

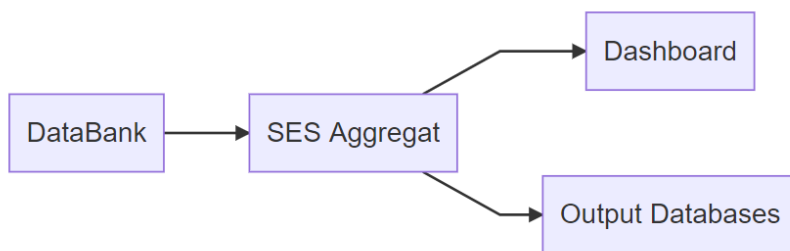
Aggregat - 4 - Technical - Data Application

- Architecture Data Flow
 - DataApp Flow
- Steps descriptions
 - SES Aggregat
 - Description
 - Tools
 - Access rights
 - Validation process
 - Configuration
 - Source locations
 - Source formats
 - Destination locations
 - Destination formats
 - Sizing
 - Scheduling
 - Timing
 - Criticality
 - Logging

Architecture Data Flow

DataApp Flow

Schema showing the different STEPS of the application flow - with the data involved at each step



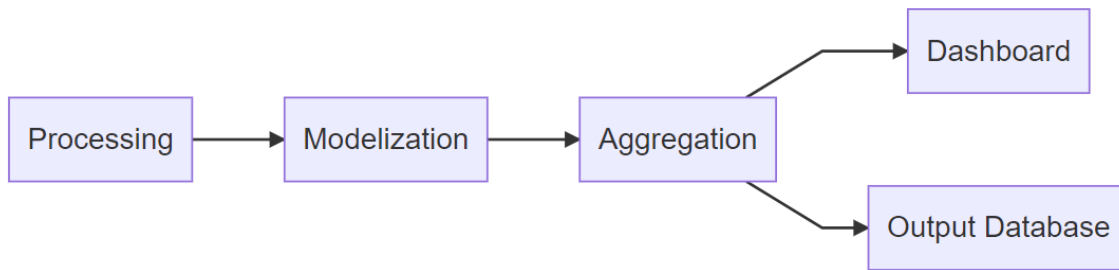
Steps descriptions

Describe the data and process involved at each step

SES Aggregat

Description

What does it do ?



Forecast energy production for solar and wind park.

1. Processing step : All datas sources are imported from the Data Bank into the project :

- Météo France / Arpege
- History Production
- Energy DB Mapping
- RTE
- EpexPrice Then these datas are combine with specifications and maintenance files from business to filter on required park only. There are also processed to get the current date, or windows of 14 days up to the current date.

2. Modelization step :

- a. Solar : Datas are prepared to make a train dataset for solar energy production forecast. All production history is used to train a model. Then the model is used to predict on current day datas to predict the next 48 hours of energy production.
- b. Wind : Same principle with wind energy production. Each model solar and wind is executed in parallel.

3. Aggregation step : A sequence of aggregation an stacking is performed by park.

4. Dashboard step : Datas are prepared to be send to a Dashboard for visualization.

5. Output step : Final Datas are send to the output database.

Tools

The tool used

Dataiku DSS, Python, GitLab

Access rights

Is there any credentials used ? Where are they stored ?

No credentials except for Data Bank access. Database connection settings are already set up in Dataiku DSS

Validation process

How to validate that this output match business expectations

Multiple check are done on output dataset :

- consistency : check that schema are matching expectation
- Forecast range : check that forecast are in a acceptable range of values
- Date : check that the right date is forecasted
- minimum count : check that forecasted rows are not under a lower bound. A check is performed on input/ouput tables through a scenario.

Configuration

which kind of information or configuration this step uses ?

Process steps is fully automated, a date argument is automatically fetched day by day. No configuration is required.

Source locations

Where is the data coming from ?

Data is coming from the DataBank on GCP. The input tables are pre-computed from the DataPrep project automatically. There is also 2 configuration files coming from Google sheet where business rules are defined

Source formats

The format of the source datasets

SQL Database from GBQ in connection CX_GBQ_A_PRJ-BDA-DATABANK-DEV_ALL :

- History_Production_DATABANK
- EpecPrice_DATABANK
- Energy_DB_Mapping_DATABANK
- Hours_Limit_Mapping_DATABANK
- RTE_Imbalance_UnitCost_DATABANK
- RTE_ImbalanceData_DATABANK
- Arpege_History
- Arpege

File csv for google sheet sources :

- Park_Specifications
- Park_Maintenance

Destination locations

Where is the data stored ?

Output data is stored in GBQ

Destination formats

The format of the data output

SQL Database from GBQ in connection CX_GBQ_A_SOLVAY-ENERGY-AGGREGATION-DEV_DATAIKU :

- insertPrevisionElec_GBQ
- AGR_PARK_DATAOPS_GBQ
- Park_Forecast_For_Dashboards_historical_GBQ

Sizing

Expected data volume

Data Volume :

- table insertPrevisionElec_GBQ : ~130 rows per day
- table AGR_PARK_DATAOPS_GBQ : ~250 rows per day

Scheduling

Is there an automatic schedule ? At what frequency ? What is the trigger ?

2 automatic schedule :

- A train scenario executed once per week
- A predict scenario executed once per day

Timing

The average time expected for :

- *full process*
- *incremental process*

There are only incremental scenario:

- Train scenario : 1h10
- Predict scenario : 40min

Criticality

High / Medium / Low

Medium

Logging

Logging location

- Train : <https://dss-test.solvay.com/projects/SESAGGREGAT/scenarios/TRAIN/runs/list/>
- Predict : <https://dss-test.solvay.com/projects/SESAGGREGAT/scenarios/PREDICT/runs/list/>