



1

Create schema and tables in PostgreSQL database

PythonOperator: create_schema

2

Get twitter data via API call and save data in S3

PythonOperator: get_twitter_data

3a

Create EMR cluster and start Spark and Hadoop application

PythonOperator: create_emr_cluster

3b

Add steps to EMR cluster

PythonOperator: step_adder

3c

Add step checker to EMR cluster

PythonOperator: step_checker

3d

Terminate EMR cluster

PythonOperator: terminate_emr_cluster

4

Save results to PostgreSQL database

PythonOperator: save_result_to_postgres_db