DFS TD - Synthesis Reaction Raw Data

Technical documentation for the data coming from reaction tests

Summary

Sum-up

Equipment / Scale	Reaction 25L (A and B)	Reaction 80L	Reaction 170L (France and Korea)	Reaction 2500L
Data Sources	ELN, Raw Data on file share	ELN, Raw Data on Google Drive	ELN, Raw Data on file share	ELN, Raw Data on Google Drive
Raw Data File type	xlsx	xlsx	xlsx	xlsx
Scale Name on ELN	FR-25L-A FR-25L-B	FR-80L	FR-170L KR-170L	FR-2500L
Data Collection	Talend: • J010_Download_Synthesis_LabServers Python : • download_synthesis_25L.py • download_synthesis_25L_B.py	Talend: • R013_Download_Synthesis_gDrive_Reaction	Talend: • J010_Download_Synthesis_LabServers Python : • download_synthesis_170L.py • download_synthesis_170L_KR.py	Talend: • R013_Download_Synthesis_gDrive_Reaction
Parse	Python: parse_synthesis_25L.py	Python: parse_synthesis_2500L.py	Python: • parse_synthesis_170L.py • parse_synthesis_170L_KR.py	Python: parse_synthesis_2500L.py
Compute	Python: compute_synthesis_25L.py	Python: compute_synthesis_80L.py	Python: compute_synthesis_170L.py	Python: compute_synthesis_2500L.py
BigQuery	*Target tables:* • *raw_data_synthesis.ReactionDetails* • *raw_data_synthesis.ReactionSummary* • *raw_data_synthesis.ReactionMaterialBalance (table to be documented)*			
Mapping spreadsheet	Mapping Spreadsheet			

Data Sources

- ELN
- Raw Data

Data Collection

The talend jobs **J010_Download_Synthesis_LabServers** and **J011_Download_Synthesis_gDrive_Reaction** extract the raw data files listed on the ELN table **synthesis_raw_data_link** for which the field “synthesis_equipment_name” is the scale name, i.e. “FR-80L”. For information of how these job works, check the following page :

Talend - Jobs - Synthesis - Download:

Exemples

Lab servers source

- \\FRPH2-labpc-backup\labo\W-522649\DATAS DATALAKE
- \\FRPH2-LABPC-BACKUP\LABO\W-509931

Python files

- download_filtration170L.py
- download_synthesis25L.py

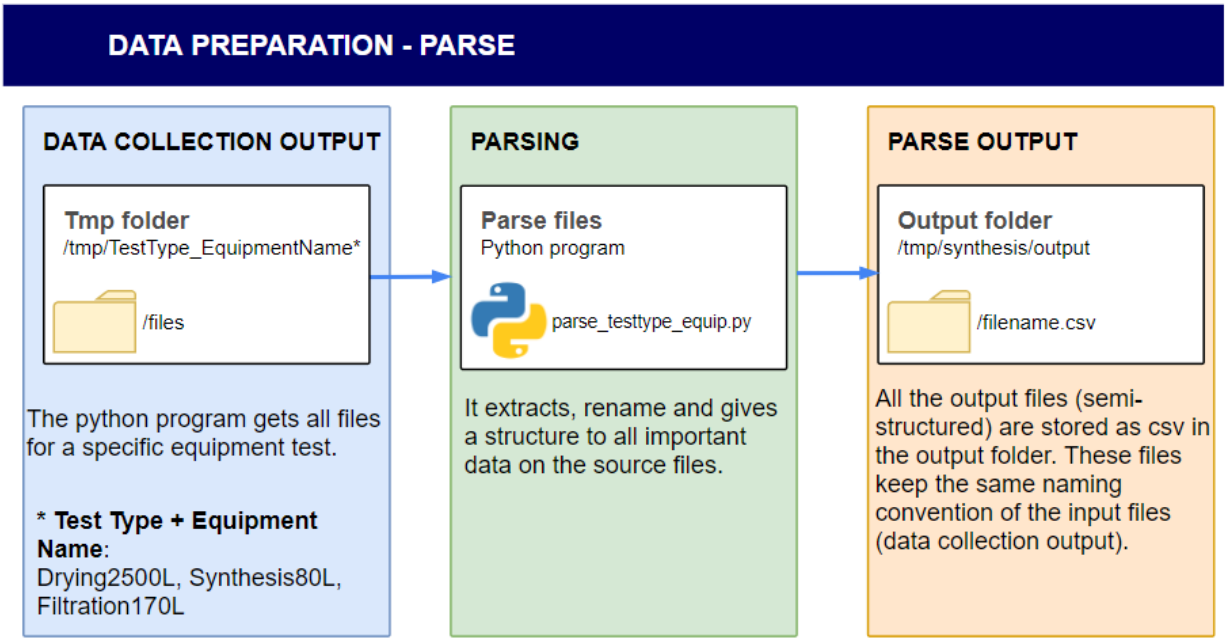
Output folders

- D:\DATA\[ENV]\RnI\Silica\tmp\Synthesis25L
- D:\DATA\[ENV]\RnI\Silica\tmp\Synthesis170L

Data Preparation

Parse

The parsing python scripts extracts from the raw data files the needed columns.



Columns List

For each sample, the script extracts the many fields from the raw data files and outputs a .csv file. For the mapping details, please refers to the sheet "Pars e Mapping " on the Reaction Mapping spreadsheet (link to the spreadsheet on the Sum-up section).

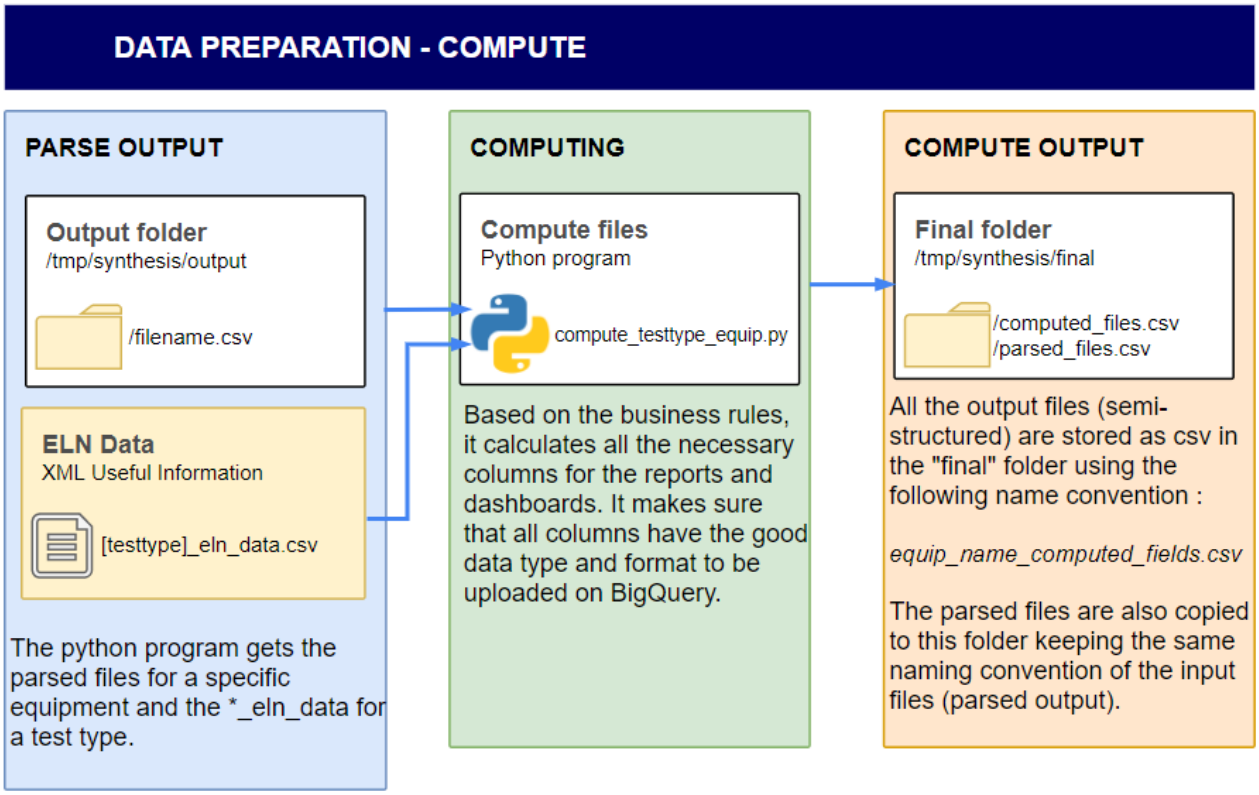
T he following columns extracted from the raw data file:

- unique_id
- study_id
- sample_id
- operator
- reactor
- date
- time (in minutes)
- ph
- temperature
- acid_mass_one
- silicate_mass
- additive_mass
- acid_mass_two
- variable_product_mass (empty for 170L scale)
- percent_acid_one
- percent_silicate
- percent_additive
- percent_acid_two
- percent_pump_pH_control
- percent_variable_product (empty for 170L scale)
- turbidity

Compute

The compute python script uses as input the parsed .csv files previously created and the tables **synthesis_eln_data** and **operating_procedure** . It computes the new columns and values from raw data and regenerates new files.

If the output files already exist the script will **NOT** replace them.



In the beginning of the script, for each product, we extract the following values for each product listed in the table ***synthesis_eln_data*** . Each of the following values will be used in later computations as constants:

Product	**Variables**

Silicate	• density_silicate_eln = density_silicate (from ENL)
	• density_silicate
	  ◦ density_silicate extracted from ELN is replaced by the following computation:
	  ◦ 144 * density_silicate (from ELN) / (144 + 0.035 * density_silicate (from ELN) * temperature_max)
	  ◦ In the previous formula, temperature_max = maximum of the column [Temperature] from the raw data file
	  ◦ This correction is necessary for the computation of the total volume
	• rp_silicate
	• silicate_qty
	• concentration_sio2
	• concentration_na2o

Water	• water_qty • density_water • concentration_water
Aluminate	• add_qty • density_add • concentration_add_al • concentration_add_na2o
Other	• other_qty • density_other • concentration_add_oo (product name = Other and compound name = Other) • concentration_add_ou (product name = Other and compound name = Unknown) • concentration_hplus_o (product name = Other and compound name = H+) • concentration_na2o_o (product_name = Other and compound name = Na20) • nb_hplus_hplus_o

R66	• r66_qty • density_r66 • concentration_add_rma (product name = R66 and compound name = 2-methylglutaric acid) • concentration_hplus_r (product name = R66 and compound name = H+) • nb_hplus_hplus_r
Sodium Sulfate	• sodium_sulfate_qty • concentration_sodium_sulfate
Sodium Hydroxide	• sodium_hydroxide_qty • concentration_sodium_hydroxide • density_sodium_hydroxide

Sulfuric Acid Concentrate	• h2so4_c_qty
	• concentration_h2so4_c
	• density_h2so4_c_eln = density_h2so4 (from ELN)
	• density_h2so4_c
	  ◦ density_h2so4_c extracted from ELN is replaced by the following computation:
	  ◦ ((-0.3119 * (concentration_h2so4_c * 100) ** 2 + 61.569 * (concentration_h2so4_c * 100) - 1200.4) - (0.5133 * temperature_max)) / 1000
	  ◦ In the previous formula, temperature_max = maximum of the column [Temperature] from the raw data file
	  ◦ This correction is necessary for the computation of the total volume
	• nb_hplus_h2so4_c

Sulfuric Acid	• h2so4_d_qty
	• density_h2so4_d_eln = density_h2so4_d_eln (from ELN)
	• density_h2so4_d
	  ◦ density_h2so4_d extracted from ELN is replaced by the following computation:
	  ◦ (density_h2so4_d (from ELN) * 1000 - 0.5133 * temperature_max) / 1000
	  ◦ In the previous formula, temperature_max = maximum of the column [Temperature] from the raw data file
	  ◦ This correction is necessary for the computation of the total volume
	• concentration_h2so4_d
	• nb_hplus_h2so4_d

Nitric Acid Concentrate	• hno3_c_qty • density_hno3_c • concentration_hno3_c • nb_hplus_hno3_c
Nitric Acid	• hno3_d_qty • density_hno3_d • concentration_hno3_d • nb_hplus_hno3_d
Chlorhydric Acid Concentrate	• hcl_c_qty • density_hcl_c • concentration_hcl_c • nb_hplus_hcl_c
Chlorhydric Acid	• hcl_d_qty • density_hcl_d • concentration_hcl_d • nb_hplus_hcl_d

Molar masses for the following elements are also defined and used in later computations (mm_ stands for molar mass):

- mm_na2o = 61.98
- mm_h2so4 = 98.079
- mm_sio2 = 60.084
- mm_na2so4 = 142

- mm_hcl = 36.46
- mm_hno3 = 63.02
- mm_hplus = 1.01

Next, we define the **activity on each pump** as follows:

- We first define “by default” activity on each pump:
  - acid_one pump (concentrated acid) Sulfuric Acid Concentrate
  - silicate pump Silicate
  - additive pump Aluminate
  - acid_two pump (diluted acid) Sulfuric Acid
- Next, from the table ***operating_procedure*** , we extract the changes in the activity for each pump. The next table lists the products that can be present on each pump:

Any other element (other than those listed in the table for each pump) will not be considered for later computations

raw_mass_acid_one (conc)	raw_mass_silicate	raw_mass_additive	raw_mass_acid_two (dil)
WIR811_Masse	**WIR611_Masse**	**WIR711_Masse**	**WIR511_Masse**
Sulfuric Acid Concentrate (default)	Silicate (default)	Aluminate (default)	Sulfuric Acid (default)
Nitric Acid Concentrate		Other	Nitric Acid
Chlorhydric Acid Concentrate		R66	Chlorhydric Acid
Other			Other
R66			R66
Water			Water

For each sample, the compute scripts create **three different tables:**

ReactionDetails

The **first table** is composed of the columns previously extracted from the raw data files and the new columns calculated during the execution.

Dataset : raw_data_synthesis_mig

For the columns details, please refers to the sheets " **Details Mappings** " on the **Reaction Mapping** spreadsheet (link to the spreadsheet on the Sum-up section).

ReactionSummary

The **second table** is composed of the new values computed from raw data. This is a atomic table and it aggregates the values by **unique_id, study_id and sample_id** which represents one line per data raw file.

Dataset : raw_data_synthesis_mig

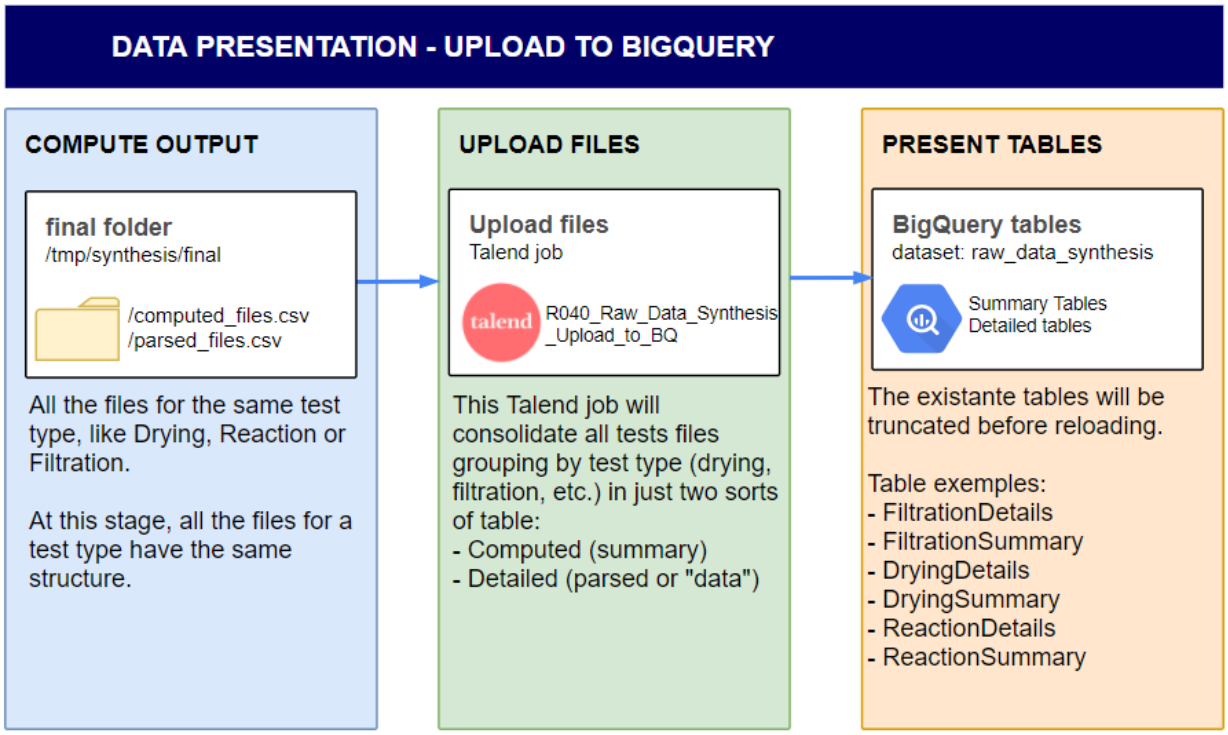
For the columns details, please refers to the sheets " **Summary Mapping** " on the **Reaction Mapping** spreadsheet (link to the spreadsheet on the Sum-up section).

ReactionMaterialBalance

To be documented

Presentation

The raw data (already parsed) and the computed columns are created as tables on **BigQuery**. A Talend job is responsable to push all this data to a dataset called **raw_data_synthesis_mig.**


DATA PRESENTATION - UPLOAD TO BIGQUERY
COMPUTE OUTPUT
final folder
/tmp/synthesis/final
/computed_files.csv
/parsed_files.csv
All the files for the same test type, like Drying, Reaction or Filtration.
At this stage, all the files for a test type have the same structure.
UPLOAD FILES
Upload files
Talend job
talend
R040_Raw_Data_Synthesis_Upload_to_BQ
This Talend job will consolidate all tests files grouping by test type (drying, filtration, etc.) in just two sorts of table:
- Computed (summary)
- Detailed (parsed or "data")
PRESENT TABLES
BigQuery tables
dataset: raw_data_synthesis
Summary Tables
Detailed tables
The existante tables will be truncated before reloading.
Table exemples:
- FiltrationDetails
- FiltrationSummary
- DryingDetails
- DryingSummary
- ReactionDetails
- ReactionSummary