

# 4. Data Loading with Talend

Responsible & contact points:

- Alessandro Mainardi - Project Owner
- Simon Bourguignon - Delivery Manager
- Alba Carrero/ Gaetan Frenoy - Product Owner
- Rui Ferraz - Project Manager

## 4.1 - Talend Integration

Source data integration with Talend ETL tool

Connection details:

- **FTP Server (Meteologica):** <ftp.meteologica.com> (incremental load)
  - Talend connects to the FTP server where files containing current and future electricity prices for Spain, Italy, France, and Germany are located.
  - Talend retrieves these files from the FTP server.
  - The retrieved files are then loaded into Google Cloud Storage.
  - We need to use GCP RE on this source because in order to access Meteologica, it is required statistic IP in order to register the whitelist for data access.
- **Postgres Database (Vendohm):** ~~ses-gcp-sandbox.eu.vendohm--~~ **Decommission replaced by CSV from Dataiku**
  - ~~Talend establishes a direct connection with the Vendohm PostgreSQL database that stores sensor values identified by specific curve IDs.~~
- **Dataiku files (full load)**
  - Vendohm is decommissioned and IRM energy data need to calculated many rules. Therefore, project team decide to get the final calculation of energy from Dataiku.
    - Dataiku will generate energy files (Spot hour, Spot day and Forward combines to 1 consolidation file) as .gz file to DataOcean Industrial bucket (cs-ew1-prj-data-dm-industrial-[env]-staging/in/enr/irm-conso) twice a day
    - Dataiku will generate a file to replace IRM energy deal to DataOcean Industrial bucket (cs-ew1-prj-data-dm-industrial-[env]-staging/in/enr/irm-deals) once a day
  - Talend gets the files considering GCS as data source
  - This extracted data is loaded into Google Cloud Storage as files again in the standard way.
  - [Dataiku project](#)
- **Oracle Database (IRM):** ~~acew1pirmdb01.prod.aws.cloud.solvay.com~~
  - ~~Talend establishes a direct connection with the IRM Oracle database that stores energy deals identified by specific deal IDs.~~
  - ~~It extracts relevant data from the Oracle database.~~
  - ~~This extracted data is loaded into Google Cloud Storage as files.~~
- **Google Sheets (Hedges, wap solid fuels, CO2): (full load)**
  - Talend integrates with Google Sheets, where solid fuel wap, CO2 emissions, and hedging information are stored.
  - It retrieves this data from Google Sheets.
  - Similar to the other sources, this data is also loaded into Google Cloud Storage as files.
    - [CO2](#)
    - [WAP Solid Fuels](#)
    - ~~[Energy Deals Hubs](#)~~
    - ~~[Energy Deals Counter Parties](#)~~
    - ~~[Energy Deals Sites Hedges](#)~~
    - ~~[Vendohm Forwards](#)~~
    - ~~[Vendohm Meteo](#)~~
    - ~~[Vendohm Spot](#)~~

Data Transformation and Loading to Google BigQuery:

- Once data from all four sources (FTP server, ~~PostgreSQL database~~, ~~Oracle database~~, Google Sheets, Dataiku file) is available in Google Cloud Storage as files, Talend proceeds with data transformation and loading.
- Talend performs data transformations as needed, including cleansing, mapping, and structuring the data for consistency.
- The transformed data is loaded into various stages, operational data stores (ODS), and data mart tables within Google BigQuery.
- These tables are organized to facilitate efficient querying and reporting for energy optimization purposes.

By utilizing Talend for data extraction, transformation, and loading (ETL), the web app ensures that data from diverse sources is collected, processed, and structured for analysis and reporting within Google BigQuery, enabling users to make informed decisions based on up-to-date and accurate data.

- 4.1 - Talend Integration
  - J001\_FTP\_to\_GCS-METEOLÓGICA\_FRANCE
  - J002\_FTP\_to\_GCS-METEOLÓGICA\_ITALY
  - J003\_FTP\_to\_GCS-METEOLÓGICA\_GERMANY
  - J004\_FTP\_to\_GCS-METEOLÓGICA\_SPAIN
  - J005\_FTP\_to\_GCS-METEOLÓGICA\_SPAIN\_ENS
  - J006\_FTP\_to\_GCS-METEOLÓGICA\_SPAIN\_OBS
  - J011\_GSheet\_to\_GCS\_VENDOHEM\_FORWARDSDS
  - J012\_GSheet\_to\_GCS\_VENDOHEM\_SPOT
  - J013\_GSheet\_to\_GCS\_VENDOHEM\_METEO
  - J015\_GSheet\_to\_GCS\_IRM\_HUBS
  - J016\_GSheet\_to\_GCS\_IRM\_COUNTERPARTIES
  - J017\_GSheet\_to\_GCS\_IRM\_SITES\_HEDGES
  - J009\_GSheet\_to\_GCS\_CO2\_EMISSIONS
  - J010\_Postgres\_to\_GCS\_VENDOHEM Decommission
  - J010\_FIL\_Energy\_FO\_TO\_ODS
  - J014\_Oracle\_to\_GCS\_IRM
  - J014\_GCS\_to\_GCS\_IRM\_ENERGY\_DEAL
- 4.3 - Data Loading to Google BigQuery
  - J001\_Extraction\_till\_ODS\_METEOLÓGICA\_FRANCE
  - J002\_Extraction\_till\_ODS\_METEOLÓGICA\_ITALY
  - J003\_Extraction\_till\_ODS\_METEOLÓGICA\_GERMANY
  - J004\_Extraction\_till\_ODS\_METEOLÓGICA\_SPAIN
  - J005\_Extraction\_till\_ODS\_METEOLÓGICA\_SPAIN\_ENS
  - J006\_Extraction\_till\_ODS\_METEOLÓGICA\_SPAIN\_OBS
  - J009\_Extraction\_till\_ODS\_GSHEET\_CO2
  - J011\_Extraction\_till\_ODS\_GSHEET\_VENDOHEM\_FORWARDSDS
  - J012\_Extraction\_till\_ODS\_GSHEET\_VENDOHEM\_SPOT

#### GCP Details:

| Domain         | GCP Project                  | Bucket                                                |
|----------------|------------------------------|-------------------------------------------------------|
| Industrial     | prj-data-dm-industrial-[env] | cs-ew1-prj-data-dm-industrial-[env]-staging/ROBUSTIFY |
| Sustainability | prj-data-dm-sust-[env]       | cs-ew1-prj-data-dm-sust-[env]-staging/ROBUSTIFY       |
| Robustify      | prj-data-robustify-[env]     | N/A                                                   |

Cloud Remote Engine = prj-talend-[dev] / ce-ew1-b-talend-gcp-remote-engine-[env]

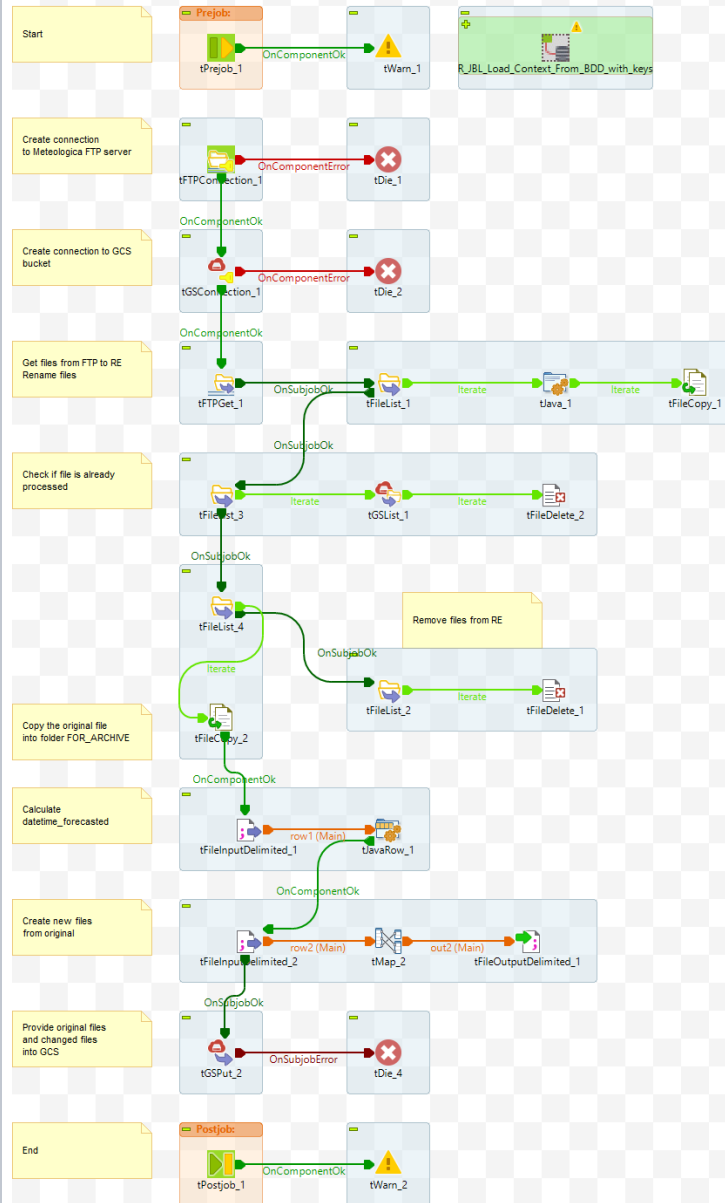
- J013\_Extraction\_till\_ODS\_GSHEET\_VENDOHM\_METEO
- J015\_Extraction\_till\_ODS\_GSHEET\_IRM\_HUBS
- J016\_Extraction\_till\_ODS\_GSHEET\_IRM\_COUNTER\_PARTIES
- J017\_Extraction\_till\_ODS\_GSHEET\_IRM\_SITES\_HEDGES
- J010\_Extraction\_till\_ODS\_VENDOHM
- 4.4 - Load to DM (calculations and transformations)
  - J001\_ODS\_TO\_DM\_FACT\_ENERGY\_PRICE\_FORECAST
  - J002\_ODS\_TO\_DM\_DIM\_METEO\_INFO
  - J003\_ODS\_TO\_DM\_DIM\_ENERGY\_PRICE\_INFO
  - J004\_ODS\_TO\_DM\_FACT\_METEO\_DATA
  - J005\_ODS\_TO\_DM\_FACT\_ENERGY\_PRICE replace with view in Robustify prj-data-robustify-dev.DM.V\_FACT\_energy\_price\_hourly
  - J006\_ODS\_TO\_DM\_FACT\_CO2\_EMISSIONS
  - J007\_ODS\_TO\_DM\_FACT\_SOLID\_FUEL\_WAP
  - J009\_ODS\_TO\_DM\_FACT\_ENERGY DEALS
  - J009\_ODS\_TO\_DM\_FACT\_IRM\_DEALS
  - J001\_DM\_TO\_DM\_FACT\_ENERGY\_PRICE\_HOURLY
- 4.5 - Scheduling and Automation
- 4.6 - Remark

## 4.2 - Source Data Extraction

|                                 |                                                                                                                                                                           |                                                                                              |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| Main jobs for source extraction | <ul style="list-style-type: none"> <li>• J001_FTP_to_GCS-METEO LOGIC A_FRA NCE</li> <li>• J002_FTP_to_GCS-METEO LOGIC A_ITALY</li> <li>• J003_FTP_to_GCS-METEO</li> </ul> | <p>--to the top --</p> <p>GCP Remote Engine(RE)</p> <p>Industrial Data Ocean(DO) project</p> |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                       |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
|  | <div>LOGIC<br/>A_GER<br/>MANY</div> <div><div>• J004_F</div><div>TP_to_</div><div>GCS-</div><div>METEO</div><div>LOGIC</div><div>A_SPAIN</div></div> <div><div>• J005_F</div><div>TP_to_</div><div>GCS-</div><div>METEO</div><div>LOGIC</div><div>A_SPAIN</div><div>N_ENS</div></div> <div><div>• J006_F</div><div>TP_to_</div><div>GCS-</div><div>METEO</div><div>LOGIC</div><div>A_SPAIN</div><div>N_OBS</div></div>                                                             |                       |
|  | <div>Job<br/>description<br/>by steps</div>                                                                                                                                                                                                                                                                                                                                                                                                                                        | <div>Job design</div> |
|  | <div>1. FTP<br/>Server<br/>Connecti<br/>on:</div> <div><div>a. Esta<br/>blish<br/>es a<br/>secu<br/>re<br/>conn<br/>ectio<br/>n to<br/>the<br/>Mete<br/>ologi<br/>ca<br/>FTP<br/>serv<br/>er.</div><div>b. Retri<br/>eves<br/>files<br/>from<br/>a<br/>pred<br/>efine<br/>d<br/>folde<br/>r<br/>locat<br/>ed<br/>on<br/>the<br/>FTP<br/>serv<br/>er.</div></div> <div>2. File<br/>Consolid<br/>ation and<br/>Renamin<br/>g:</div> <div><div>a. After<br/>fetch<br/>ing</div></div> |                       |

- the files, the job consolidates and renames them according to predefined naming conventions.
- b. This step ensures uniformity and consistency in file naming for further processing.
3. Creating File Backups:
- a. As a best practice for data integrity, the job generates copies of the original files.
- b. These copies are stored in an archive folder on the rem



ote  
engi  
ne  
for  
back  
up  
and  
histo  
rical  
refer  
ence.

4. Checking  
for  
Previousl  
y  
Processe  
d Files:

a. The  
job  
perf  
orms  
a  
chec  
k to  
dete  
rmin  
e if  
the  
pulle  
d  
files  
have  
alrea  
dy  
been  
proc  
esse  
d.

b. This  
verifi  
catio  
n is  
cruci  
al to  
prev  
ent  
the  
repr  
oces  
sing  
of  
files  
from  
previ  
ous  
days  
, as  
older  
files  
may  
still  
exist  
in  
the  
folde  
r.

5. Data  
Extraction  
and  
Transfor  
mation:

a. For  
files  
that  
have  
not  
been  
proc  
esse

d  
earli  
er,  
the  
job  
proc  
eeds  
with  
data  
anal  
ysis.

b. It  
extra  
cts  
a  
datet  
ime  
valu  
e  
from  
each  
file,  
locat  
ed  
at a  
spec  
ified  
posit  
ion  
withi  
n  
the  
file.

c. Usin  
g  
this  
datet  
ime  
infor  
mati  
on,  
the  
job  
calc  
ulate  
s a  
new  
colu  
mn,  
enab  
ling  
preci  
se  
time  
stam  
ping  
of  
the  
data.

6. CSV  
Output  
and  
Google  
Storage:

a. The  
proc  
esse  
d  
data,  
now  
enric  
hed  
with  
the  
new  
time  
stam  
p  
colu

mn,  
is  
conv  
erte  
d  
into  
a  
CSV  
form  
at.

- b. The  
resul  
ting  
CSV  
outp  
ut  
files  
are  
then  
push  
ed  
to  
Goo  
gle  
Clou  
d  
Stor  
age  
for  
furth  
er  
stora  
ge  
and  
acce  
ss.

7. File  
Deletion:

- a. Onc  
e  
the  
files  
have  
been  
succ  
essf  
ully  
proc  
esse  
d  
and  
their  
data  
store  
d in  
Goo  
gle  
Clou  
d  
Stor  
age,  
they  
are  
delet  
ed  
from  
the  
rem  
ote  
engi  
ne.
- b. This  
delet  
ion  
step  
help  
s  
man  
age

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                          |  |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | <p>storage resources efficiently and ensures that only processed data is retained.</p> <p>This Talend job ensures the efficient handling of data from the Meteorologica FTP server, including consolidation, backup, and validation steps. It also extracts and enriches the data, making it ready for use in downstream processes, ultimately contributing to the accuracy and usability of the energy optimization project's data.</p> |  |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

|                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                              |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Main jobs for source extraction | <ul style="list-style-type: none"><li>• <del>J008_GS</del><br/><del>heet_t</del><br/><del>heet_to_</del><br/><del>GCS_WA</del><br/><del>P_SOLID</del><br/><del>_FUELS</del></li><li>• <del>J011_G</del><br/><del>Sheet_t</del><br/><del>e_GCS</del><br/><del>_VEND</del><br/><del>OHM_F</del><br/><del>ORWA</del><br/><del>RDS</del></li><li>• <del>J012_G</del><br/><del>Sheet_t</del><br/><del>e_GCS</del><br/><del>_VEND</del><br/><del>OHM_S</del><br/><del>POT</del></li><li>• <del>J013_G</del><br/><del>Sheet_t</del><br/><del>e_GCS</del><br/><del>_VEND</del><br/><del>OHM_</del><br/><del>METEO</del></li><li>• <del>J015_G</del><br/><del>Sheet_t</del></li></ul> | <p>--to the top --</p> <p>GCP Remote Engine</p> <p>Industrial DO project</p> |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|

~~e\_GGS~~  
~~\_IRM\_H~~  
~~UBS~~  
 • ~~J016\_G~~  
~~Sheet\_t~~  
~~e\_GGS~~  
~~\_IRM\_G~~  
~~OUNTE~~  
~~R\_PAR~~  
~~TIES~~  
 • ~~J017\_G~~  
~~Sheet\_t~~  
~~e\_GGS~~  
~~\_IRM\_S~~  
~~ITES\_H~~  
~~EDGES~~

|  | Job description by steps                                                                                                                                                                                                                                                                               | Job design                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p>1. Google Sheets Connection:</p> <ol style="list-style-type: none"> <li>The job initiates a connection to a specified Google Sheets file.</li> <li>It targets a specific sheet within the Google Sheets document, identified by its unique ID.</li> </ol> <p>2. Data Extraction and Formatting:</p> | <pre> graph TD     Start([Start]) --&gt; Prejob[Prejob]     subgraph Prejob         tPrejob_1[tPrejob_1]     end     tPrejob_1 -- OnComponentOk --&gt; tWarn_1[tWarn_1]     tPrejob_1 -- OnComponentOk --&gt; RJL_log_catcher[RJL_log_catcher]     ReadHedging[Read Hedging table<br/>put the file in RE] --&gt; ReadGoogleSheet[ReadGoogleSheet]     ReadGoogleSheet -- OnSubjobOk --&gt; Upload[Upload the ...]     subgraph UploadBlock [Upload the ...]         tGSheet_1[tGSheet_1]     end     tGSheet_1 -- OnComponentError --&gt; ErrorGCS[Error GCS]     tGSheet_1 -- OnSubjobOk --&gt; Delete[Delete the ...]     subgraph DeleteBlock [Delete the ...]         tFileDelete_1[tFileDelete_1]     end     DeleteBlock --&gt; Postjob[Postjob]     subgraph PostjobBlock [Postjob]         tPostjob_1[tPostjob_1]     end     tPostjob_1 -- OnComponentOk --&gt; tWarn_2[tWarn_2]     tPostjob_1 -- OnComponentOk --&gt; End([End])   </pre> |

- a. The job extracts data from the designated Google Sheets sheet.
- b. The extracted data is converted into a CSV format with a predefined fixed schema.
- c. The schema format is consistent and predefined to ensure data uniformity and structure.
- d. The resulting CSV file includes a predefined filename that incorporates the date of

extra  
ction.  
3. Copying  
to  
Google  
Storage:

a. After  
the  
succ  
essf  
ul  
extra  
ction  
and  
form  
attin  
g of  
the  
data,  
the  
job  
copi  
es  
the  
CSV  
file  
to  
Goo  
gle  
Clou  
d  
Stor  
age.

b. This  
step  
serv  
es  
as a  
secu  
re  
and  
relia  
ble  
mea  
ns  
of  
stori  
ng  
the  
proc  
esse  
d  
data  
in  
the  
clou  
d.

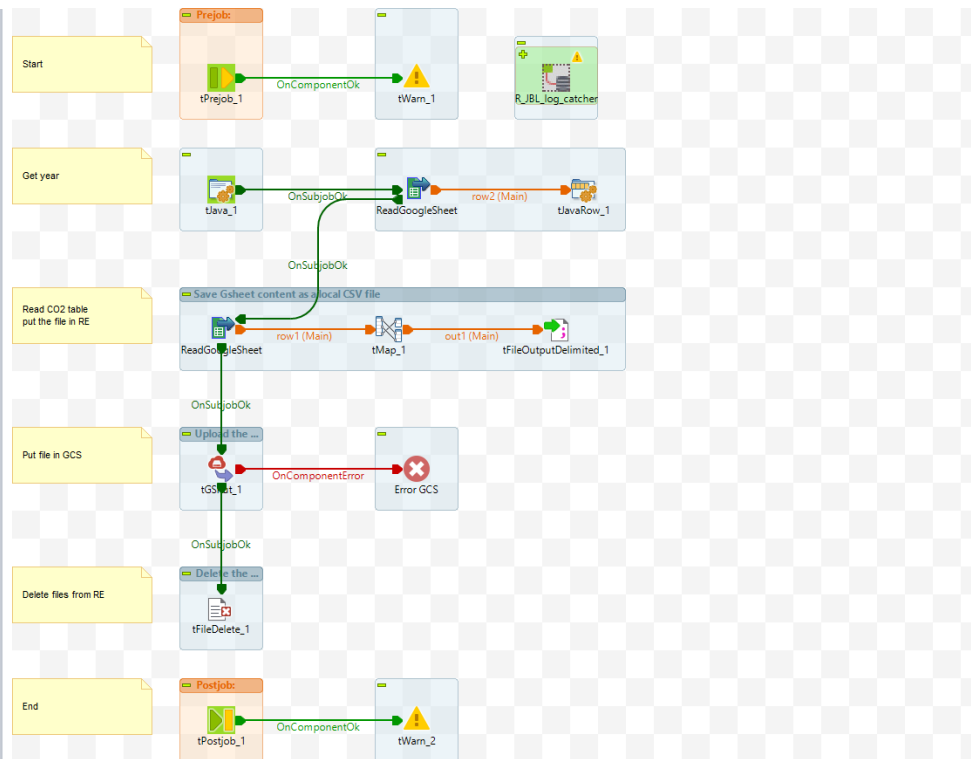
4. File  
Deletion:

a. Follo  
wing  
the  
succ  
essf  
ul  
copy  
ing  
of  
the  
CSV  
file  
to  
Goo  
gle  
Clou  
d  
Stor  
age,

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | <p>the original file located on the remote engine is deleted.</p> <p>b. This cleanup step helps manage storage resources efficiently and maintains data integrity.</p> <p>This Talend job efficiently extracts, formats, and securely stores data from a Google Sheets document into Google Cloud Storage, ensuring that the data is readily available for further analysis and processing while adhering to a fixed schema and naming conventions.</p> |  |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

|                                 |                                                                                             |                                                                                                  |
|---------------------------------|---------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Main jobs for source extraction | <ul style="list-style-type: none"> <li>• <b>J009_GSheet_to_GCS_CO2_EMISSIONS</b></li> </ul> | <p><a href="#">--to the top --</a></p> <p>GCP Remote Engine</p> <p>Sustainability DO project</p> |
|                                 | <p><b>Job description by steps</b></p>                                                      | <p><b>Job design</b></p>                                                                         |
|                                 | <p>1. Google Sheets Connection:</p>                                                         |                                                                                                  |

- b. It focuses on a particular sheet within the Google Sheets document, identified by its unique ID.



## 2. Data Extraction and Formatting:

- The job extracts data from the designated Google Sheet's sheet.
- The extracted data is transformed into a CSV format, adhering to a predefined, fixed schema.
- The schema format is consistent and predefined.

to ensure uniformity and data structure.

- d. The resulting CSV file follows a predefined naming convention, which incorporates the date of extraction.

3. Data Reading:

- a. The job involves two distinct data reading steps.

- i. First Reading: The job reads only a single cell with this

file, specifically the cell contentaining the yearly arrival value. This is set to capture yearly arrival data. Second Reading

g : These second read of the file is reserved for capturing the remaining data, defined by the schema and for

m  
a  
t.  
T  
h  
i  
s  
s  
t  
e  
p  
c  
o  
m  
p  
r  
e  
h  
e  
n  
s  
i  
v  
e  
l  
y  
r  
e  
a  
d  
s  
t  
h  
e  
d  
a  
t  
a  
s  
p  
e  
c  
i  
f  
i  
e  
d  
i  
n  
t  
h  
e  
s  
c  
h  
e  
m  
a.

4. Copying to Google Storage:

- a. After successfully extracting and formatting the data, the job copies the CSV file to Google Cloud Storage (inside Sustainability

- domain).
- b. This step serves as a secure and reliable means of storing the processed data in the cloud.

5. File

Deletion:

- a. Following the successful copying of the CSV file to Google Cloud Storage, the original file located on the remote engine is deleted.
- b. This cleanup step helps manage storage resources efficiently and maintains data integrity.

This Talend job effectively extracts, formats, and securely stores data from a Google Sheets document into Google Cloud Storage. It features specialized data reading

|  |                                                                                                                                                                                                                 |  |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | steps, ensuring the capture of year-related data separately, while adhering to predefined schemas and naming conventions. The deletion of the original file enhances data management and resource optimization. |  |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

|                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                              |
|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| <p><b>Main jobs for source extraction</b></p> | <ul style="list-style-type: none"> <li><del>J010_Postgr</del><br/><del>es to GCS</del><br/><del>VENDOHM</del><br/><b>Decommission</b></li> <li><b>J010_FIL_Energy_F_O_T O_ODS</b></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                       | <p>--to the top --</p> <p>AWS Remote Engine</p> <p>Industrial DO project</p> |
|                                               | <p><b>Job description by steps</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <p><b>Job design</b></p>                                                     |
|                                               | <p>GCS Connection:</p> <ul style="list-style-type: none"> <li>Dataiku will generate energy a file that calculated many rules from the Energy team and save to GCS at DataOcean Industrial bucket (cs-ew1-prj-data-dm-industrial-[env]-staging/in /enr) twice a day to get the historical data to future hourly.</li> </ul> <p>The mapping file and fields</p> <ol style="list-style-type: none"> <li>First loading the data from cs-ew1-prj-data-dm-industrial-[env]-staging/in /enr to local drive and unzip the file</li> <li>Upload the file to cs-ew1-prj-data-dm-industrial-[env]-</li> </ol> |                                                                              |

|                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                              |
|------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
|                                          | <p>staging<br/>/ENERGY/irm-<br/>conso</p> <p>3. Storage in<br/>Google Cloud<br/>Storage (GCS) :</p> <ul style="list-style-type: none"><li>◦ After the<br/>successful<br/>extraction<br/>and<br/>formatting<br/>of the data,<br/>the job<br/>securely<br/>stores the<br/>CSV file in<br/>Google<br/>Cloud<br/>Storage.</li><li>◦ The files<br/>are placed<br/>within a<br/>designated<br/>"ENERGY"<br/>folder in<br/>the GCS<br/>location for<br/>efficient<br/>storage<br/>and<br/>accessibilit<br/>y.</li><li>◦ Also move<br/>the source<br/>file .gz to<br/>folder<br/>"Processed"</li><li>◦ Following<br/>a<br/>successful<br/>data<br/>extraction<br/>and<br/>storage<br/>process,<br/>the original<br/>file located<br/>on the<br/>remote<br/>engine is<br/>deleted to<br/>manage<br/>storage<br/>resources<br/>efficiently.</li><li>◦ The<br/>connection<br/>to the<br/>Oracle<br/>database<br/>is closed to<br/>ensure<br/>proper<br/>resource<br/>manageme<br/>nt and<br/>security.</li></ul> |                                                                              |
| Main jobs<br>for<br>source<br>extraction | <ul style="list-style-type: none"><li>• <del>J014_Oracle</del><br/><del>_to_GCS_IR</del><br/><del>M-</del><br/>J014_GCS_t<br/>o_GCS_IRM</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | <p>--to the top --</p> <p>AWS Remote Engine</p> <p>Industrial DO project</p> |

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <b>_ENERGY_D<br/>EAL</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|  | <b>Job description<br/>by steps</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <b>Job design</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|  | <ol style="list-style-type: none"><li>1. Download file from GCS (output of Dataiku) cs-ew1-prj-data-dm-industrial-[env]-staging/in/enr/irm-deals</li><li>2. Unzip the file and keep to the GCS cs-ew1-prj-data-dm-industrial-[env]-staging/ENERGY/irm-deals<ul style="list-style-type: none"><li>• The data is structured according to a predefined and fixed schema to maintain data consistency.</li></ul></li><li>3. Data Formatting and CSV Output:<ul style="list-style-type: none"><li>• The extracted data is formatted and transformed into CSV format.</li><li>• Each resulting CSV file includes the table name and the date of extraction for reference and organization.</li></ul></li><li>4. Storage in Google Cloud Storage (GCS):<ul style="list-style-type: none"><li>• After the successful extraction and formatting of the data, the job securely stores the CSV file in Google Cloud Storage.</li><li>• The files are placed</li></ul></li></ol> | <pre>graph TD     Start([Start]) --&gt; tPrejob_1[tPrejob_1]     tPrejob_1 -- OnComponentOk --&gt; LoadContext[DATA_OCEAN_R_JBL_Load_Context_From_BDD]     LoadContext -- OnSubjobOk TRIGGER_OUTPUT_1 --&gt; tWarn_1[tWarn_1]     LoadContext --&gt; tSetGlobalVar_1[tSetGlobalVar_1]     tSetGlobalVar_1 -- OnSubjobOk --&gt; tGSList_1[tGSList_1]     tGSList_1 -- Iterate --&gt; tJava_1[tJava_1]     tJava_1 -- Iterate --&gt; tGGet_1[tGGet_1]     tGGet_1 --&gt; tFileUnarchive_1[tFileUnarchive_1]     tFileUnarchive_1 --&gt; tFileList_1[tFileList_1]     tFileList_1 --&gt; tJava_3[tJava_3]     tFileList_1 --&gt; tGSet_1[tGSet_1]     tGSet_1 -- OnComponentError --&gt; ErrorGCS[Error GCS]     tGSet_1 -- OnSubjobOk --&gt; tFileDelete_1[tFileDelete_1]     tFileDelete_1 -- OnSubjobOk --&gt; tFileDelete_2[tFileDelete_2]     tFileDelete_2 -- OnSubjobOk --&gt; GSCopy[GCS_CopyFile_StagingFolder]     GSCopy -- OnSubjobOk --&gt; tJava_2[tJava_2]     tPostjob_1[tPostjob_1] -- OnComponentOk --&gt; tWarn_2[tWarn_2]</pre> |

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |  |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | <p>within a designated "IRM" folder in the GCS location for efficient storage and accessibility.</p> <p>5. File Deletion and Connection Closure:</p> <ul style="list-style-type: none"> <li>• Following a successful data extraction and storage process, the original file located on the remote engine is deleted to manage storage resources efficiently.</li> <li>• Source file from Dataiku will move to folder Process and rename to the file + timestamps</li> </ul> |  |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

## 4.3 - Data Loading to Google BigQuery

|                                 |                                                                                                                                                                                                                                          |                                                                                              |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| Main jobs for source extraction | <ul style="list-style-type: none"> <li>• J001_Extraction_til_ODS_ME_TEOLOGICA_FRANCE</li> <li>• J002_Extraction_til_ODS_ME_TEOLOGICA_ITALY</li> <li>• J003_Extraction_til_ODS_ME_TEOLOGICA_GERMANY</li> <li>• J004_Extraction</li> </ul> | <p><a href="#">--to the top --</a></p> <p>GCP Remote Engine</p> <p>Industrial DO project</p> |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|

|  |                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <div>n_til_O<br/>DS_ME<br/>TEOLO<br/>GICA_<br/>SPAIN<br/>• J005_E<br/>xtractio<br/>n_til_O<br/>DS_ME<br/>TEOLO<br/>GICA_<br/>SPAIN_<br/>ENS<br/>• J006_E<br/>xtractio<br/>n_til_O<br/>DS_ME<br/>TEOLO<br/>GICA_<br/>SPAIN_<br/>OBS</div>                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|  | <div>Job<br/>description<br/>by steps</div>                                                                                                                                                                                                                                                                                                                                                     | <div>Job design</div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|  | <div><div>Job<br/>Initialization<br/>and Logging:<br/><br/>1. Job Run<br/>Initializati<br/>on (Subjob<br/>1):<br/><br/>a. The<br/>job start<br/>s with<br/>the first<br/>subjob,<br/>where<br/>the job<br/>ID is<br/>retrieved<br/>, and the<br/>extra<br/>ction date<br/>is calcu<br/>late d.<br/><br/>b. This cruci<br/>al step<br/>initia<br/>lizes the<br/>job and<br/>captures</div></div> | <div><pre>graph TD     subgraph Prejob         tPrejob_1[tPrejob_1]     end     subgraph Init         tSetGlobalVar_1[tSetGlobalVar_1]         tStatCatcher_1[tStatCatcher_1]         tFlowTolterate_1[tFlowTolterate_1]         tSetGlobalVar_2[tSetGlobalVar_2]     end     subgraph ComputeMetaRunId         ComputeMetaRunId[Compute meta_run_id]     end     subgraph ExtractionToBucket         ExtractionToBucket[Extraction to Bucket]     end     subgraph IterateThroughFiles         IterateThroughFiles[Iterate through the files]     end     subgraph BucketToODS         BucketToODS[Bucket to ODS]     end     subgraph WriteRunLogs         WriteRunLogs[Write run logs]     end      tPrejob_1 -- OnComponentOk --&gt; tSetGlobalVar_1     tPrejob_1 -- OnSubjobOk (TRIGGER_OUTPUT) --&gt; tStatCatcher_1     tStatCatcher_1 -- row23 (Main) --&gt; tFlowTolterate_1     tFlowTolterate_1 -- Iterate --&gt; tSetGlobalVar_2     tSetGlobalVar_1 -- OnSubjobOk --&gt; ComputeMetaRunId     ComputeMetaRunId -- row1 (Main) --&gt; tFlowTolterate_2[tFlowTolterate_2]     tFlowTolterate_2 -- Iterate --&gt; tSetGlobalVar_3[tSetGlobalVar_3]     ComputeMetaRunId -- OnSubjobOk --&gt; ExtractionToBucket     ExtractionToBucket -- OnSubjobError --&gt; WriteRunLogs     ExtractionToBucket -- OnSubjobOk --&gt; IterateThroughFiles     IterateThroughFiles -- Iterate --&gt; tGSLList_1[tGSLList_1]     tGSLList_1 -- OnComponentOk --&gt; BucketToODS     BucketToODS -- OnSubjobOk --&gt; WriteRunLogs</pre></div> |

essential  
meta  
data.

Data  
Extraction  
and  
Processing:

1. Files  
Extraction  
from FTP  
Server  
(Subjob  
2):

a. In  
the  
next  
subj  
ob,  
data  
files  
are  
extra  
cted  
from  
the  
Mete  
ologi  
ca  
FTP  
serv  
er.

2. Error  
Handling  
and  
Logging  
(Subjob  
3 - On  
Failure):

a. If  
the  
data  
extra  
ction  
subj  
ob  
(Sub  
job  
2)  
enco  
unte  
rs  
an  
issu  
e or  
does  
not  
finis  
h  
succ  
essf  
ully,  
the  
job  
trans  
ition  
s to  
a  
subj  
ob  
for  
writi  
ng  
error  
logs.

b. Logging is essential for tracking and diagnosing issues in the data extraction process.

3. Data Loading and Transformation (Subjob 4 - On Success):

a. When the data extraction subjob (Subjob 2) successfully completes, the job proceeds with data loading and transformation.

b. In this stage, each extracted file is processed one by one.

c. For each file, a subjob

- is called to load data from the CSV file into a stage table, preparing it for further processing.
- d. Subsequently, data from the stage table is loaded into an operational data store (ODS) table.
- e. All necessary parameters, such as table names and connection parameters, are provided to ensure accurate data extraction and loading.

Logging and Reporting:

- 1. Logging (Subjob 5):
  - a. At the end of the data loading and transformation process, a subjob is called to write logs.
  - b. Logging captures critical information about the data processing, ensuring transparency and traceability.

This Talend job orchestrates the data extraction, loading, and transformation process. It begins with initialization and metadata capture, proceeds to extract data from the FTP server, handles errors if they occur, and performs data loading and transformation for energy optimization purposes. Detailed logging and

|                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                         |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                 | error handling mechanisms enhance job monitoring and maintain data integrity.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                         |
| Main jobs for source extraction | <ul style="list-style-type: none"> <li>• <del>J008_Extraction_til_ODS_GSHEET_WAP_SOLID_FUEL</del></li> <li>• <del>J009_Extraction_til_ODS_GSHEET_CO2</del></li> <li>• <del>J011_Extraction_til_ODS_GSHEET_VENDORM_FORWARDDS</del></li> <li>• <del>J012_Extraction_til_ODS_GSHEET_VENDORM_SPORT</del></li> <li>• <del>J013_Extraction_til_ODS_GSHEET_VENDORM_MEETO</del></li> <li>• <del>J014_Extraction_til_ODS_GSHEET_RACLERM_J014_FIL_IRM_ENERGY_DEAL_F_O_T_O_ODS</del></li> <li>• <del>J015_Extraction_til_ODS_GSHEET_IRMHUBS</del></li> <li>• <del>J016_Extraction_til_ODS_GSHEET_I</del></li> </ul> | <p><a href="#">--to the top --</a></p> <p>GCP Remote Engine</p> <p>Industrial DO project</p> <p>Exception :</p> <ul style="list-style-type: none"> <li>• J009_Extraction_til_ODS_GSHEET_CO2 on GCP RE and on Sustainability DO project</li> <li>• J014_FIL_IRM_ENERGY_DEAL_F_O_TO_ODS on AWS RE and on Industrial DO project</li> </ul> |

RM\_CO  
UNTER  
\_PARTI  
ES  
• J017\_E  
xtractio  
n\_till\_O  
DS\_CS  
HEET\_I  
RM\_SIT  
ES\_HE  
DGES

Job  
description  
by steps

Job design

Job  
Initialization  
and Logging:

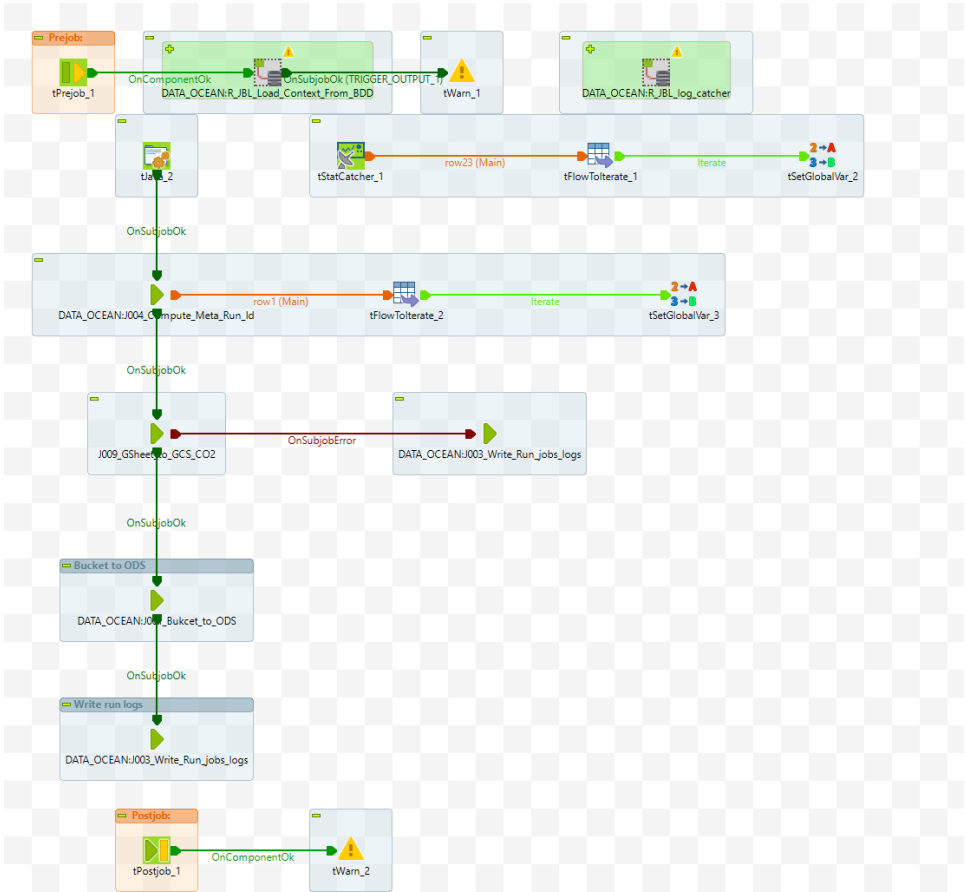
1. File  
Name  
Configura  
tion  
(Global  
Variable):

a. The  
job  
begi  
ns  
by  
confi  
gurin  
g  
the  
file  
nam  
e  
as  
a  
glob  
al  
varia  
ble.

b. This  
varia  
ble  
is  
used  
thro  
ugh  
out  
the  
job  
to  
ensu  
re  
cons  
isten  
cy  
in  
file  
hand  
ling.

2. Job Run  
Initializati  
on  
(Subjob  
1):

a. In  
the  
first  
subj  
ob,



- the  
job  
ID  
is  
retri  
eved  
,  
and  
the  
extra  
ction  
date  
is  
calc  
ulate  
d.
- b. This  
step  
sets  
the  
stag  
e  
for  
the  
data  
proc  
essi  
ng  
and  
capt  
ures  
esse  
ntial  
meta  
data.

Data  
Extraction  
from Google  
Sheets / IRM  
database:

1. Files  
Extraction  
from  
Google  
Sheets /  
IRM  
database  
(Subjob  
2):

- a. The  
next  
subj  
ob  
is  
resp  
onsi  
ble  
for  
extra  
cting  
data  
from  
Goo  
gle  
She  
ets/  
IRM  
data  
base.
- b. It  
targ  
ets  
the  
spec  
ified  
file  
base

d on  
the  
glob  
al  
varia  
ble  
(file  
nam  
e)  
confi  
gure  
d  
earli  
er.

2. Error  
Handling  
and  
Logging  
(Subjob  
3 - On  
Failure):

a. In  
case  
the  
data  
extra  
ction  
subj  
ob  
(Sub  
job  
2)  
enco  
unte  
rs  
any  
issu  
es  
or  
does  
n't  
finis  
h  
succ  
essf  
ully,  
the  
job  
trans  
ition  
s to  
a  
subj  
ob  
for  
writi  
ng  
error  
logs.

b. Logg  
ing  
is  
cruci  
al  
for  
track  
ing  
and  
diag  
nosi  
ng  
probl  
ems  
durin  
g  
the  
data  
extra  
ction

process.

Data Loading  
and  
Transformation:

1. Data Loading and Transformation (Subjob 4 - On Success):
  - a. When the data extraction subjob (Subjob 2) successfully completes for the provided filename, the job proceeds with data loading and transformation.
  - b. A subjob is called to load data from the CSV file into a stage table, preparing it for further processing.
  - c. Subsequently

- , data from the stage table is loaded into an operational data store (ODS) table.
- d. All necessary parameters, such as table names and connection parameters, are provided to ensure accurate data extraction and loading.

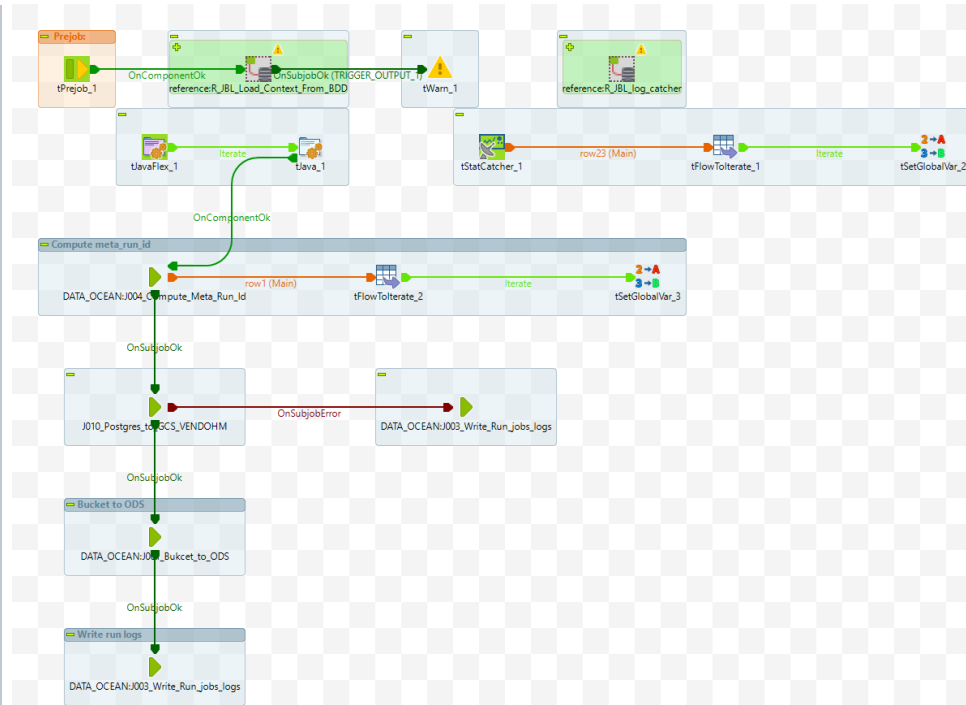
#### Logging and Reporting:

1. Logging (Subjob 5):
  - a. At the end of the data loading and transformation process, a subjob is called to

|                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                              |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
|                                 | <p>write logs.</p> <p>b. Logging captures crucial information about the data processing, enhancing transparency and traceability.</p> <p>This Talend job streamlines the data extraction, loading, and transformation process. It begins with file name configuration, proceeds with metadata capture, extracts data from Google Sheets, handles errors if they occur, and performs data loading and transformation, all while maintaining detailed logs for monitoring and maintaining data integrity.</p> |                                                                              |
| Main jobs for source extraction | <ul style="list-style-type: none"> <li><del>J010-Extraction till ODS-VE NDOHM</del></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                              | <p>--to the top --</p> <p>GCP Remote Engine</p> <p>Industrial DO project</p> |
|                                 | Job description by steps                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Job design                                                                   |
|                                 | <p><del>Global Variable Configuration:</del></p> <p>1. <del>File Names</del></p>                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                              |

and Table Names Configuration (Global Variables, postgresTableNames):

- The job starts by configuring global variables for file names and table names.
- These variables are populated by reading a list of all tables and files that subjects need to load.
- This dynamic configuration ensures that the job adapts to the specific files and tables required for processing.



Data-  
Extraction-  
and-  
Processing in  
Iterations:

1. Iteration-  
Loop-  
(One-by-  
One):

- a. The-  
job-  
proce-  
sses-  
table-  
s-  
and-  
files-  
iterat-  
ively-  
r-  
one-  
by-  
one-  
base-  
d on-  
the-  
confi-  
gure-  
d-  
glob-  
al-  
varia-  
bles.

2. Job Run-  
Initializati-  
on-  
(Subjob-  
1):

- a. Withi-  
n-  
each  
iterat-  
ion-  
the-  
first-  
subj-  
ob-  
is-  
calle-  
d to-  
retri-  
eve-  
the-  
job-  
ID-  
and-  
calc-  
ulate  
the-  
extra-  
ction  
date.
- b. This-  
initia-  
lizati-  
on-  
step-  
capt-  
ures-  
vital-  
meta-  
data-  
for-  
the-  
ongo

ing-  
data-  
proc  
essi  
ng-

3. Files-  
Extraction  
from-  
Postgres  
QL-  
Database  
(Subjob-  
2)-

a. In-  
the-  
next-  
subj  
ob-  
data-  
files-  
are-  
extra  
cted-  
from  
the-  
Post  
gres  
QL-  
data  
base-

b. For  
h-  
iterat  
ion-  
targ  
ets-  
a-  
spec  
ific-  
file-  
base  
d on-  
the-  
glob  
al-  
varia  
bles-  
(file  
nam  
es)-  
confi  
gure  
d-  
earli  
er-

4. Error-  
Handling-  
and-  
Logging-  
(Subjob-  
3-On-  
Failure)-

a. If  
the-  
data-  
extra  
ction  
subj  
ob-  
(Sub  
job-  
2)-  
enco  
unte  
rs-  
any-  
issu  
es-

or does not finish successfully, the job transitions to a subject for writing error logs.

b. Logging is crucial for tracking and diagnosing problems during the data extraction process.

5. Data Loading and Transformation (Subjob 4 On Success):

a. When the data extraction subject (Subjob 2) successfully completes for the provided filename, the job proceeds with

data-  
loadi-  
ng-  
and-  
trans-  
form-  
ation-

b. A-  
subj-  
ob-  
is-  
calle-  
d-to-  
load-  
data-  
from-  
the-  
CSV-  
file-  
into-  
a-  
stag-  
e-  
table  
-  
prep-  
aring-  
it-for-  
further-  
proce-  
ssing-

c. Sub-  
sequ-  
ently  
-  
data-  
from-  
the-  
stag-  
e-  
table  
is-  
load-  
ed-  
into-  
an-  
oper-  
ation  
at-  
data-  
store  
(OD-  
S)-  
table-

d. All-  
neces-  
sary-  
para-  
mete-  
rs,  
such-  
as-  
table  
nam-  
es-  
and-  
conn-  
ectio-  
n-  
para-  
mete-  
rs,  
are-  
provi-  
ded-  
to-

ensu  
re-  
accu  
rate-  
data-  
extra  
ction  
and-  
loadi  
ng-

Logging and  
Reporting:

1. Logging-  
(Subjob-  
5):
  - a. At-  
the-  
end-  
of-  
each  
iterat  
ion,-  
a-  
subj  
ob-  
is-  
calle  
d-to-  
write  
logs-
  - b. Logg  
ing-  
capt  
ures-  
critic  
al-  
infor  
mati  
on-  
abou  
t the-  
data-  
proce  
ssi  
ng,-  
ensu  
ring-  
trans  
pare  
ncy-  
and-  
trace  
abilit  
y-for-  
each  
proce  
sse  
d-  
file-  
and-  
table-

This Talend  
job is-  
designed for-  
dynamic data-  
extraction and-  
processing,-  
adapting to-  
different  
tables and-  
files. It  
initializes-  
each iteration-  
with metadata-  
capture,-  
extracts data-  
from the-

|  |                                                                                                                                                                                                                                 |  |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | PostgreSQL database, handles errors when necessary, performs data loading and transformation with parameterized configurations, and maintains detailed logs for monitoring and data integrity across multiple files and tables. |  |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

#### 4.4 - Load to DM (calculations and transformations)

|                                 |                                                                                                                                                                                                                                                       |                                                                               |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| Main jobs for source extraction | <ul style="list-style-type: none"> <li>J001_ODS_TO_DM_FACT_ENERGY_PRICE_FORECAST</li> </ul>                                                                                                                                                           | <a href="#">--to the top --</a><br>GCP Remote Engine<br>Industrial DO project |
|                                 | Job description by steps                                                                                                                                                                                                                              | Job design                                                                    |
|                                 | Metadata Calculation and Cache Initialization:<br><br>1. Job Metadata Calculation:<br><ul style="list-style-type: none"> <li>The job begins by calculating essential job metadata, which include information such as job identifiers, time</li> </ul> |                                                                               |

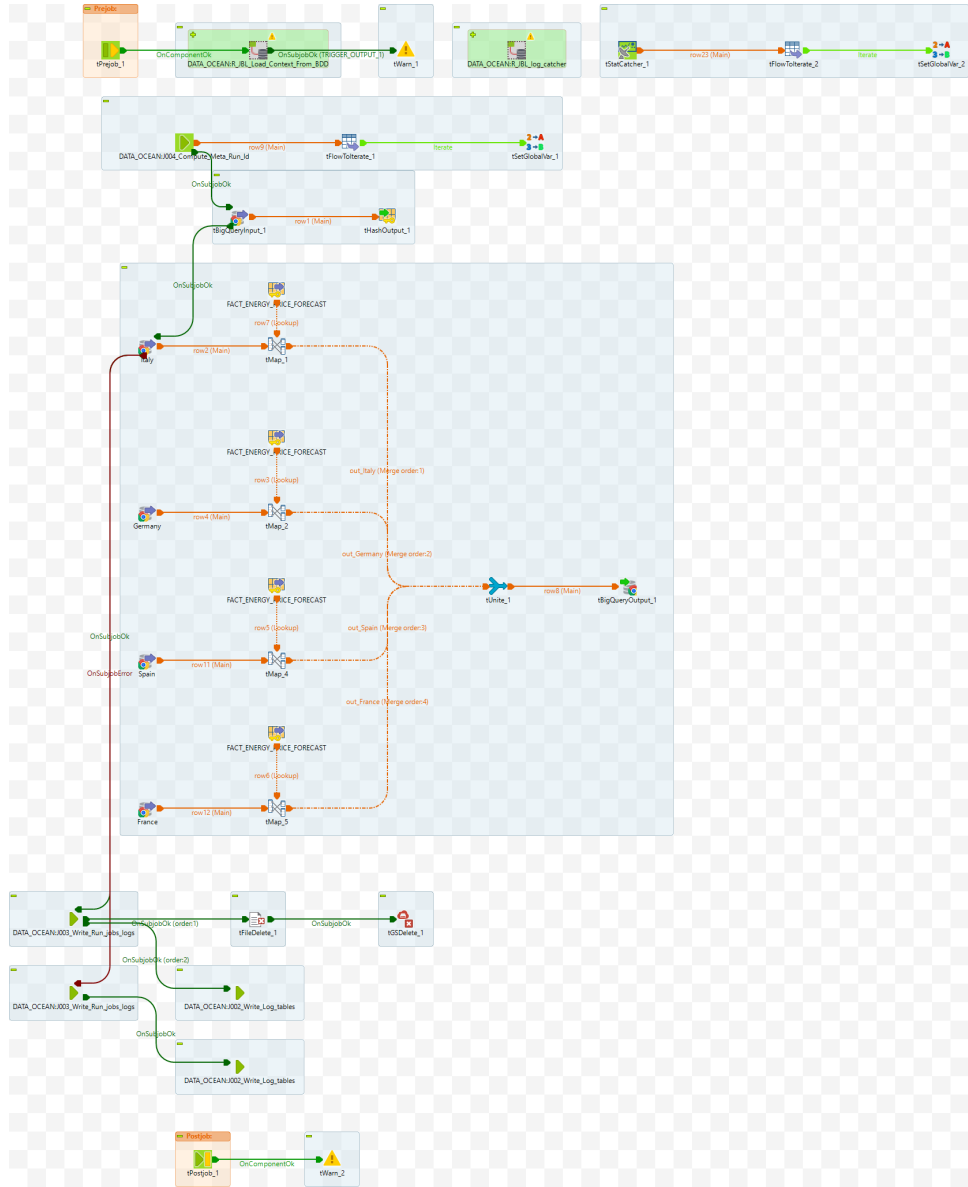
- stamps, or configuration details.
- This step sets the stage for the subsequent data processing.

## 2. Fact Table Keys Calculation and Caching:

- Following meta data calculation, the job proceeds to compute and store keys associated with the current Fact table.
- These keys are cached for efficient reference during data processing.

## Data Extraction and Import:

1. Energy Forecast



Data

Extraction:

- The job extracts energy forecast data from four Operational Data Store (ODS) tables, each containing information for a specific region (Italy, Germany, Spain, and France).

2. Data Comparison and Import:

- The extracted energy forecast data is compared with the cached keys from the Fact table.
- The primary purpose of this step

is to identify and import only new rows of data for the latest forecasted date.

- By importing only new data, the job ensures that the data set remains up to date and avoids unnecessary duplication.

#### Logging and Cleanup:

##### 1. Log Generation:

- At the conclusion of data import and processing, the job generates logs.
- These logs provide a record of the

|                                 |                                                                                                                                                                                                                                                                                                                                                 |                                                                                          |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
|                                 | <div>job's activities, facilitating monitoring and troubleshooting.</div> <div>2. Temporary File Deletion:<ul style="list-style-type: none"><li>As a final step, the job deletes any temporary files created during the data transfer process.</li><li>This cleanup helps manage storage resources and maintain data integrity.</li></ul></div> |                                                                                          |
| Main jobs for source extraction | <div><ul style="list-style-type: none"><li><del>J002_O</del><br/><del>DS_TO</del><br/><del>_DM_DI</del><br/><del>M_MET</del><br/><del>EO_INFO</del></li></ul></div>                                                                                                                                                                             | <div>--to the top --</div> <div>GCP Remote Engine</div> <div>Industrial DO project</div> |
|                                 | <div><del>Job-</del><br/><del>description-</del><br/><del>by-steps</del></div>                                                                                                                                                                                                                                                                  | Job design                                                                               |
|                                 | <div>Metadata Calculation:</div> <div>1. <del>Job-</del><br/><del>Metadata-</del></div>                                                                                                                                                                                                                                                         |                                                                                          |

## Calculation

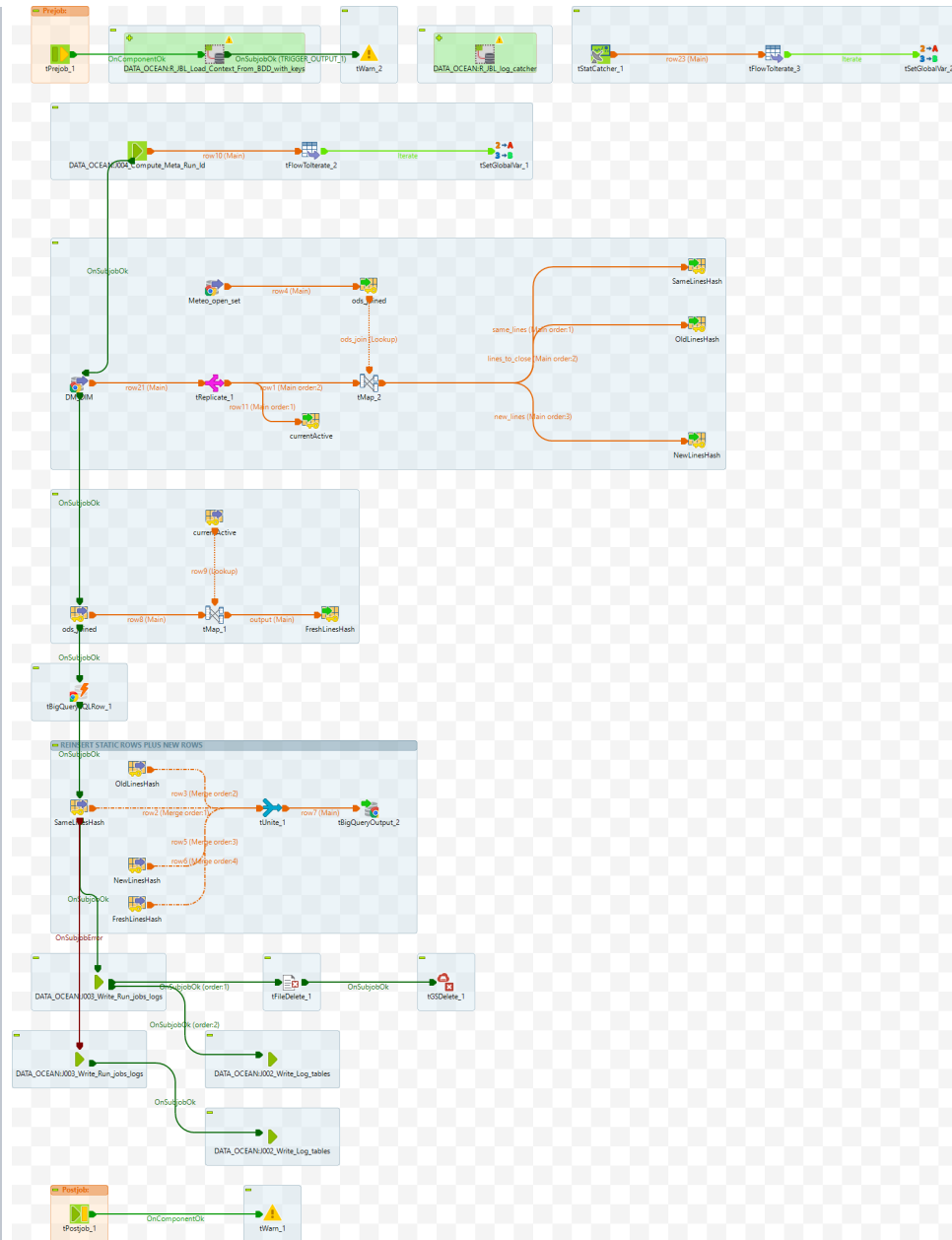
1. The job

- initiates by calculating essential job meta data. This meta data includes information on job identifiers, timestamps, or job specific configuration details.
- This step sets the foundation for the subsequent data processing.

## Slowly-Changing-Dimension-(SCD2)-Data-Loading:

- SCD2-Logic for Dimension Data Loading:

- In this step, the job implements Slowly-Changing



g-  
Dim  
ensi  
on-  
Type  
2-  
(SC  
D2)-  
logie.

- The  
purp  
ose-  
of  
this  
logie  
is to  
load  
data-  
into-  
a  
dime  
nsio  
n-  
table  
-  
spee  
ificall  
y-  
from  
the  
Oper  
ation  
at  
Data  
Stor  
e  
(OD  
S)-  
mete  
o-  
table  
into-  
the  
Data  
Mart  
(DM)  
mete  
o-  
table.
- The  
SCD  
2-  
logie  
is-  
appli  
ed-  
by-  
com  
parin  
g-  
multi  
ple-  
field  
s-  
acro  
ss-  
the  
two-  
table  
s,-  
such  
as-  
nam  
e,-  
data  
base  
nam  
e,-le-  
grou

p-,  
asse  
t  
clas  
s-,  
asse  
t  
sub-  
clas  
s-,  
asse  
t  
type-,  
capa  
city-  
unit-,  
com  
modi  
ty-,  
curv  
e-  
publi  
shin  
g-,  
curv  
e-  
type-,  
hub-,  
mark  
et-,  
obje  
ct  
type-,  
own  
er-,  
regio  
n-,  
unit-,  
and  
volu  
me-  
unit.

#### Logging and Cleanup:

1. Log-  
Generatio  
n:
  - Upo  
n  
com  
pleti  
on  
of  
the  
data-  
loadi  
ng-  
proce  
ss-,  
the  
job  
gene  
rates  
logs. Thes  
e  
logs  
serv  
e as  
a  
reco  
rd  
of  
the  
job's  
activ  
ities,  
provi

|                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                             |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|
|                      | <div>ding-<br/>visibi-<br/>lity-<br/>into-<br/>the-<br/>data-<br/>loadi-<br/>ng-<br/>proc-<br/>ess.</div> <div>2. Temporary File Deletion:<ul style="list-style-type: none"><li>As a final step, the job takes care of deleting any temporary files created during the data transfer process. This cleanup ensures the efficient management of storage resources.</li></ul></div> <div>3. In case of error to save data in BQ, if the last data to save to BQ will save to /DATA /DEV/IND /ROBUST /FY/nOut /DM/ [meta_run_id].csv</div> |                                                             |
| Main jobs for source | <div><ul style="list-style-type: none"><li>• J003_QDS_TO</li></ul></div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <div><a href="#">to the top</a><br/>GCP Remote Engine</div> |

|            |                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| extraction | <div><div>DM-DI</div><div>M-ENE</div><div>RGY-P</div><div>RICE-I</div><div>NFO</div></div>                                                                                                                                                                                                                                                                                                                                                   | Industrial-DO-project                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|            | Job-description-by-steps                                                                                                                                                                                                                                                                                                                                                                                                                     | Job-design                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|            | <div>Metadata-Calculation:</div> <div>1. Job-Metadate-Calculation:</div> <div><ul style="list-style-type: none"><li>The job-initiates-by-calculation-essential-job-metadate. This-metadate-includes-information-like-job-identifier-time-stamp-or-job-speed-the-configurational-detail.</li><li>This-step-sets-the-foundations-for-the-subsequent-data-processing.</li></ul></div> <div>Slowly-Changing-Dimension-(SCD2)-Data-Loading:</div> | <p>The diagram illustrates a metadata calculation job flow. It begins with a 'Prestage' block (tPrestage_1) which triggers an 'OnSubjobOk' event and initiates the loading of content from a source (DATA_OCEAN002_Load_Content_from_BSD_with_Maps). This leads to a block for computing metadata (DATA_OCEAN002_Compute_Meta_Run_Id), which then flows through a tolerance check (tFlowTolerate_2) and an iteration loop (Iterate) to set global variables (tSetGlobalVar_1). The core of the job is a large processing block starting with 'OnSubjobOk'. It branches into two main paths: 'Forwards' and 'Spot'. The 'Forwards' path involves merging data (row4), joining (tJoin_3), mapping (tMap_1), and joining (ods_join) to calculate 'lines_to_close' and update 'NewLinesHash'. The 'Spot' path involves merging (row19), mapping (tMapCalc_1), and joining (tMap_2) to also calculate 'lines_to_close' and update 'NewLinesHash'. Other components include a 'tBigQueryOutput_1' block, a 'tMap_3' block for 'output (Main)' leading to 'FreshLinesHash', and a 'tBigQueryOutput_2' block. The flow concludes with 'DATA_OCEAN002_Write_Run_Jobs_Logs', a 'tFileDelete_1' block, 'DATA_OCEAN002_Write_Log_tables', and a final 'tWarn_1' block.</p> |

1. SCD2-  
Logic for  
Dimension  
Data  
Loading:

- In this step, the job applies Slowly Changing Dimension Type 2 (SCD2) logic to load data into a dimension table.
- The source of the data is the Operational Data Store (ODS) energy price metadata table, which is a union of tables named "forwards" and "spot". The data is loaded into the

- SCD 2 methodology involves partitioning specific fields, such as name, company, modify, contract, proprietary, time span, rolling, and time zone, to identify changes and maintain historical records in the DM table.

#### Logging and Cleanup:

1. Log Generation:
  - Upon the completion of the data load

ng-  
proe  
ess,  
the  
job  
gene  
rates  
logs.  
Thes  
e  
logs  
serv  
e as  
a  
reco  
rd  
of  
the  
job's  
activ  
ities  
and  
provi  
de  
visibi  
lity  
into  
the  
data  
loadi  
ng  
proe  
ess.

## 2. Temporary File Deletion:

- As  
a  
final  
step,  
the  
job  
take  
s  
care  
of  
delet  
ing  
any  
temp  
orar  
y  
files  
creat  
ed  
durin  
g  
the  
data  
trans  
fer  
proe  
ess.  
This  
clea  
nup  
ensu  
res  
the  
effici  
ent  
man  
age  
ment  
of  
stora  
ge.

|  |                                                                                                                                                                                                                                                |  |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | <div>resou<br/>urces<br/>s-<br/>3. In case<br/>of error<br/>to save<br/>data in<br/>BQ, if the<br/>last data<br/>to save<br/>to BQ<br/>will save<br/>to /DATA<br/>/DEV/IND<br/>/ROBUST<br/>/FY/InOut<br/>/DM/<br/>[meta_ru<br/>n_id].csv</div> |  |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

|                                 |                                                                                                                      |                                                                                       |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Main jobs for source extraction | <div><ul style="list-style-type: none"><li>• J004_O<br/>DS_TO<br/>_DM_F<br/>ACT_M<br/>ETEO_<br/>DATA</li></ul></div> | <div><a href="#">to the top</a><br/>GCP Remote Engine<br/>Industrial DO project</div> |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|

|  |                          |            |
|--|--------------------------|------------|
|  | Job description by steps | Job design |
|--|--------------------------|------------|

|  |                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <div>Metadata Calculation:<br/>1. Job Metadata Calculation:<br/><ul style="list-style-type: none"><li>• The job initiates by calculating essential job metadata. This metadata includes information like job identifier, time stamps, or job specific configuration</li></ul></div> | <p>The diagram illustrates a metadata calculation pipeline. It begins with a 'PreJob' block containing 'tPostJob_1' and 'OnComponentOk' leading to 'DATA_OCEAN_RBL_Load_Content_From_BDO_with_Java'. This triggers 'OnSubjobOk (TRIGGER_OUTPUT_1)' leading to 'tWarn_1'. The main flow goes to 'DATA_OCEAN_XXX_Compute_Meta_Run_Id' (row1 Main), then 'tFlowToIterate_1', 'Iterate', and 'tSetGlobalVar_1'. A parallel path goes through 'tStatCatcher_1', 'row23 (Main)', 'tFlowToIterate_2', 'Iterate', and 'tSetGlobalVar_2'. The main path continues to 'DATA_OCEAN_XXX_Compute_Meta_Run_Id' (row2 Main), then 'tMap_1', 'out (Main)', 'tMap_2', 'out1 (Main)', and 'tBigQueryOutput_1'. This triggers 'OnSubjobOk' leading to a 'Data' block with 'DM_DIM_METEO_INFO' (row2 Backup), 'DM_FACT_METEO_DATA' (row3 Backup), and 'DM_DIM_METEO_DATA' (row1 Main). This triggers 'OnSubjobOk' leading to a 'Data' block with 'DATA_OCEAN_XXX_Write_Run_Job_Logs' (order:1), 'tFileDelete_1', 'OnSubjobOk', and 'tGSDDelete_1'. This triggers 'OnSubjobOk' leading to a 'Data' block with 'DATA_OCEAN_XXX_Write_Run_Job_Logs' (order:2), 'DATA_OCEAN_XXX_Write_Log_tables', and 'DATA_OCEAN_XXX_Write_Log_tables'. This triggers 'OnSubjobOk' leading to a 'PostJob' block with 'tPostJob_1', 'OnComponentOk', and 'tWarn_2'.</p> |
|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

- This step establishes the foundation for the subsequent data processing.

#### Data Enrichment and Loading:

1. Data Enrichment from DIM\_metadata\_info:
  - In this step, the job enriches data from the Vendor metadata table with attributes obtained from the DIM\_metadata\_info table.
2. Loading into FACT\_metadata\_data:
  - The enriched data is then loaded into the FACT\_metadata\_data table.

- This table serves as a fact table containing data relevant to meteorological information.

#### Logging and Cleanup:

##### 1. Log Generation:

- Upon the completion of the data enrichment and loading process, the job generates logs. These logs serve as a record of the job's activities and provide visibility into the data processing steps.

##### 2. Temporary File Deletion:

- As a

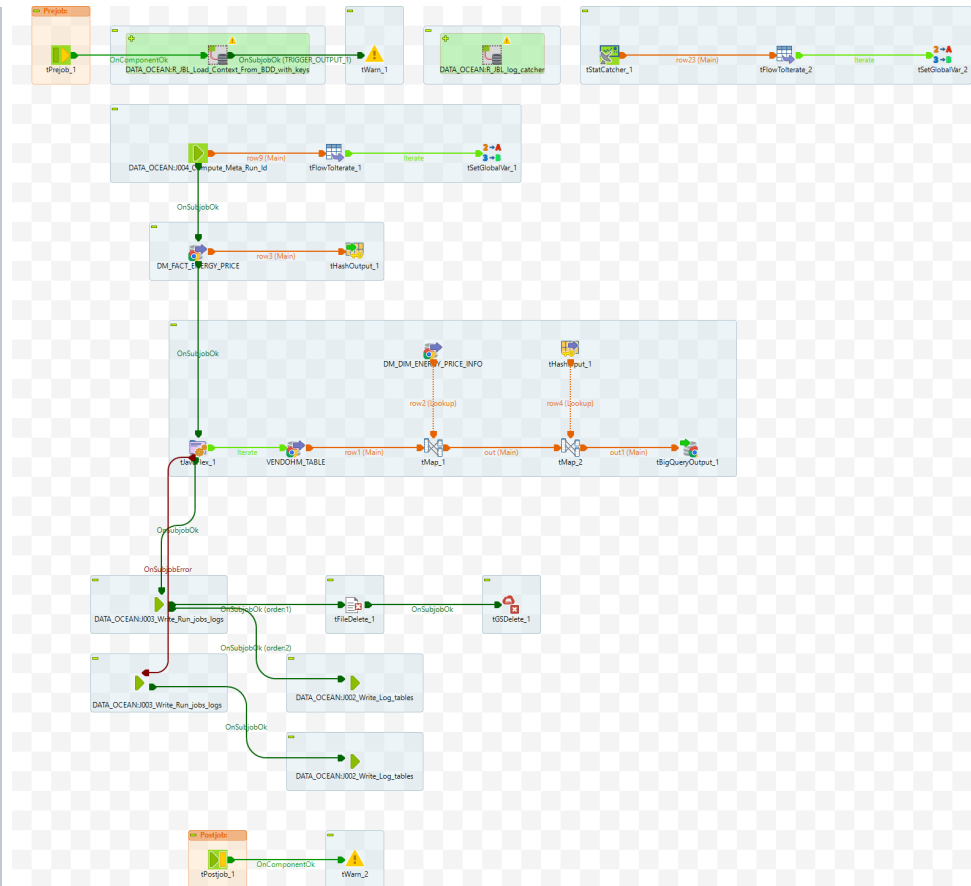
|                                          |                                                                                                                                                                                                                                                                                                                                                                                                                 |            |
|------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
|                                          | <div>final<br/>step,<br/>the<br/>job<br/>take<br/>s-<br/>care-<br/>of<br/>delet<br/>ing-<br/>any-<br/>temp<br/>orar<br/>y-<br/>files-<br/>creat<br/>ed-<br/>durin<br/>g-<br/>the-<br/>data-<br/>trans<br/>fer-<br/>and<br/>enrie<br/>hme<br/>nt<br/>proe<br/>ess-<br/>This-<br/>clea<br/>nup-<br/>ensu<br/>res-<br/>effici<br/>ent<br/>stora<br/>ge-<br/>reso<br/>urce-<br/>man<br/>age<br/>ment.</div>         |            |
| Main<br>jobs for<br>source<br>extraction | <div><div><div>• <del>J005_Q</del><br/><del>DS_TO</del><br/><del>_DM_F</del><br/><del>ACT_E</del><br/><del>ENERGY</del><br/><del>_PRICE</del></div><div>replace<br/>with<br/>view in<br/>Robust<br/>ify prj-<br/>data-<br/>robusti<br/>fy-dev.<br/>DM.<br/>V_FAC<br/>T_ener<br/>gy_pric<br/>e_hourly</div></div></div> <div>--to the top --</div> <div>GCP Remote Engine</div> <div>Industrial DO project</div> |            |
|                                          | Job<br>description<br>by steps                                                                                                                                                                                                                                                                                                                                                                                  | Job design |
|                                          | Metadata<br>Calculation:                                                                                                                                                                                                                                                                                                                                                                                        |            |

# 1. Job- Metadata- Calculation- #:

- The job commences with the calculation of essential job metadata. This metadata includes information like job identifiers, timestamps, or specific configuration details.
- This initial step provides a foundation for subsequent data processing.

## Data- Enrichment- and Loading:

1. Data-  
Enrichment-  
from-  
DIM\_ener-  
gy\_price\_-  
info:
  - In this step, the job enrie



hes-  
data-  
from  
the-  
ener-  
gy-  
price  
table  
by-  
incor-  
pora-  
ting-  
attrib-  
utes-  
sour-  
ced-  
from  
the-  
DIM  
\_ene-  
rgy\_  
price  
\_info  
table-

2. Loading-  
into-  
FACT\_en-  
ergy\_pric-  
e\_data:

- The-  
enrie-  
hed-  
data-  
is-  
subs-  
eque-  
ntly-  
load-  
ed-  
into-  
the-  
FAG  
T\_e-  
norg  
y\_pri-  
ce\_d  
ata-  
table-
- The-  
FAG  
T\_e-  
norg  
y\_pri-  
ce\_d  
ata-  
table  
is-  
desi-  
gned  
to-  
store  
data-  
relev-  
ant  
to-  
ener-  
gy-  
price  
s-

Logging and-  
Cleanup:

1. Log-  
Generatio-  
n:

- Follo-  
wing  
the-  
com-

pletion of the data entry and loading process, the job generates logs. These logs serve as a record of the job's activities and provide visibility into the data processing steps.

## 2. Temporary File Deletion:

- As a final step, the job manages the deletion of any temporary files that were created during the data transfer and entry process.

|  |                                                                                                                    |  |
|--|--------------------------------------------------------------------------------------------------------------------|--|
|  | ess-<br>This<br>clea<br>rup<br>ensu<br>res<br>effici<br>ent<br>stora<br>ge-<br>reso<br>urce-<br>man<br>age<br>ment |  |
|--|--------------------------------------------------------------------------------------------------------------------|--|

|                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Main<br>jobs for<br>source<br>extraction | <ul style="list-style-type: none"><li>J006_O<br/>DS_TO<br/>_DM_F<br/>ACT_C<br/>O2_EMI<br/>SSIONS</li></ul>                                                                                                                                                                                                                                                                                                                                            | --to the top --<br><br>GCP Remote Engine<br><br>Industrial DO project                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|                                          | Job<br>description<br>by steps                                                                                                                                                                                                                                                                                                                                                                                                                        | Job design                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|                                          | Metadata<br>Calculation:<br><br>1. Job<br>Metadata<br>Calculation:<br><ul style="list-style-type: none"><li>The<br/>job<br/>com<br/>men<br/>ces<br/>by<br/>calcula<br/>ting<br/>esse<br/>ntial<br/>job<br/>meta<br/>data. This<br/>meta<br/>data<br/>inclu<br/>des<br/>infor<br/>mati<br/>on<br/>such<br/>as<br/>job<br/>ident<br/>ifiers<br/>,<br/>time<br/>stamps,<br/>or<br/>spec<br/>ific<br/>confi<br/>gura<br/>tion<br/>detai<br/>ls.</li></ul> | <pre>graph TD     Prejob[tPrejob_1] -- OnComponentOk --&gt; Load[DATA_OCEAN:R_IBL_Load_Context_From_BDD_with_keys]     Load -- OnSubjobOk (TRIGGER_OUTPUT_1) --&gt; Warn1[tWarn_1]     Warn1 --&gt; Catcher[DATA_OCEAN:R_IBL_log_catcher]     Catcher --&gt; StatCatcher[tStatCatcher_1]     StatCatcher -- row23 (Main) --&gt; FlowTol2[tFlowTolterate_2]     FlowTol2 -- Iterate --&gt; SetVar2[tSetGlobalVar_2]     SetVar2 --&gt; Compute[DATA_OCEAN:J004_Compute_Meta_Run_Id]     Compute -- row9 (Main) --&gt; FlowTol1[tFlowTolterate_1]     FlowTol1 -- Iterate --&gt; SetVar1[tSetGlobalVar_1]     SetVar1 -- OnSubjobOk --&gt; SQLRow[tBigQuerySQLRow_1]     SQLRow -- OnComponentOk --&gt; CO2[CO2]     CO2 -- row1 (Main) --&gt; Map2[tMap_2]     Map2 -- out1 (Main) --&gt; BigQueryOut[tBigQueryOutput_1]     BigQueryOut -- OnSubjobOk --&gt; WriteLogs[DATA_OCEAN:J003_Write_Run_Jobs_Logs]     WriteLogs -- OnSubjobError --&gt; FileDelete[tFileDelete_1]     FileDelete -- OnSubjobOk --&gt; GDelete[tGDelete_1]     WriteLogs -- OnSubjobOk (order:1) --&gt; WriteLogTables1[DATA_OCEAN:J002_Write_Log_tables]     WriteLogs -- OnSubjobOk (order:2) --&gt; WriteLogTables2[DATA_OCEAN:J002_Write_Log_tables]     WriteLogTables1 -- OnSubjobOk --&gt; End[End]     WriteLogTables2 -- OnSubjobOk --&gt; End     End --&gt; Postjob[tPostjob_1]     Postjob -- OnComponentOk --&gt; Warn2[tWarn_2]</pre> |

- This initial step provides a foundation for subsequent data processing.

Latest CO2  
Data Refresh:

1. Refreshing/Providing Latest CO2 Data:

- In this step, the job focuses on refreshing or providing the latest CO2 emissions data.
- The source of this data is the Operational Data Store (ODS) table, which serves as a central repository for CO2-related

ed  
infor  
mati  
on.

Logging and  
Cleanup:

1. Log  
Generatio  
n:

- Follo  
wing  
the  
com  
pleti  
on  
of  
the  
CO2  
data  
refre  
sh,  
the  
job  
gene  
rates  
logs. Thes  
e  
logs  
serv  
e as  
a  
reco  
rd  
of  
the  
job's  
activ  
ities  
and  
provi  
de  
visibi  
lity  
into  
the  
data  
refre  
sh  
proc  
ess.

2. Temporar  
y File  
Deletion

- As  
a  
final  
step,  
the  
job  
man  
ages  
the  
delet  
ion  
of  
any  
temp  
orar  
y  
files  
that  
were  
creat  
ed  
durin  
g  
the

|  |                                                                                    |  |
|--|------------------------------------------------------------------------------------|--|
|  | data transfer process. This cleanup ensures efficient storage resource management. |  |
|--|------------------------------------------------------------------------------------|--|

|                                 |                                                                                          |                                                               |
|---------------------------------|------------------------------------------------------------------------------------------|---------------------------------------------------------------|
| Main jobs for source extraction | <ul style="list-style-type: none"><li>J007_O DS_TO _DM_F ACT_S OLID_F UEL_W AP</li></ul> | --to the top --<br>GCP Remote Engine<br>Industrial DO project |
|---------------------------------|------------------------------------------------------------------------------------------|---------------------------------------------------------------|

|  |                          |            |
|--|--------------------------|------------|
|  | Job description by steps | Job design |
|--|--------------------------|------------|

|  |                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p>Metadata Calculation:</p> <p>1. Job Metadata Calculation:</p> <ul style="list-style-type: none"><li>The job commences by calculating essential job metadata. This metadata includes information such as job identifiers, time stamps, or specific confi</li></ul> | <pre>graph TD     PreJob_1[PreJob_1] -- OnComponentOk --&gt; LoadContext[DATA_OCEANR_JBL_Load_Context_From_BDD_with_keys]     LoadContext -- OnSubjobOk TRIGGER_OUTPUT_1 --&gt; Warn_1[Warn_1]     Warn_1 --&gt; LogCatcher[DATA_OCEANR_JBL_log_catcher]     LogCatcher --&gt; tStatCatcher_1[tStatCatcher_1]     tStatCatcher_1 -- row23 (Main) --&gt; tFlowTolerate_2[tFlowTolerate_2]     tFlowTolerate_2 -- Iterate --&gt; tSetGlobalVar_2[tSetGlobalVar_2]     LogCatcher --&gt; row9(row9 (Main))     row9 --&gt; tFlowTolerate_1[tFlowTolerate_1]     tFlowTolerate_1 -- Iterate --&gt; tSetGlobalVar_1[tSetGlobalVar_1]     tSetGlobalVar_1 -- OnSubjobOk --&gt; tQuerySQLRow_1[tQuerySQLRow_1]     tQuerySQLRow_1 -- OnComponentOk --&gt; MARGINAL_COST[MARGINAL_COST]     MARGINAL_COST -- row1 (Main) --&gt; tMap_2[tMap_2]     tMap_2 -- out1 (Main) --&gt; tBigQueryOutput_1[tBigQueryOutput_1]     tQuerySQLRow_1 -- OnSubjobOk --&gt; WriteJobsLogs[DATA_OCEANR_JBL_Write_Run_Jobs_Logs]     WriteJobsLogs -- OnSubjobOk (order:1) --&gt; tFileDelete_1[tFileDelete_1]     tFileDelete_1 -- OnSubjobOk --&gt; tGDelete_1[tGDelete_1]     WriteJobsLogs -- OnSubjobOk (order:2) --&gt; WriteLogTables1[DATA_OCEANR_JBL_Write_Log_Tables]     WriteLogTables1 -- OnSubjobOk --&gt; WriteLogTables2[DATA_OCEANR_JBL_Write_Log_Tables]     PostJob_1[PostJob_1] -- OnComponentOk --&gt; Warn_2[Warn_2]</pre> |
|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

- guration details.
- This initial step provides a foundation for subsequent data processing.

Latest wap  
solid fuel data  
Refresh:

1. Refreshing /Providing Latest Cost Data:
  - In this step, the job focuses on refreshing or providing the latest weighted average fuel prices data.
  - The source of this data is the Operational Data Store (ODS) table, which serves as a

central repository for wap solid fuel - related information.

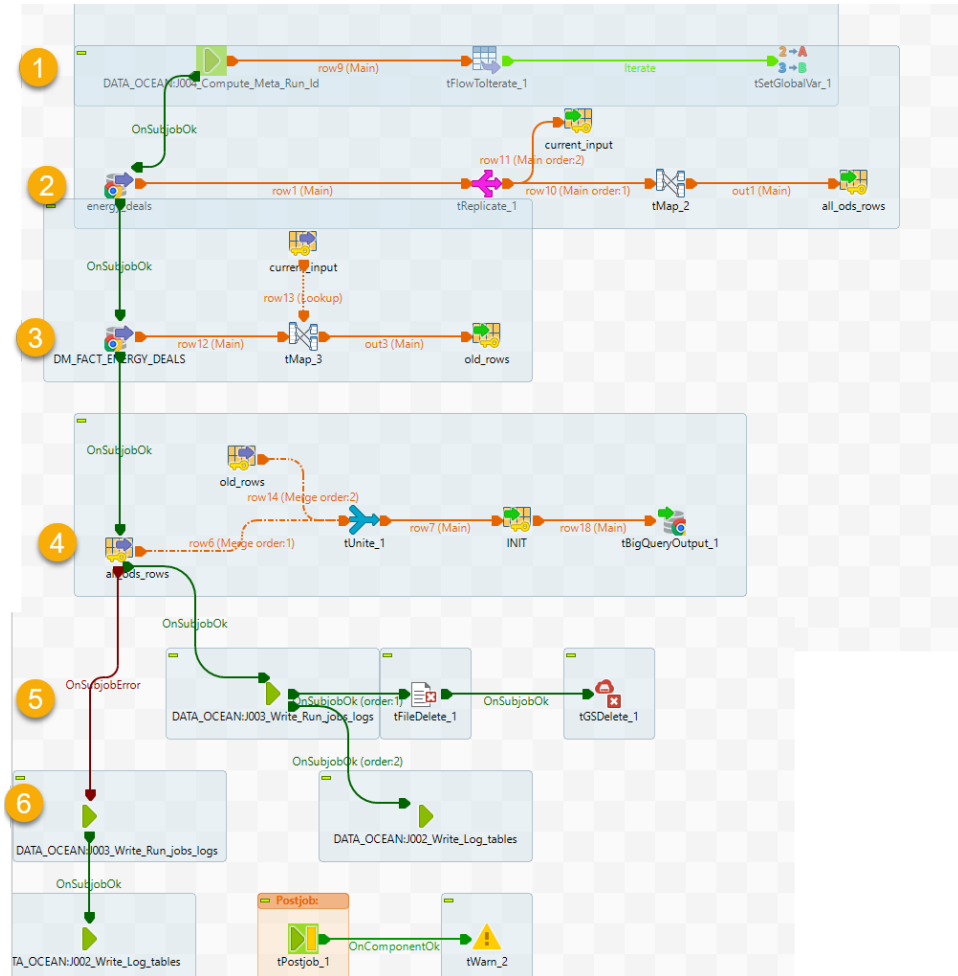
Logging and Cleanup:

1. Log Generation:
  - Following the completion of the Wap solid fuel data refresh, the job generates logs. These logs serve as a record of the job's activities and provide visibility into the data refresh process.
2. Temporary File Deletion
  - As a final step, the job manages the deletion of any

|                                 |                                                                                                                                                                     |                                                                              |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
|                                 | <p>temporary files that were created during the data transfer process. This cleanup ensures efficient storage resource management.</p>                              |                                                                              |
| Main jobs for source extraction | <ul style="list-style-type: none"><li><del>J009_ODS_TO_DM_FACT_ENERGY_DEALS</del> J009_ODS_TO_DM_FACT_IRM_DEALS</li></ul> <p>(DATA_OCEAN_DOMAIN_SUSTAINABILITY)</p> | <p>--to the top --</p> <p>GCP Remote Engine</p> <p>Industrial DO project</p> |
|                                 | Job description by steps                                                                                                                                            | Job design                                                                   |
|                                 | <p>1. Job Metadata Calculation:</p> <ul style="list-style-type: none"><li>The job initiates by calculating essential job metadata. This metadata</li></ul>          |                                                                              |

include information like job identifiers, timestamps, or job-specific configuration details.

- This step establishes the foundation for the subsequent data processing.
2. Get full load from ODS
  3. Get data from DM. FACT\_irm\_energy\_deals\_daily in case there is some data, which not exist in ODS
  4. Combine data from 2 and 3 to DM. FACT\_irm\_energy\_deals\_daily with truncate option
  5. Log Generation:
    - Upon the completion of the data



enrichment and loading process, the job generates logs. These logs serve as a record of the job's activities and provide visibility into the data processing steps.

6. Temporary File Deletion:

- As a final step, the job takes care of deleting any temporary files created during the data transfer and enrichment process. This cleanup ensures effici

|  |                                                           |  |
|--|-----------------------------------------------------------|--|
|  | ent<br>stora<br>ge<br>reso<br>urce<br>man<br>age<br>ment. |  |
|--|-----------------------------------------------------------|--|

Main jobs for source extraction

• J001\_DM\_TO\_DM\_FACT\_ENERGY\_PRICE\_HOURLY

(ROBUSTIFY)

Job description by steps

Data Extraction and Transformation:

1. The code extracts data related to energy prices and contracts from a database.

The source table is DataOcean.V\_FACT\_energy\_irm

To get the raw data from IRM in ODS and transform to the format that Dataiku can consume

Split data into hourly

The past to current date will get from Spot

The current month will get from the first date of the next month

Next month data will be all the same within the month by the first date of each month

Explanation the job:

1. Setup global variable on output file / source table and define the value of gas = 0.0

2. Get data from the source with format below

| delivery_date           | elec_DE_EPEX | elec_FR_EPEX | elec_IT_PUN | elec_SP_OMEI | gas_DE_THE | gas_FR_PEO | gas_IT_PSV | gas_SP_PVB | gas_TTF |
|-------------------------|--------------|--------------|-------------|--------------|------------|------------|------------|------------|---------|
| 2024-06-21 21:00:00 UTC | 105.26       | 105.14       | 135.3151    | 114.86       | 0.0        | 0.0        | 0.0        | 0.0        | 0.0     |
| 2024-06-21 22:00:00 UTC | 95.99        | 96.49        | 115.0116    | 115.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0     |
| 2024-06-21 23:00:00 UTC | 82.14        | 83.26        | 112.0       | 105.74       | 0.0        | 0.0        | 0.0        | 0.0        | 0.0     |
| 2024-07-01 00:00:00 UTC | 73.15        | 50.94        | 106.9       | 74.56        | 34.5       | 34.31      | 36.15      | 34.45      | 34.48   |
| 2024-08-01 00:00:00 UTC | 76.32        | 50.1         | 106.04      | 77.89        | 34.92      | 34.64      | 35.44      | 34.43      | 34.87   |

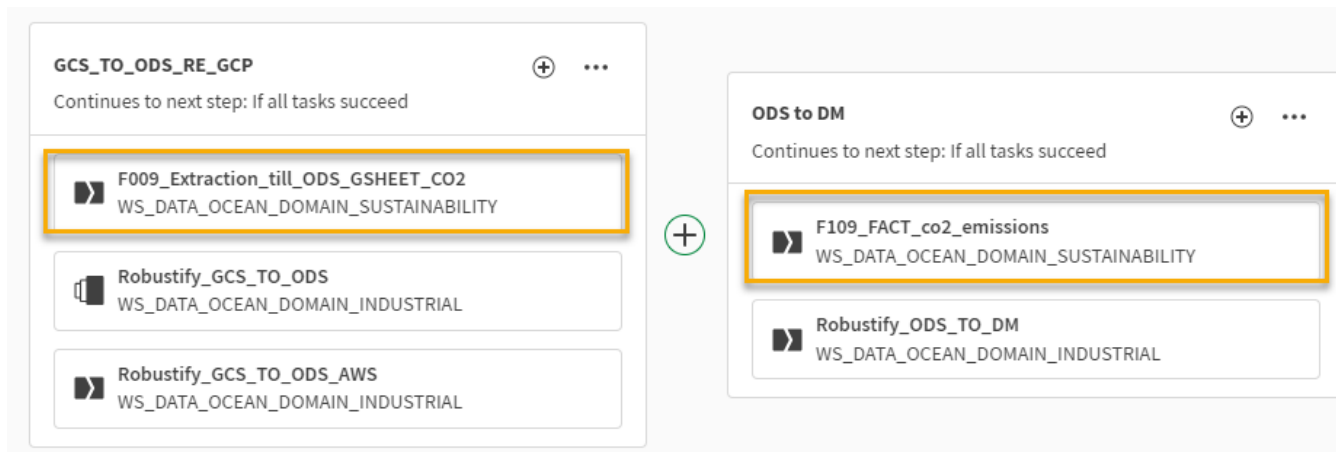
3. Apply the business rule and save into the output file by tJavaFlex

4. Transform to the Dataiku format

| date                    | elec_DE_EPEX | elec_FR_EPEX | elec_IT_PUN | elec_SP_OMEI | gas_DE_THE | gas_FR_PEO | gas_IT_PSV | gas_SP_PVB | gas_TTF | elec_IT_MGSP | inserted_date           |
|-------------------------|--------------|--------------|-------------|--------------|------------|------------|------------|------------|---------|--------------|-------------------------|
| 2024-06-21 21:00:00 UTC | 105.26       | 105.14       | 135.3151    | 114.86       | 33.963     | 34.028     | 36.875     | 34.49      | 33.867  | null         | 2024-06-21 10:09:55 UTC |
| 2024-06-21 22:00:00 UTC | 95.99        | 96.49        | 115.0116    | 115.0        | 33.963     | 34.028     | 36.875     | 34.49      | 33.867  | null         | 2024-06-21 10:09:55 UTC |
| 2024-06-21 23:00:00 UTC | 82.14        | 83.26        | 112.0       | 105.74       | 33.963     | 34.028     | 36.875     | 34.49      | 33.867  | null         | 2024-06-21 10:09:55 UTC |
| 2024-06-22 00:00:00 UTC | 73.15        | 50.94        | 106.9       | 74.56        | 34.5       | 34.31      | 36.15      | 34.45      | 34.48   | null         | 2024-06-21 10:09:55 UTC |
| 2024-06-22 01:00:00 UTC | 73.15        | 50.94        | 106.9       | 74.56        | 34.5       | 34.31      | 36.15      | 34.45      | 34.48   | null         | 2024-06-21 10:09:55 UTC |
| 2024-06-22 02:00:00 UTC | 73.15        | 50.94        | 106.9       | 74.56        | 34.5       | 34.31      | 36.15      | 34.45      | 34.48   | null         | 2024-06-21 10:09:55 UTC |
| 2024-06-22 03:00:00 UTC | 73.15        | 50.94        | 106.9       | 74.56        | 34.5       | 34.31      | 36.15      | 34.45      | 34.48   | null         | 2024-06-21 10:09:55 UTC |
| 2024-06-22 04:00:00 UTC | 73.15        | 50.94        | 106.9       | 74.56        | 34.5       | 34.31      | 36.15      | 34.45      | 34.48   | null         | 2024-06-21 10:09:55 UTC |
| 2024-06-22 05:00:00 UTC | 73.15        | 50.94        | 106.9       | 74.56        | 34.5       | 34.31      | 36.15      | 34.45      | 34.48   | null         | 2024-06-21 10:09:55 UTC |

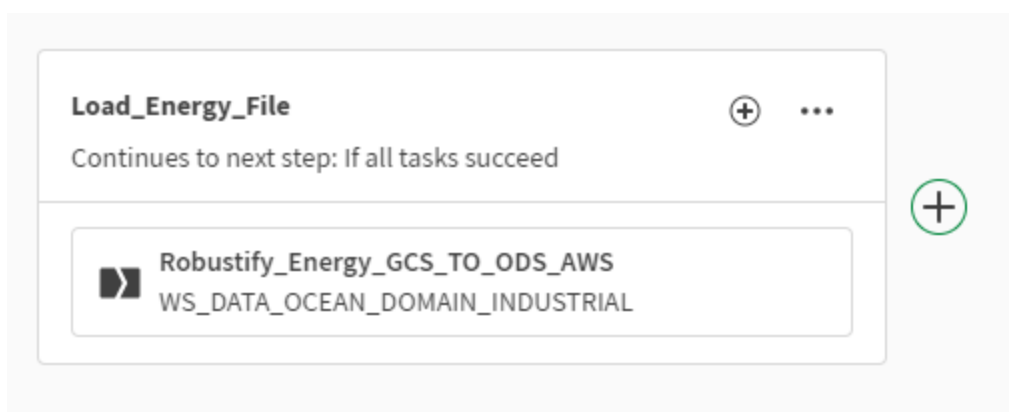
5. Delete temporary file

Although there are 3 projects involve but main plan will be in TMC on WS\_DATA\_OCEAN\_DOMAIN INDUSTRIAL with plan PL\_INDUS\_ROBUSTIFY\_LOAD load - Daily at 12:15 PM.



\*\*Yellow box is the job in Sustainability domain.

PL\_INDUS\_ROBUSTIFY\_ENERGY load hourly at xx:30



Monitor the loading in the log tables on prj-data-dm-industrial-[env] by

```
select job.job_name , job.meta_start_date , 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 1000 )
and (
job_name like '%METEOLOGICA%' or
upper(job_name) like '%ENERGY%' or
job_name like '%WAP%' or
job_name like '%IRM%' )
and meta_start_date > DATE_SUB ( CURRENT_TIMESTAMP () , INTERVAL 1 DAY )
order by job.meta_start_date desc
```

## 4.6 - Remark

1. Most of Talend jobs will be on project DATA\_OCEAN\_DOMAIN\_INDUSTRIAL except:

- Gsheet Co2 (J009\_ODS\_TO\_DM\_FACT\_ENERGY\_DEALS) will be in DATA\_OCEAN\_DOMAIN\_SUSTAINABILITY prj-data-dm-sust-[env].STG.  
STG\_FIL\_0000\_0000\_F001\_F\_D\_co2\_emissions

- DM to DM (J001\_DM\_TO\_DM\_FACT\_ENERGY\_PRICE\_HOURLY) will be in ROBUSTIFY project prj-data-robustify-[env].DM.  
FACT\_energy\_price\_hourly

This is following [data architecture](#)

2. Most of Talend jobs are required to use remote engine on Cloud except J014\_Extraction\_till\_ODS\_ORACLE\_IRM is required AWS. This is because of security reason