**BASH COMMANDS:**

**Implementation with the Dataset:**  
  
**Download the dataset:**

wget <https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2023-01.csv>

**Simulate a streaming environment:**

split -l 1000 yellow\_tripdata\_2023-01.csv trip\_chunk\_

**Kafka for Data Ingestion:**

**Create a Kafka Topic:**

kubectl exec -it my-cluster-kafka-0 -- kafka-topics --create --topic taxi\_data --bootstrap-server my-cluster-kafka-bootstrap:9092

**Final Testing**

**Kafka producer to simulate data ingestion:**

python kafka\_producer.py

**Trigger the Airflow DAG:**

airflow dags trigger streaming\_pipeline

**S3 for the processed data:**

aws s3 ls s3://your-bucket/taxi\_streaming/