In [1]: `print('mukesh')`

Starting Spark application

ID	YARN Application ID	Kind	State	Spark UI
0	application_1547533712041_0004	pyspark3	idle	[Link (https://suyalmukesh.azurehdinsight.net/y](https://suyalmukesh.azurehdinsight.net/y)

SparkSession available as 'spark'.
mukesh

In [2]: `read_input_file = spark.read.csv("wasb:///data/movies.csv")`

In [3]: `read_input_file.printSchema()`
`read_input_file.show()`

```
root
 |-- _c0: string (nullable = true)
 |-- _c1: string (nullable = true)
 |-- _c2: string (nullable = true)

+-------+--------------------+--------------------+
|    _c0|                 _c1|                 _c2|
+-------+--------------------+--------------------+
|movieId|               title|              genres|
|      1|    Toy Story (1995)|Adventure|Animati...|
|      2|      Jumanji (1995)|Adventure|Childre...|
|      3|Grumpier Old Men ...|      Comedy|Romance|
|      4|Waiting to Exhale...|Comedy|Drama|Romance|
|      5|Father of the Bri...|              Comedy|
|      6|         Heat (1995)|Action|Crime|Thri...|
|      7|      Sabrina (1995)|      Comedy|Romance|
|      8| Tom and Huck (1995)|  Adventure|Children|
|      9| Sudden Death (1995)|              Action|
|     10|    GoldenEye (1995)|Action|Adventure|...|
|     11|American Presiden...|Comedy|Drama|Romance|
|     12|Dracula: Dead and...|      Comedy|Horror|
|     13|        Balto (1995)|Adventure|Animati...|
|     14|        Nixon (1995)|               Drama|
|     15|Cutthroat Island ...|Action|Adventure|...|
|     16|       Casino (1995)|        Crime|Drama|
|     17|Sense and Sensibi...|       Drama|Romance|
|     18|   Four Rooms (1995)|              Comedy|
|     19|Ace Ventura: When...|              Comedy|
+-------+--------------------+--------------------+
only showing top 20 rows
```

```
In [69]:   from pyspark.sql.types import *
           schema = StructType([StructField('movieId', IntegerType(), True),
                                StructField('Title', StringType(), True),
                                StructField('Generes', StringType(), True)])
           read_file2 = spark.read.csv("wasb:///data/movies.csv",schema=schema)
           #read_file2 = spark.read.csv("wasb:///data/movies.csv",inferSchema=True)
           #Header = read_file2.first()
           #read_file2.show(truncate=False)

           read_file2.registerTempTable('movies')

           #spark.sql("create table movies_new (movieId string,title string ,generes string
           #            stored as orc  ")

           #read_file2.show()
           #spark.sql("select Title,Generes from movies").show()
           #spark.sql("insert into movies_new select * from movies")
           #spark.sql("select * from movies_new").show()
           spark.sql("select count(*) from movies_new").show()
```

```
+--------+
|count(1)|
+--------+
|   58099|
+--------+
```

```
In [9]:    from pyspark.sql import HiveContext
           hc = HiveContext(sc)
           read_input_file.write.format("orc").saveAsTable("employees")
```

In [19]:
```python
read_file2.show(truncate=False)
#spark.sql("select * from employees").show(truncate=False)
```

```
+----+--------+----+
|Name|DateTime|Age |
+----+--------+----+
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
|null|null    |null|
+----+--------+----+
only showing top 20 rows
```

In [13]:
```python
spark.sql("select count(*) from employees").show()
```

```
+--------+
|count(1)|
+--------+
|   58099|
+--------+
```

```
In [17]: spark.sql("select * from employees where _c2 like '%Action%'").show(truncate=False
         spark.sql("create view Action_movies as select * from employees where _c2 like '%
         spark.sql("select * from Action_movies").show(truncate=False)
```

```
+---+------------------------------------------+-------------------------------
---+
|_c0|_c1                                       |_c2
   |
+---+------------------------------------------+-------------------------------
---+
|6  |Heat (1995)                               |Action|Crime|Thriller
   |
|9  |Sudden Death (1995)                       |Action
   |
|10 |GoldenEye (1995)                          |Action|Adventure|Thriller
   |
|15 |Cutthroat Island (1995)                   |Action|Adventure|Romance
   |
|20 |Money Train (1995)                        |Action|Comedy|Crime|Drama|Thril
ler|
|23 |Assassins (1995)                          |Action|Crime|Thriller
   |
|42 |Dead Presidents (1995)                    |Action|Crime|Drama
   |
|44 |Mortal Kombat (1995)                      |Action|Adventure|Fantasy
   |
|51 |Guardian Angel (1994)                     |Action|Drama|Thriller
   |
|66 |Lawnmower Man 2: Beyond Cyberspace (1996) |Action|Sci-Fi|Thriller
   |
|70 |From Dusk Till Dawn (1996)                |Action|Comedy|Horror|Thriller
   |
|71 |Fair Game (1995)                          |Action
   |
|76 |Screamers (1995)                          |Action|Sci-Fi|Thriller
   |
|78 |Crossing Guard, The (1995)                |Action|Crime|Drama|Thriller
   |
|86 |White Squall (1996)                       |Action|Adventure|Drama
   |
|89 |Nick of Time (1995)                       |Action|Thriller
   |
|95 |Broken Arrow (1996)                       |Action|Adventure|Thriller
   |
|98 |Shopping (1994)                           |Action|Thriller
   |
|110|Braveheart (1995)                         |Action|Drama|War
   |
|112|Rumble in the Bronx (Hont faan kui) (1995)|Action|Adventure|Comedy|Crime
   |
+---+------------------------------------------+-------------------------------
---+
only showing top 20 rows

+---+------------------------------------------+-------------------------------
---+
```

```
|_c0|_c1                                        |_c2
   |
+---+------------------------------------------+------------------------------
---+
|6  |Heat (1995)                               |Action|Crime|Thriller
   |
|9  |Sudden Death (1995)                       |Action
   |
|10 |GoldenEye (1995)                          |Action|Adventure|Thriller
   |
|15 |Cutthroat Island (1995)                   |Action|Adventure|Romance
   |
|20 |Money Train (1995)                        |Action|Comedy|Crime|Drama|Thril
ler|
|23 |Assassins (1995)                          |Action|Crime|Thriller
   |
|42 |Dead Presidents (1995)                    |Action|Crime|Drama
   |
|44 |Mortal Kombat (1995)                      |Action|Adventure|Fantasy
   |
|51 |Guardian Angel (1994)                     |Action|Drama|Thriller
   |
|66 |Lawnmower Man 2: Beyond Cyberspace (1996) |Action|Sci-Fi|Thriller
   |
|70 |From Dusk Till Dawn (1996)                |Action|Comedy|Horror|Thriller
   |
|71 |Fair Game (1995)                          |Action
   |
|76 |Screamers (1995)                          |Action|Sci-Fi|Thriller
   |
|78 |Crossing Guard, The (1995)                |Action|Crime|Drama|Thriller
   |
|86 |White Squall (1996)                       |Action|Adventure|Drama
   |
|89 |Nick of Time (1995)                       |Action|Thriller
   |
|95 |Broken Arrow (1996)                       |Action|Adventure|Thriller
   |
|98 |Shopping (1994)                           |Action|Thriller
   |
|110|Braveheart (1995)                         |Action|Drama|War
   |
|112|Rumble in the Bronx (Hont faan kui) (1995)|Action|Adventure|Comedy|Crime
   |
+---+------------------------------------------+------------------------------
---+
only showing top 20 rows
```

In [ ]: