OEE

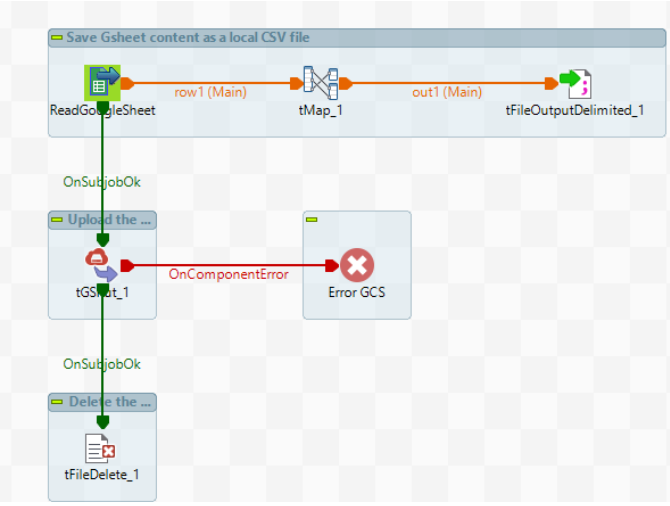
Description

The source data is the gsheet from user, we load to big query (BQ) on project prj-data-dm-industrial-[env].

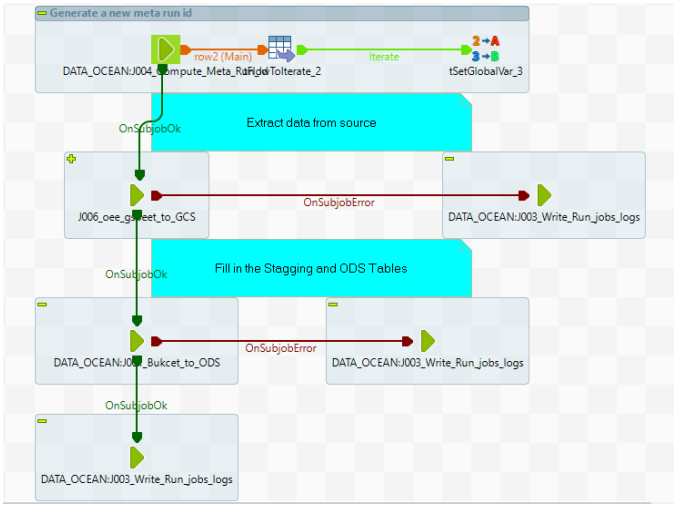
Talend

Source to ODS (DATA_OCEAN_DOMAIN_INDUSTRIAL)

J006_oee_gsheet_to_GCS



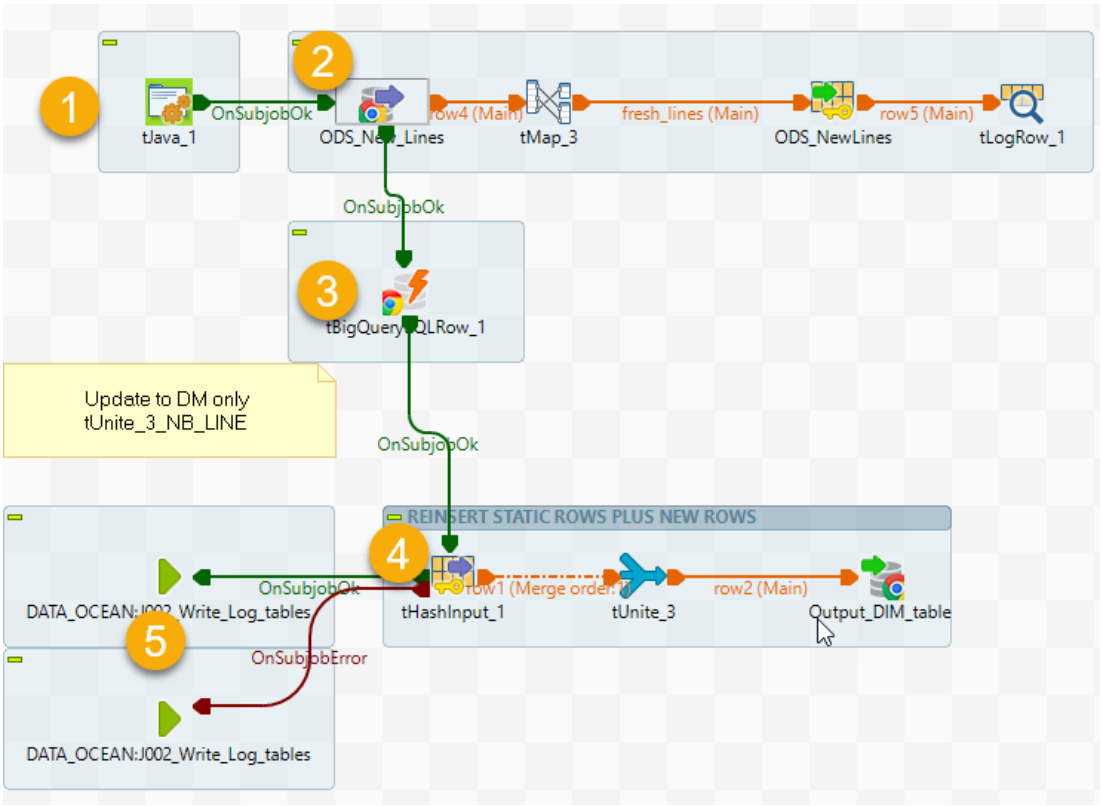
F006_Gsheet_OEE_TO_BQ



1. Read data from
    a. gsheet l_VAR_GOOGLE_SHEET_oee_id
    b. tabname l_VAR_GOOGLE_SHEET_oee_tab
    c. start row l_VAR_GOOGLE_SHEET_oee_row_start = 2
    d. end row l_VAR_GOOGLE_SHEET_oee_row_end = 0 means all rows
2. Save data to bucket cs-ew1-prj-data-dm-industrial-dev-staging/Gsheet_OEE
3. Load data to prj-data-dm-industrial-dev
    a. STG_FIL_0000_0000_F001_F_H_oee
    b. ODS_FIL_0000_F001_F_H_oee

ODS to DM

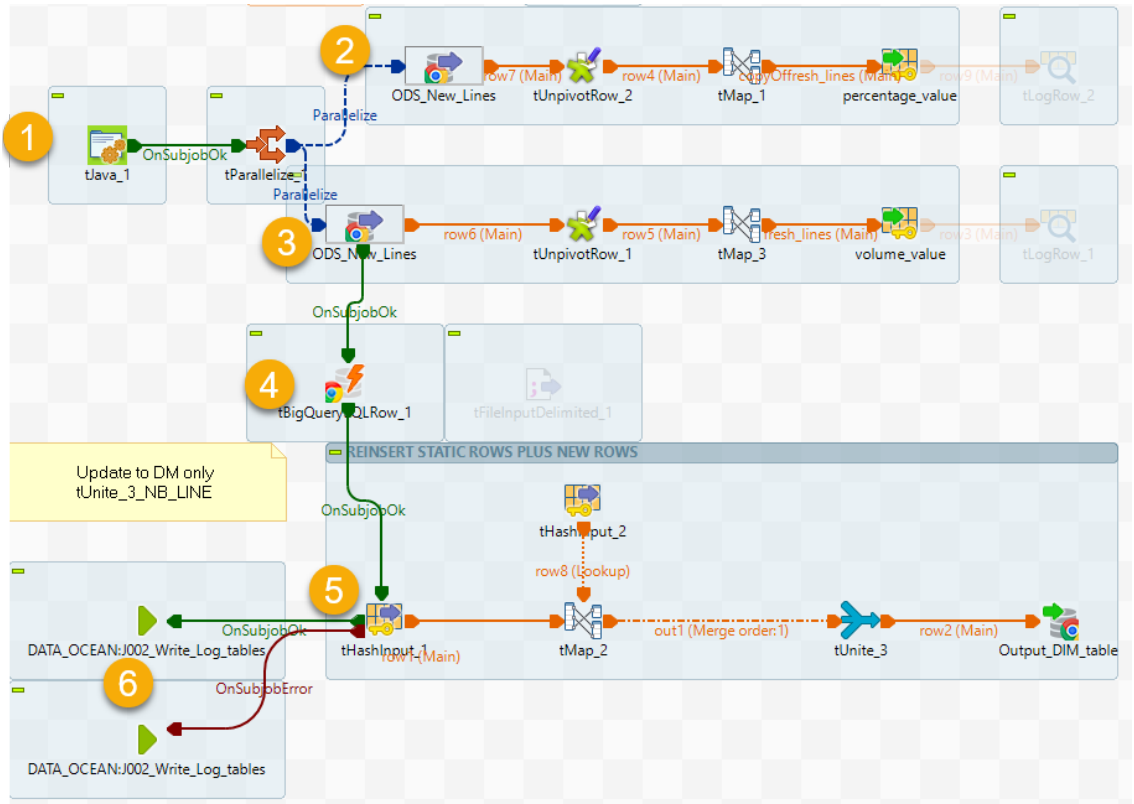
F016_FACT_oee_main (J016_FACT_oee_main)



1. Assigned extraction time and set the loading getting from last load from ODS_FIL_0000_F001_F_H_oee max meta_business_date
2. Select data from ODS only the main fields regarding requirement (the rest of the measure except the column starting wit number 1 - 6 are kept in losses table)
3. Delete data on DM.FACT_oee_main before load data from ODS because there is a case that user enter date 1/1/2024 and change to 01/01 /2024, which cause duplicate records

4. Load the data from point 2 to DM.FACT_oee_main
5. Write the log

F016_FACT_oee_losses (J016_FACT_oee_losses)



1. Assigned extraction time and set the loading getting from last load from ODS_FIL_0000_F001_F_H_oee max meta_business_date
2. Select data from ODS only the losses fields on percentage and unpivot the data regarding requirement
3. Select data from ODS only the losses fields on volume and unpivot the data
4. Delete data on DM.FACT_oee_losses before load data from ODS because there is a case that user enter date 1/1/2024 and change to 01/01 /2024, which cause duplicate records
5. Load the data from point 2 join with point 3 to DM.FACT_oee_losses
6. Write the log

From

1. Volume External causes	1. External causes	2. Volume Planned maintenance shutdown	2. Planned maintenance shutdown	3. Volume Intercampaigns changeovers	3. Intercampaigns changeovers	4. Volume Breakdown	4. Breakdown	5. Volume Performance	5. Performance	6. Volume Quality	6. Quality
956.55724	17.28%	0	0.00%	82.42739	1.49%	101.17442	1.83%	146.55453	2.65%	5.12214	0.09%
0	0.00%	32.82276	0.60%	0	0.00%	98.46827	1.80%	339.16849	6.20%	0	0.00%
1926.21254	21.40%	45.00497	0.50%	135.0149	1.50%	117.01291	1.30%	162.01788	1.80%	0	0.00%
878.09781	21.64%	13.79636	0.34%	91.70522	2.26%	348.96678	8.60%	260.91354	6.43%	33.27358	0.82%
9.48745	0.28%	76.87526	2.29%	21.70002	0.65%	49.32128	1.47%	105.80862	3.15%	-0.87473	-0.03%
51.62664	0.75%	163.82852	2.38%	28.91092	0.42%	134.22925	1.95%	85.35604	1.24%	0	0.00%
73.89723	3.51%	606.79473	28.83%	0	0.00%	328.77478	15.62%	107.34903	5.10%	12.16109	0.58%
455.52947	11.49%	34.6309	0.87%	80.45023	2.03%	628.68394	15.86%	424.13789	10.70%	48.8601	1.23%
224.98428	13.02%	233.20352	13.50%	110.68112	6.41%	181.79844	10.52%	77.73451	4.50%	54.72859	3.17%
47.09968	0.74%	266.32728	4.18%	0	0.00%	635.41749	9.97%	802.78406	12.60%	55.62954	0.87%

To

gbu	site	unit	cluster	losses	volume_value	percentage_value
Peroxides	Longview	*null*	*null*	1_external_causes	1918.776	23.535%
Peroxides	Longview	*null*	*null*	2_planned_maintenance_shutd…	0.000	0.000%
Peroxides	Longview	*null*	*null*	3_intercampaigns_changeovers	0.000	0.0%
Peroxides	Longview	*null*	*null*	4_breakdown	0.000	0.000%
Peroxides	Longview	*null*	*null*	5_performance	0.000	0.000%
Peroxides	Deerpark	*null*	*null*	1_external_causes	1635.194	20.525%
Peroxides	Deerpark	*null*	*null*	2_planned_maintenance_shutd…	0.000	0.000%
Peroxides	Deerpark	*null*	*null*	3_intercampaigns_changeovers	0.000	0.0%
Peroxides	Deerpark	*null*	*null*	4_breakdown	93.645	1.175%
Peroxides	Deerpark	*null*	*null*	5_performance	0.000	0.000%
Peroxides	Warrington	*null*	*null*	1_external_causes	2796.789	61.000%
Peroxides	Warrington	*null*	*null*	2_planned_maintenance_shutd…	0.000	0.000%
Peroxides	Warrington	*null*	*null*	3_intercampaigns_changeovers	0.000	0.0%
Peroxides	Warrington	*null*	*null*	4_breakdown	84.821	1.850%

Access rights

Access to gsheet

Source

- OEE Dev
- OEE Prod

Format

- columnar format (Gsheet)

Destination

Location

- GCP = prj-data-dm-industrial-[env]
- STG table
  - STG_FIL_0000_0000_F001_F_H_oee
- ODS table
  - ODS_FIL_0000_F001_F_H_oee
- DM Table names
  - FACT_oee_main
  - FACT_oee_losses
- DS_industrial_dash
  - V_FACT_oee_main
  - V_FACT_oee_losses

- GCP = prj-data-industrial-dash-[env]
- DataOcean
  - V_FACT_oee_main
  - V_FACT_oee_losses

- DPL View names
  - V_FACT_oee_losses
  - V_FACT_oee_main

Format

- columnar format

Sizing

1,405 rows

Assessment

How to validate that the generated output is valid: Data same as gsheet

Loading

1.1 Incremental Load

N/A

1.2 Full load

PL_INDUS_OEE

Source to ODS

Variable to control:

l_VAR_GOOGLE_SHEET_oee_id = refer to Source session above
l_VAR_GOOGLE_SHEET_oee_tab = OEE (WIP)
l_VAR_GOOGLE_SHEET_oee_row_start = 4
l_VAR_GOOGLE_SHEET_oee_row_end = 0 means all the rows

ODS to DM : l_VAR_oee_main_dm_reload_condition = incremental or blank will load max meta_business_date from ODS

1.3. Reloading data

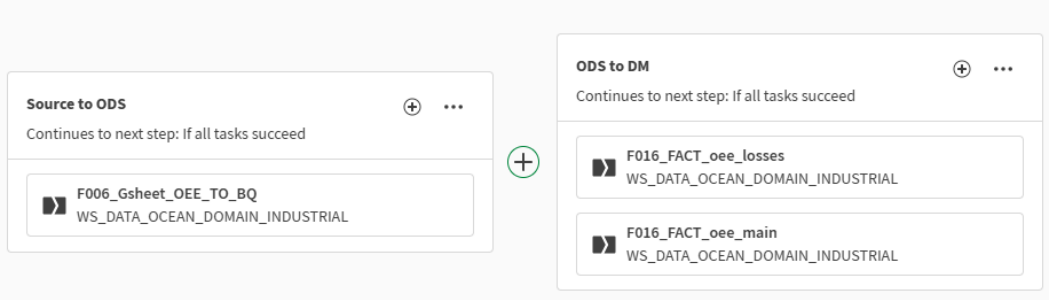
Source to ODS : just run the job again

ODS to DM : control by parameter l_VAR_oee_main_dm_reload_condition. This will impact both main and losses tables. To reload all ODS with latest data l_VAR_oee_main_dm_reload_condition =

```
QUALIFY ROW_NUMBER () OVER ( PARTITION BY date , gbu , site , unit , cluster   ORDER BY meta_ods_insert_date DESC ) = 1
```

1.4 Plan to schedule

Every 30 minutes



1.5 Timing

The average time expected for loading: 10 minutes

Criticality

Low?

Logging

```
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 , log
s.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 1000 )
```

```
and lower ( job_name ) like '%oee%'
and meta_start_date >   DATE_SUB ( CURRENT_TIMESTAMP () , INTERVAL 1 HOUR )
order by job.meta_start_date desc
```