Aggregat - 5 - Operational

- Procedures
  - DataPrep Flow
    - Start
      - Full process
    - Termination
    - Restart
    - Pause
      - Procedure
    - Stop
      - Procedure
      - Alert contacts
    - Reset
  - DataApp Flow
    - Start
      - Full process
- Scheduling
  - Trigger
  - Expected results
  - Intervention
- Monitoring
  - Runtime
  - Run history
  - Resources
  - Additional metrics
  - Logging
- Error handling
  - Alerts
  - Specificity

Procedures

> *Procedure guide on how to operate the application*

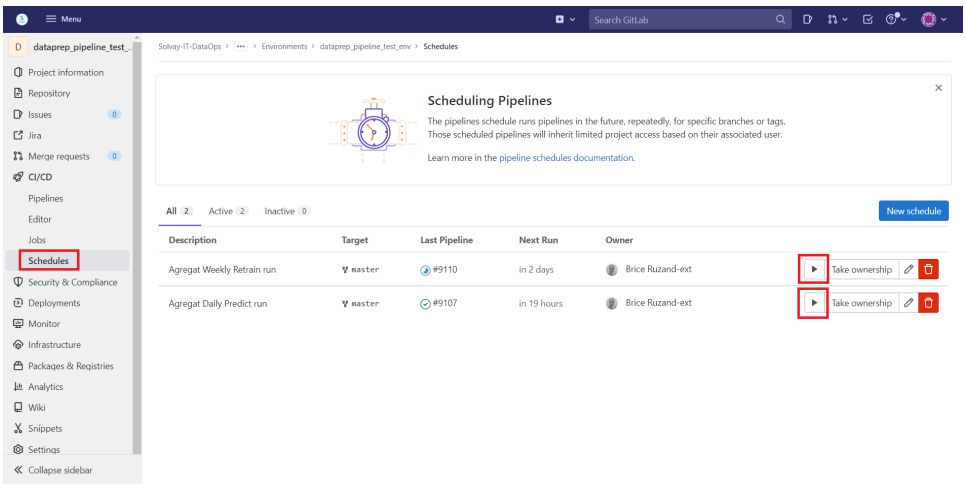
DataPrep Flow

Start

Full process

> *How to start from scratch ?*

- Go to the Gitlab Dataprep project in the section Scheduler : https://gitlab.solvay.com/solvay-it-dataops/data-ingestion/ses-agregat-dataprep/environments/dataprep_pipeline_test_env/-/pipeline_schedules
- Click on play for the line `Agregat Daily Predict run` and `Agregat Weekly Retrain run`.

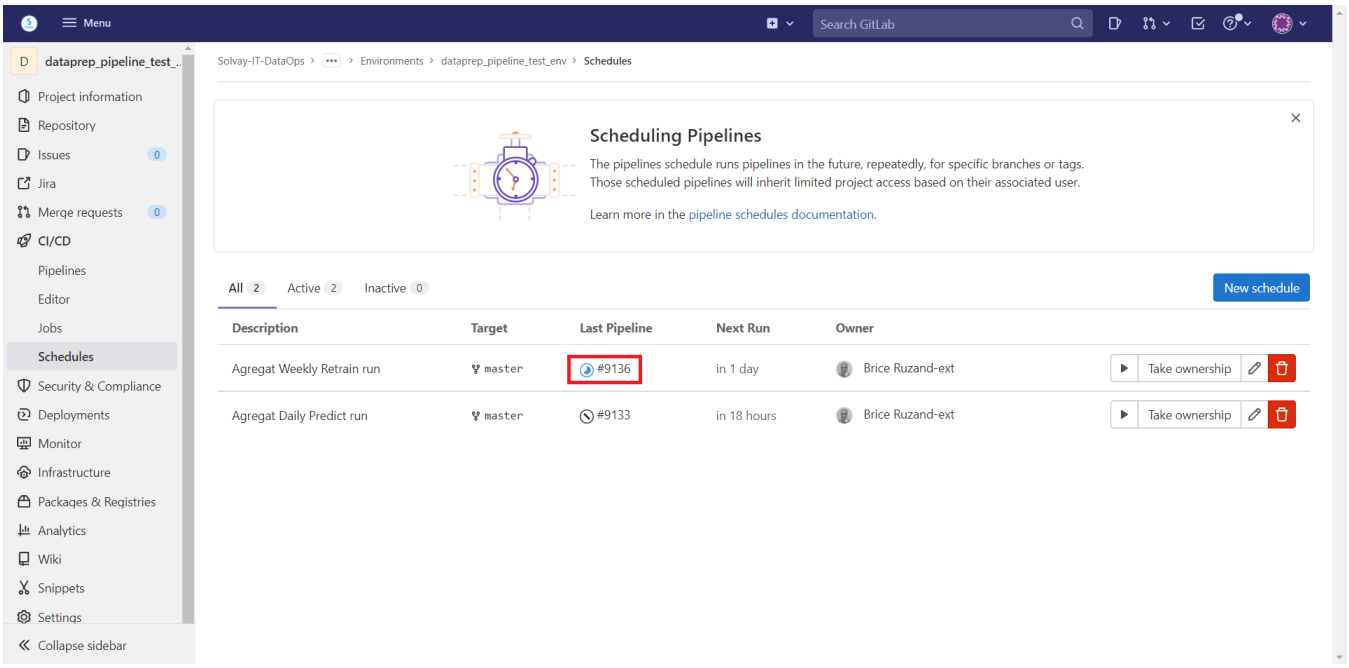


This will start the scheduling of the Dataprep each day at 7:00 am. The DataApp is triggered once the dataPrep is done so this is also the way to start the full process of DataApp as well.

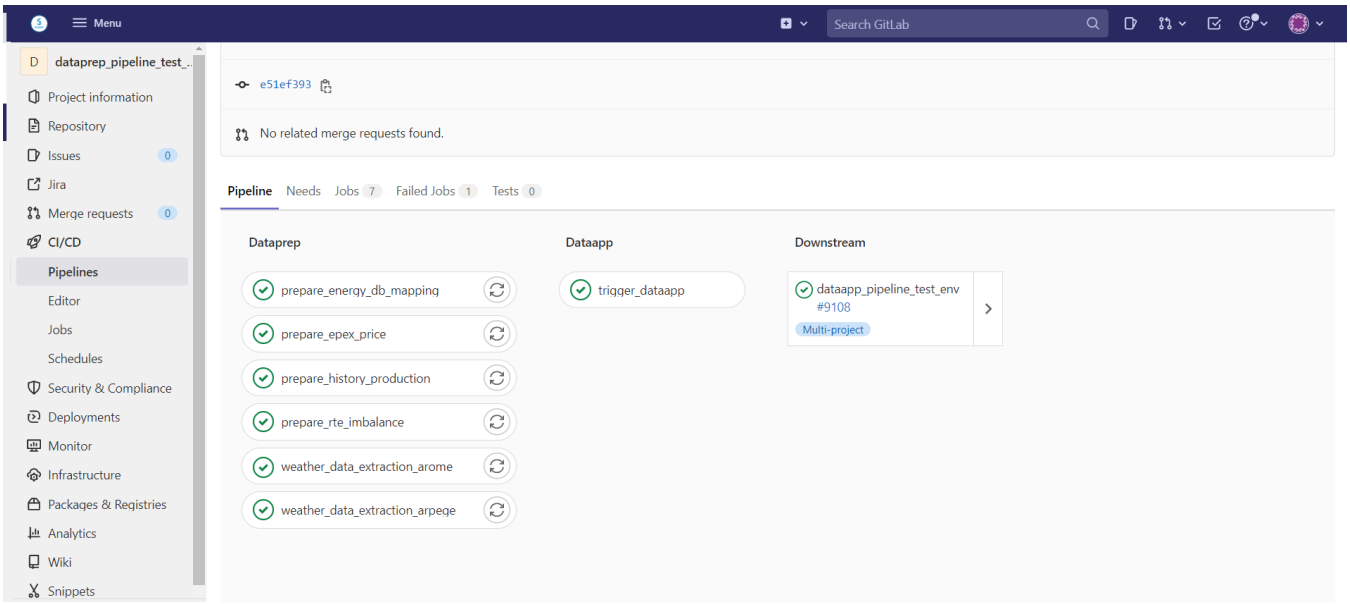
Termination

> *How to assess the application's process has terminated*

Once the process is launched you can follow the status by clicking on the pipeline number



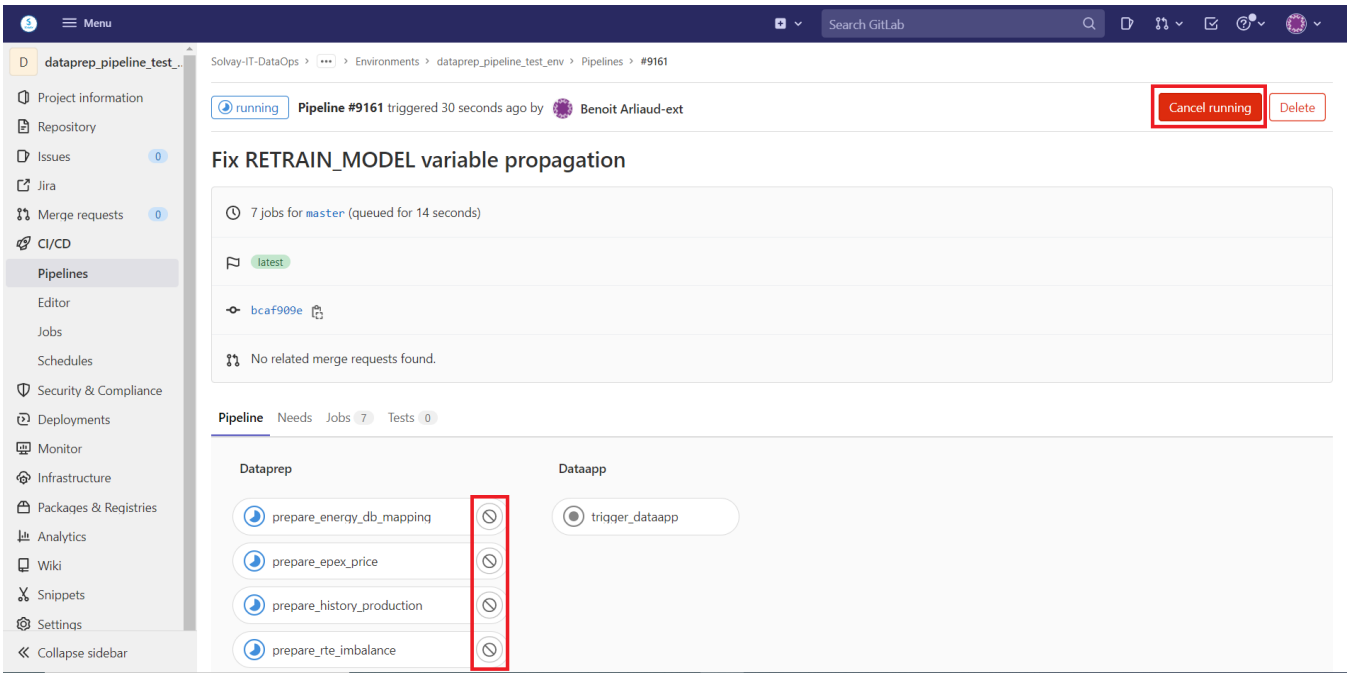
Then you can see a list of all pipelines running with there current status. This is the expected view of a pipeline that has terminated without errors. On the left are all the pipeline from the DataPrep, on the left this is the trigger toward the DataApp Pipelines.



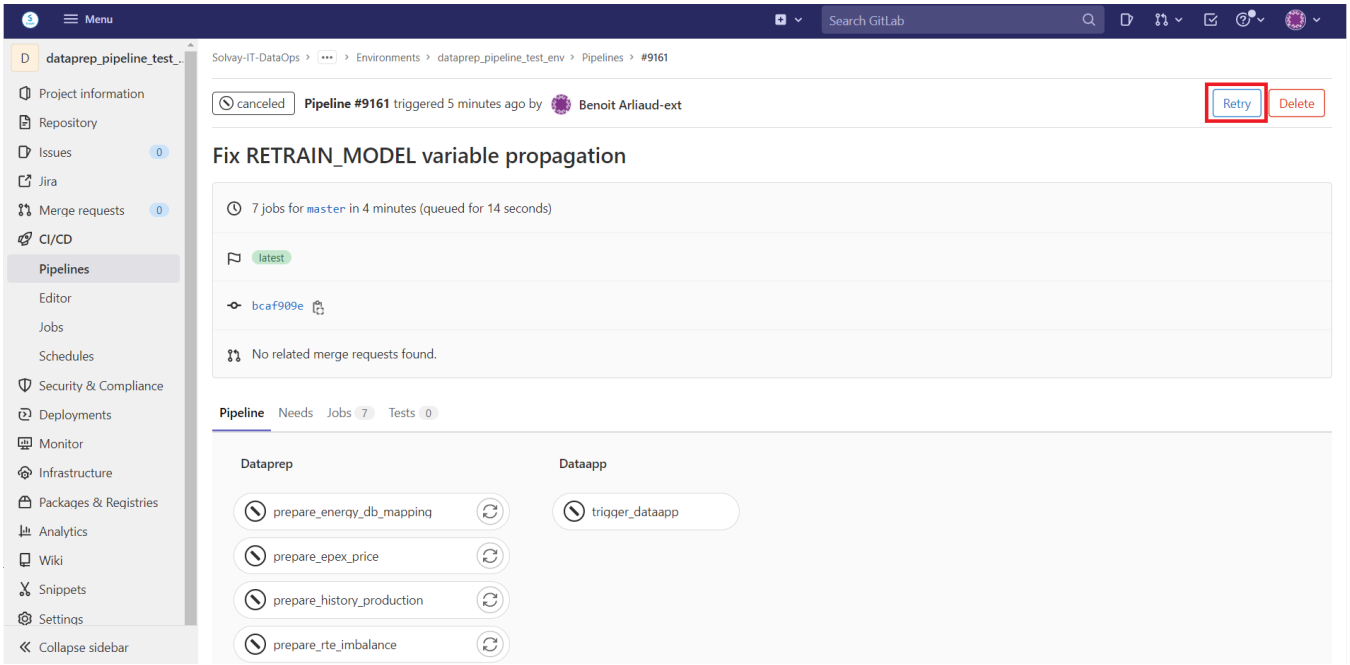
Restart

> *Is the process the same ?*
>
> *Also for restarting after an error ?*

To restart the DataPrep pipeline one must stop the running pipeline by either clicking on `Cancel running` or cancelling each pipeline one by one.



Then when once all the pipelines are stopped you can select the `Retry` option. This option also appear if everything succeeded.



Pause

Procedure

> *How ? (Resume)*

Procedure can't be paused but it can be restarted.

Stop

Procedure

> *How ?*

See section Restart

Alert contacts

> *Who should be alerted ?*

Reset

> *How ?*

DataApp Flow

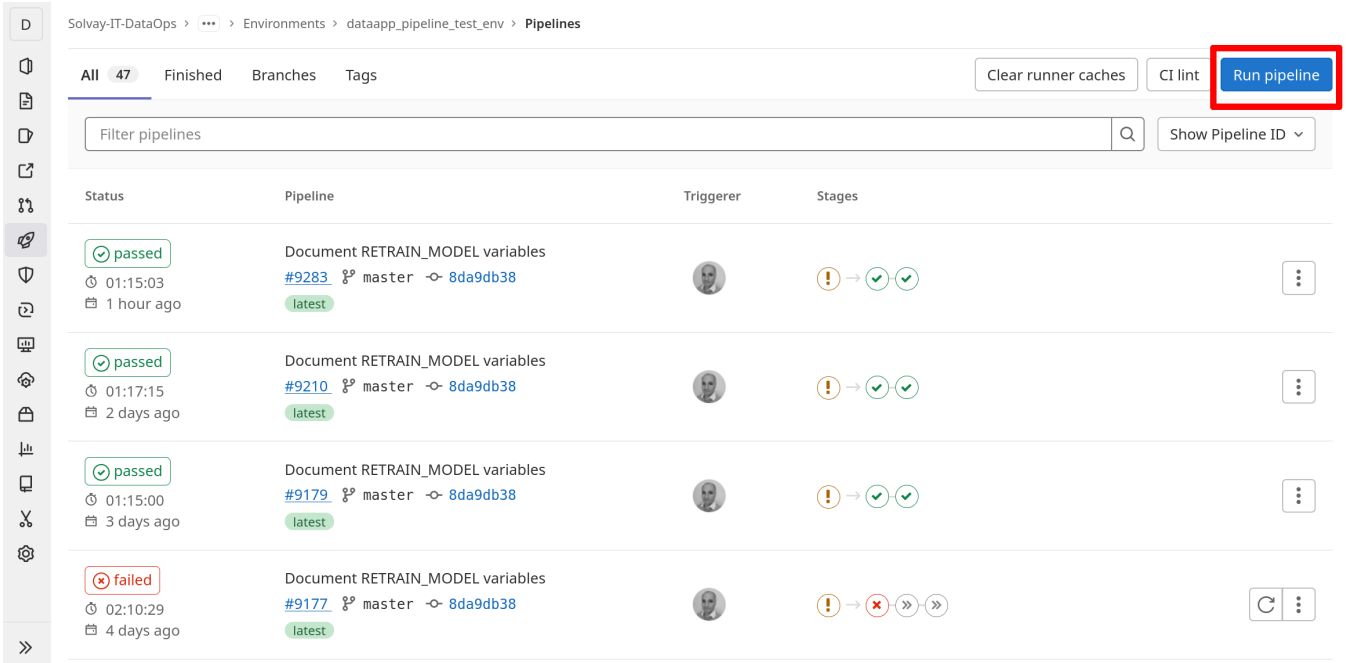
Start

Full process

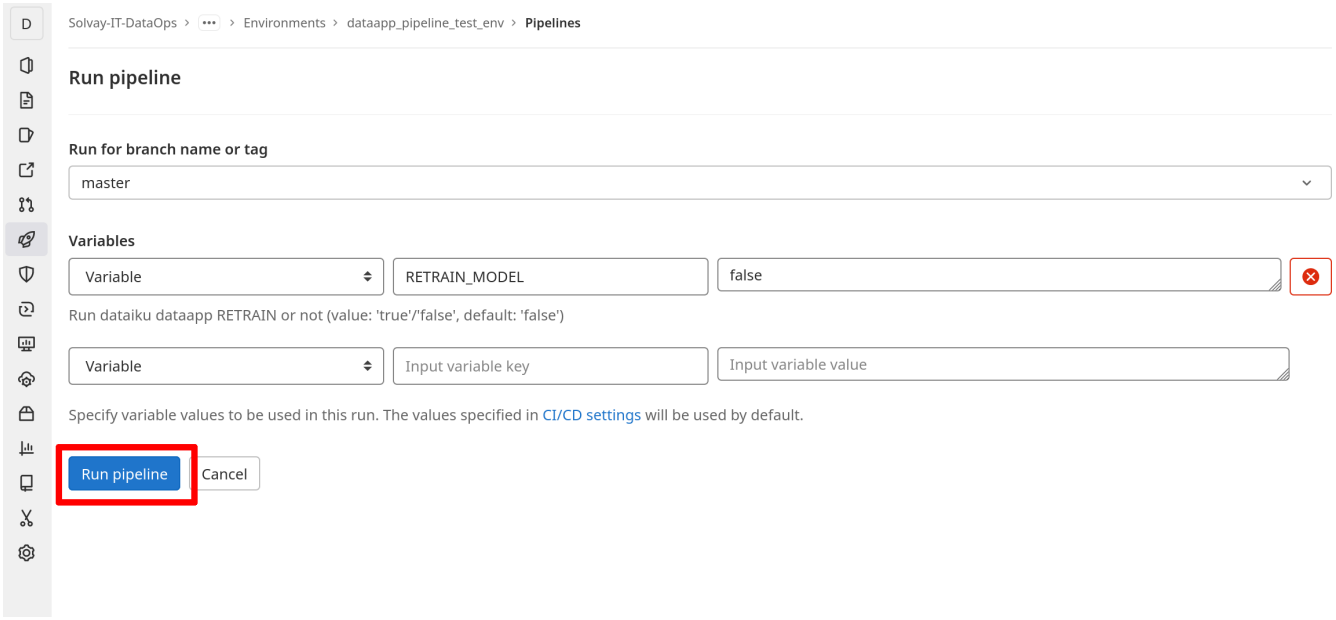
> *How to start from scratch ?*

DataApp is automatically launched after DataPrep has finished thank to a trigger, so the whole process DataPrep and DataApp is launched from DataPrep Pipeline.

However the DataApp Pipeline can be launched manually in DataApp pipelines



Just click on on blue button "Run Pipeline".



Then choose to set `RETRAIN_MODEL` to `true` if you want to retrain model.

If a pipeline fail you can select the pipeline then the job to see the logs like for the DataPrep.

Scheduling

Trigger

> *What is the start trigger ? Event based ? Time based ?*
>
> *Are there differences between DataPrep and DataApp ?*

The DataPrep is the first Trigger to set up. It launched the DataApp automatically based on a scheduler.

DataApp doesn't have a scheduler, all scheduler must be set up from DataPrep side.

Expected results

> *For each brick, what is the expected output ?*

All Pipeline must be succeeded as a resulting status which means that in dataiku the scenario has succeeded in every step.

It is recommended to click on the link in the logs window to check in the Dataiku instance that the jobs has well succeeded.

Intervention

> *When is the time frame to intervene ? (when downtime is acceptable or scheduled)*

Monitoring

Runtime

> *Where and how can we see the application status (Stopped, waiting, running, etc) ?*

The status can be seen in the logs window when selecting a specific pipeline.

Run history

> *Where are the run actions historic ?*

> *What form does it take ? Logs ?*

The history of pipeline execution can be seen here :

- DataPrep pipelines
- DataApp pipelines

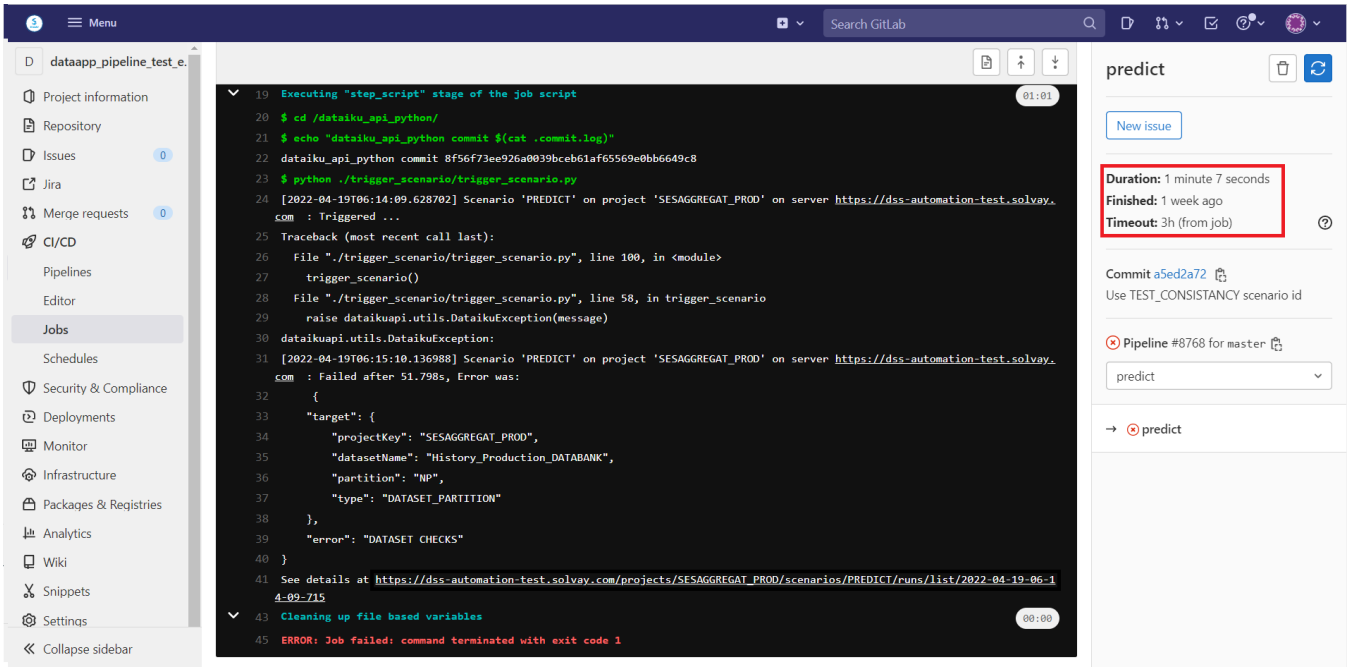
Resources

> *Memory / Disk / CPU used by application*

Additional metrics

> *According to operational requirements, detail application metrics (Processed Volume, Process duration, ...)*

Process duration can be seen from the logs window on the right panel



Logging

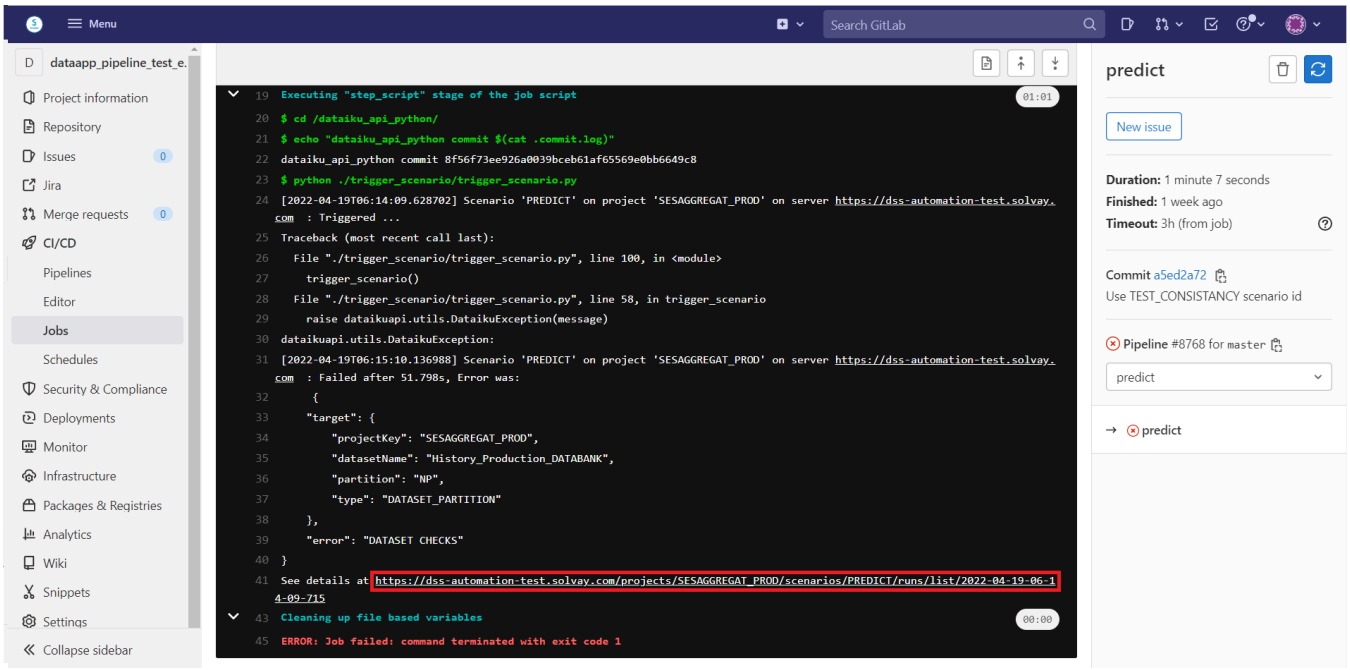
> *Where to find each step logs ?*

On the logs window.

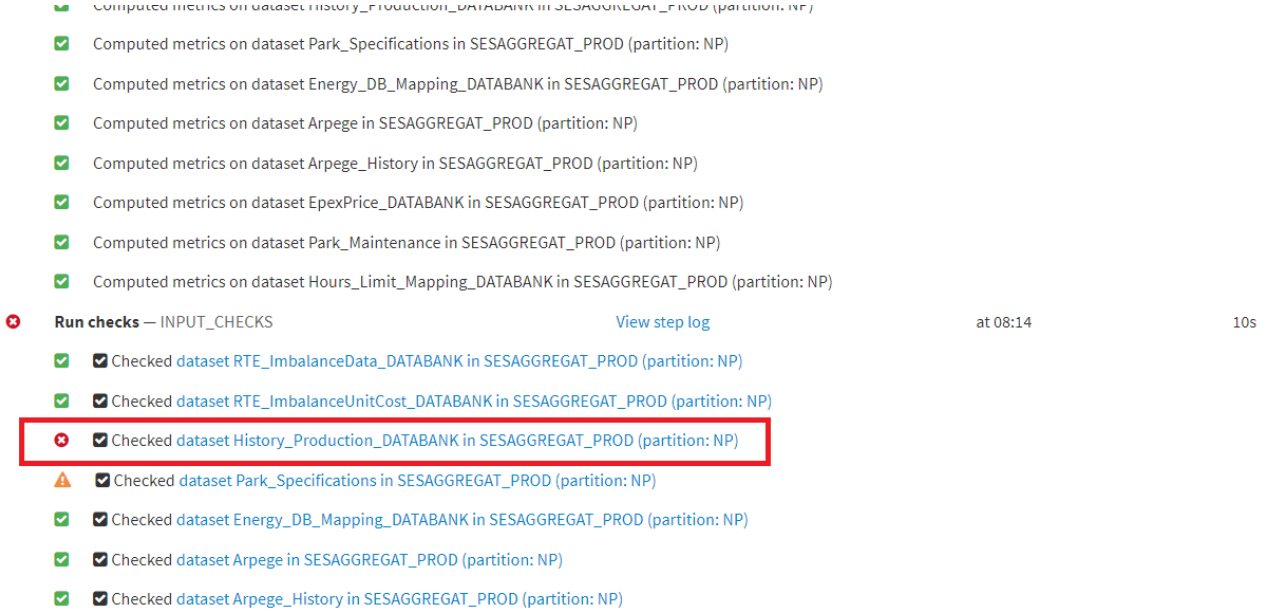
Error handling

> *As a general guideline, application should stop as soon as possible.*

In case of an error you must select the erroneous pipeline to see the logs. At the end of the logs you will have a link to the dataiku's job error.



By following link to dataiku, you will be able to se error details.



Alerts

> - *Contacts*
> - *Meaningful message (timestamps, description, criticality)*

Meaningful message is displayed on the right panel.

Specificity

> *Detail procedure for specific error cases*

Each error has a link to the dataiku job related. If the error message isn't clear, it is recommended to go access dataiku job and have more details.