**Twitter Data Pipeline**

This project focuses on building a data pipeline. Where we get data from Twitter using API (Tweepy) and using Python programming language and Apache Airflow load this data on the AWS S3 bucket.

To break into pieces:

1. We will first extract data from Twitter of @username (Elon Musk) using Twitter API called Tweepy. Using the terminal, we will download all the necessary modules we need while coding.
2. Then we will use Python to perform the transformation of the data we obtain and convert data into a CSV file.
3. In the third step, we will deploy our code to Apache Airflow, which will be our AWS EC2 machine. The focus of this tutorial is also on learning about Airflow.
4. In the end, we will store or load the data to an AWS S3 bucket using Airflow.

**Airflow:**

It’s an open-source workflow management platform for the data engineering pipeline. Which is developed by Airbnb. In simple terms, Airflow is a workflow orchestration tool. You can use it to build, schedule, and monitor data pipelines.

Workflow in Airflow means a sequence of tasks known as DAG in Airflow.

DAG is known as a Directed Acyclic Graph. It contains multiple tasks that perform operations. There are different types of DAGs, and you can also create one according to your need.

This tutorial creates a Twitter Dag for creating pipeline.