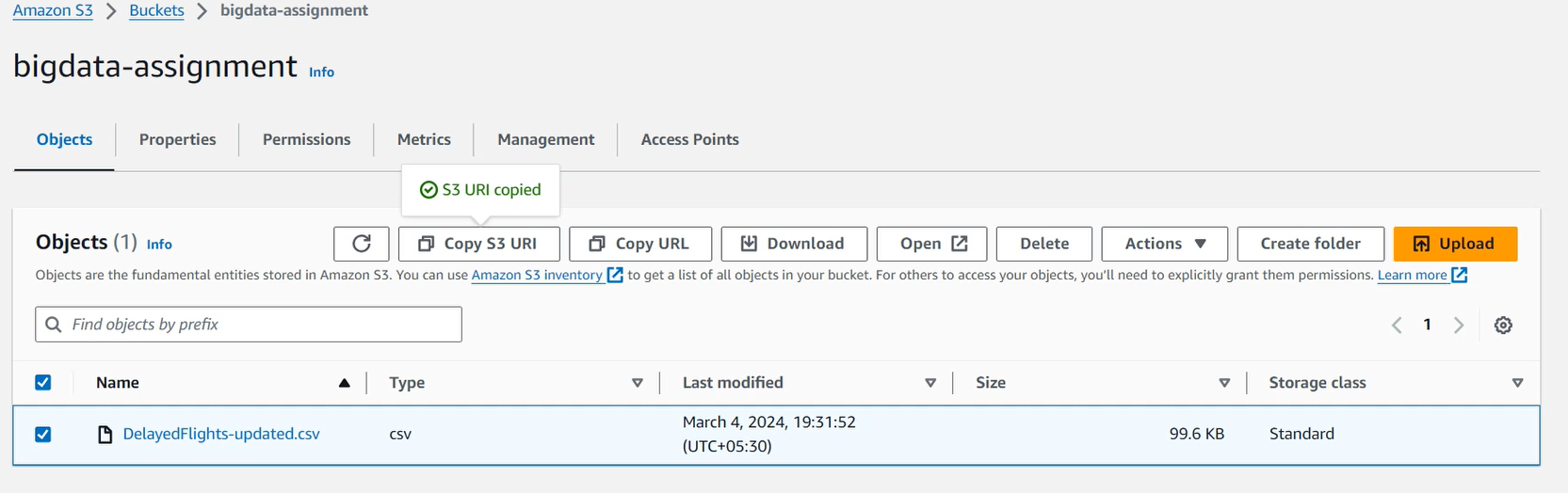
**Big Data Video Presentation**

**Question d:**

**Uploading CSV in S3**



**Spark Initialization**

A black background with white lines

Description automatically generated

**Spark Code**  
from pyspark.sql import SparkSession

# Create a Spark session

spark = SparkSession.builder.appName("DelayFlightsAnalysis").getOrCreate()

# Upload CSV and create DataFrame

csv\_path = "path/to/DelayedFlights-updated.csv"

delay\_flights\_df = spark.read.csv(csv\_path, header=True, inferSchema=True)

# Create a temporary table named delay\_flights

delay\_flights\_df.createOrReplaceTempView("delay\_flights")

# Run the first query using HiveQL

result\_hiveql = spark.sql("""

SELECT Year, AVG((CarrierDelay / ArrDelay) \* 100) AS avg\_delay\_percentage

FROM delay\_flights

GROUP BY Year""").show()

# Run the second query using Spark SQL

result\_sparksql = spark.sql("""

SELECT Year, AVG((CarrierDelay / ArrDelay) \* 100) AS avg\_delay\_percentage

FROM delay\_flights

GROUP BY Year""").show()

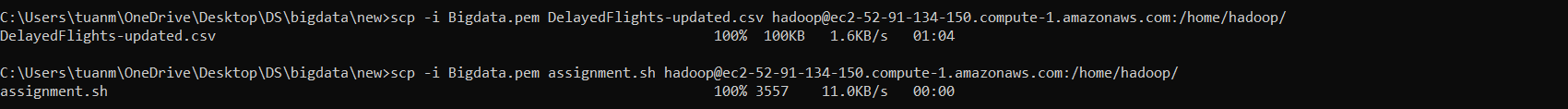
**Question e:**

**EMR Cluster Initialization**

A black screen with a black background

Description automatically generated

**Adding Shell Script and table in Hadoop directory**



**Grant permission to Shell Script and execute**



**Map Reducer Worker distribution**

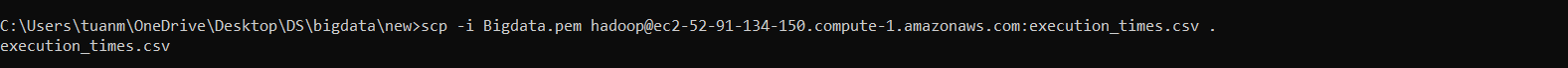
****

**Run Time Calculation**

A black screen with a black background

Description automatically generated

**Download into local path**



**Results**

A white background with black dots

Description automatically generated

A graph with blue and orange bars

Description automatically generated

### Shell Script file is [**assignment\_for\_5\_iteration.sh**](https://github.com/TuanMinhajSeedin/UoM_MapReduce-vs-Spark/blob/main/assignment_for_5_iteration.sh)

**Question f:**

### **Same command approach as above. But Shell script is different. I have attached it in the same folder** [assignment\_for\_5\_queries.sh](https://github.com/TuanMinhajSeedin/UoM_MapReduce-vs-Spark/blob/main/assignment_for_5_queries.sh)

**Results**



**Question g:**

**Results**

A graph of a graph with text

Description automatically generated with medium confidence