

Milestone 12 due Friday, 05/03.

Create an Airflow DAG that implements the data transformation workflow which you specified in Milestone 11.

Required functionality:

- DAG reads the data from the raw tables and performs the necessary transformations starting with SQL and ending with Beam.
- DAG executes dependent tasks in sequence.
- DAG executes independent tasks in parallel.
- results from the various transformations are stored in separate BQ dataset

What's **not** required:

- loading the CSV files into BQ
- creating the database views used for reporting

Testing and verification:

- DAG must produce the same results as the transforms you've been running manually up to this point.

Naming Conventions:

- Name the DAG file `workflow.py`
- Name the BQ dataset `workflow`

CS 327E Milestone 12 Rubric

Due Date: 05/03/19

<p>Create an Airflow DAG <code>workflow.py</code> of your transformations you specified in the previous milestone. The DAG should run operations in a decently efficient manner (operations that don't depend on one another should run at the same time, etc.)</p> <ul style="list-style-type: none"> -100 <code>workflow.py</code> does not exist in the repository -10 each job that runs in parallel with its dependency -10 each job that runs after another job it does not depend on -50 <code>workflow</code> dataset does not exist in BigQuery 	<p>100</p>
<p><code>submission.json</code> submitted into Canvas. Your project will not be graded without this submission. The file should have the following schema:</p> <pre>{ "commit-id": "your most recent commit ID from Github", "project-id": "your project ID from GCP" }</pre> <p>Example:</p> <pre>{ "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9", "project-id": "some-project-id" }</pre>	<p>Required</p>
<p>Total Credit:</p>	<p>100</p>