output file and

`SELECT distinct snapshot FROM `prj-data-dm-industrial-dev.ODS.ODS_WAV_0000_F004_F_D_impact_star_annual` order by snapshot desc`

The current date of snapshot should be available

## Loading

### 1.1 Incremental Load

#### Source to ODS

N/A Always full load from the file

#### ODS to DM

`I_VAR_dim_wave_project_star_reload = incremental`

### 1.2 Full load

Set `I_VAR_WAVE_reload = false`

`PL_INDUS_WAVE` will load the output excel file from Wave API and save the current time as snapshot field

### 1.3. Reloading data

#### Source to ODS

Set I\_VAR\_WAVE\_reload = true (in case of load historical data, manual file, which is not get from Wave API)

Key users will provide the historical file and load the file manually by rename the file to wave.xlsx on folder \\Acew1dt\ndeng02\data\DEV\DATA\_OCEAN\_DOMAIN\_INDUSTRIAL\InOut\WAVE. The file must have the same sequence of column same as the normal extraction.

## ODS to DM

I\_VAR\_dim\_wave\_project\_star\_reload = QUALIFY ROW\_NUMBER() OVER (PARTITION BY snapshot, projectStar ORDER BY meta\_ods\_insert\_date DESC) = 1

## 1.4 Plan to schedule

Weekday at 04:01 AM CET

## 1.5 Timing

The average time expected for loading: around 20 mins

## Criticality

High/Medium/Low

## Logging

On project prj-data-dm-industrial-[environment]

```
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 logs.meta_run_id in (SELECT meta_run_id FROM STG.run_jobs order by meta_start_date desc limit 100) and job_name like '%WAVE%' and meta_start_date > DATE_SUB(CURRENT_TIMESTAMP(), INTERVAL 24 HOUR) order by job.meta_start_date desc
```