

## Spark Activity-3

Uploading json file using spark SQL data frames

1) Create Scala notebook and upload Singleline.json data. Copy the path of the uploaded data set.

The screenshot shows the 'Upload Data' dialog box. On the left, under 'Uploaded Files', the 'Spark API Format' tab is selected, showing the file path 'red\_uploads/jg.sangeetha@gmail.com/Singleline.json'. On the right, under 'Access Files from Notebooks', the 'Scala' tab is selected, showing the same path. Both sections have a 'Copy' button. A 'Done' button is at the bottom right.

2) Read the json file

```
1 //Reading json file
2 val df=sqlContext.read.json("dbfs:/FileStore/shared_uploads/jg.sangeetha@gmail.com/Singleline.json")
```

► (1) Spark Jobs

► df: org.apache.spark.sql.DataFrame = [color: string, value: string]

df: org.apache.spark.sql.DataFrame = [color: string, value: string]

Command took 27.64 seconds -- by jg.sangeetha@gmail.com at 6/23/2021, 7:20:35 PM on My Cluster

3) Printing the schema

```

1 //Printing the schema
2 df.printSchema()

root
 |-- color: string (nullable = true)
 |-- value: string (nullable = true)

```

4) Save the file as temporary table

```

1 //saving as temporary table
2 df.createTempView("Json")

```

5) Select all data from temporary table and store it in a different variable

```

1 val data=sqlContext.sql("select * from Json")

▶ data: org.apache.spark.sql.DataFrame = [color: string, value: string]
data: org.apache.spark.sql.DataFrame = [color: string, value: string]

```

6) Display all the records

```

1 data.show()

▶ (1) Spark Jobs

+-----+-----+
| color|value|
+-----+-----+
|   red| #f00|
| green| #0f0|
|  blue| #00f|
|  cyan| #0ff|
|magenta| #f0f|
| yellow| #ff0|
|  black| #000|
+-----+-----+

```

7) Follow step 1 to step 6 for processing multiline json file.