

INFO-H420 Management of Data Science and Business Workflows

Assignment 3: Workflows with Apache Airflow (10 points)

The goal of this assignment is to write a simple workflow that analyzes a web server log file, extracts the required lines and fields, performs some transformations, and writes to another file. The log file is available on UV.

Exercise 1 (7 points)

- Define the DAG. Create a DAG named “process_web_log” that runs daily.
- Create a task to scan for a log. Create a task named “scan_for_log” that scans a folder “the_logs” for a “log.txt” file and triggers the rest of the workflow.
- Create a task to extract data. Create a task named “extract_data”. This task should extract the ipaddress field from the web server log file and save it into a file named “extracted_data.txt”.
- Create a task to transform data. Create a task named “transform_data”. This task should filter out all the occurrences of ipaddress 198.46.149.143 from “extracted_data.txt” and save the output to a file named “transformed_data.txt”.
- Create a task to load the data. Create a task named “load_data”. This task should archive the file “transformed_data.txt” into a tar file named “weblog.tar”.
- Define the workflow that execute the aforementioned tasks in sequence.

Save the DAG you defined into a file named “process_web_log.py”. In the report, include code snippets describing how you achieved this.

Exercise 2 (1 points)

Do a test run for each of the tasks you defined. Once all works as expected, do a test run for the workflow. Finally, trigger/run the workflow and monitor a few runs. In the report, document the test runs, and include any findings or observations that you may have from the runs.

Exercise 3 (2 points)

Add to the workflow an additional task after the last task. That task should send a message that the workflow was executed. You are free to define how and where this message is sent. For example, you can post to a Discord or Slack channel, write the message to a text file uploaded to Google Drive or Dropbox. In the report, document how you implemented the message task, and provide evidence that it executes.

Instructions

The assignment contributes 10% to the overall grade (i.e., 2/20).

This assignment is to be made in **groups of two** persons. You are asked to form the groups via “Groups for Assignment 3” on the Université Virtuelle (UV). If you cannot find a partner, please post a request in the “Discussion Forum” on UV. You should not form the same group as in the previous assignments.

You are asked to submit a short **report** presenting your solution to the exercises, including justifications for the choices and assumptions made.

The report and any supporting files have to be uploaded to “Assignment 3” on UV by **November 17, 2025**.