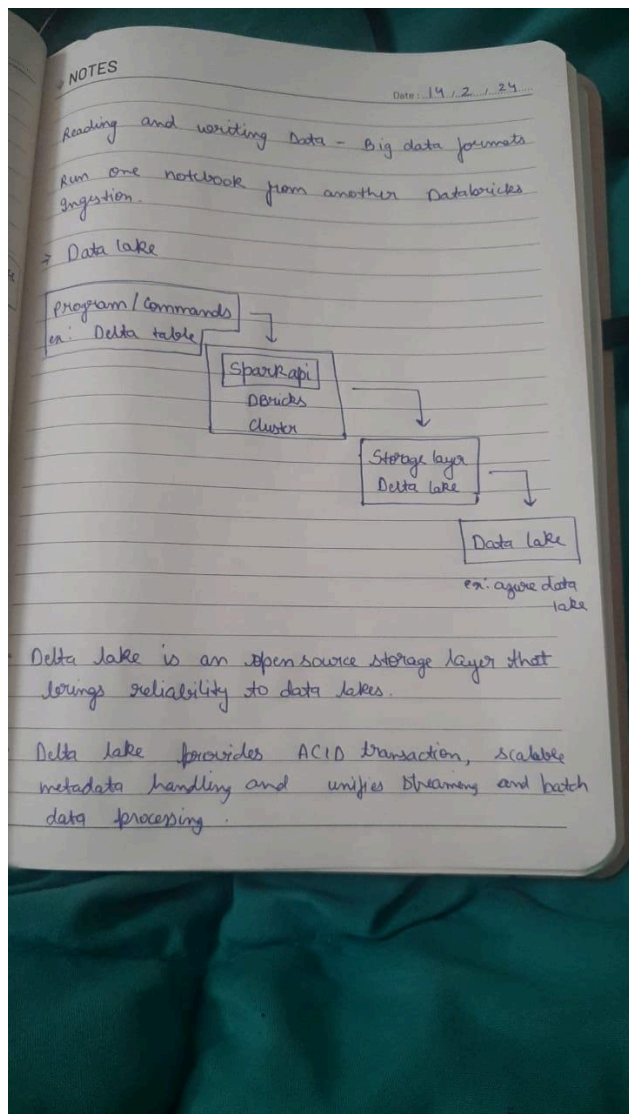


Asmita Porwal
Batch-1
Day-18
14/02/2024
Data engineering

Assignment-18

Hand written notes:



Delta lake :

Untitled Notebook 2024-02-14 15:32:02

Python



File Edit View Run Help Last edit was 11 minutes...

New cell UI: OFF

Run all

azuser1060_mml.local's...

Schedule

Share

Cmd 1

```
1 data = spark.range(0, 5)
2 data.write.format("delta").save("/tmp/delta-table")
```

▶ (6) Spark Jobs

▶ data: pyspark.sql.dataframe.DataFrame = [id: long]

Command took 3.95 seconds -- by azuser1060_mml.local@iiht1.onmicrosoft.com at 2/14/2024, 3:34:58 PM on azuser1060_mml.local's Cluster

Cmd 2

Python

```
1 df = spark.read.format("delta").load("/tmp/delta-table")
2 df.show()
```

▶ (3) Spark Jobs

▶ df: pyspark.sql.dataframe.DataFrame = [id: long]

id
3
4
0
1
2

Cmd 3

```
1 data = spark.range(5, 10)
2 data.write.format("delta").mode("overwrite").save("/tmp/delta-table")
```

▶ (6) Spark Jobs

▶ data: pyspark.sql.dataframe.DataFrame = [id: long]

Command took 2.79 seconds -- by azuser1060_mml.local@iiht1.onmicrosoft.com at 2/14/2024, 3:35:48 PM on azuser1060_mml.local's Cluster

Cmd 4

Python

```
1 df = spark.read.format("delta").option("versionAsOf", 0).load("/tmp/delta-table")
2 df.show()
```

▶ (3) Spark Jobs

▶ df: pyspark.sql.dataframe.DataFrame = [id: long]

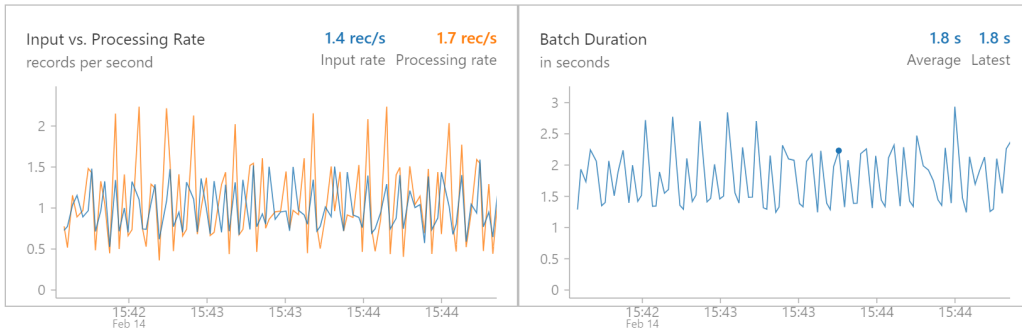
id
3
4
0
1
2

```
1 streamingDf = spark.readStream.format("rate").load()
2 stream = streamingDf.selectExpr("value as id").writeStream.format("delta").option("checkpointLocation", "/tmp/
checkpoint").start("/tmp/delta-table")
```

▶ (1) Spark Jobs

▼ 025e6c96-5104-41f5-a280-106d2b0915fd *Last updated: 4 minutes ago*

Dashboard Raw Data



▶ streamingDf: pyspark.sql.dataframe.DataFrame = [timestamp: timestamp, value: long]