

Q1

```
# sado sa
# start -all.sh
# jps
# hdfs dfs -put |home|dbda|Desktop|
```

BDT Exam / Products - 20221219 To 24048 2000 | Products | part-00000 | Dataset of Product

In browser

Local host : 50070

* Use this command to start Spark-shell.

```
# Spark-shell --packages org.apache.spark:spark-avro_2.12:3.3.1
```

```
# import org.apache.spark.sql.SQLContext
```

```
scala> val ssq = new SQLContext(sc)
```

```
# spark.read.csv(("hdfs://localhost:9000/Dataset of-product")) create Or Replace Temp View("products")
```

```
# ssq.sql("select * from products limit 5").show()
```

```
# ssq.sql("select _C0 as product_id, _C2 as product_name, _C4 as product-price from prod")
```

```
# sudo su
# start -a -sh
# jps
```

```
# hdfs dfs -put
```

```
BDT Exam | Products - 20241219 To 24548 2-7  
001 | Products | part-00000 | Dataset of Product
```

In browser

Local host : 50070

Use this command to start Spark-shell.

```
# Spark-shell --packages org.apache.spark:  
spark-avro_2.12:3.3.1
```

```
# import org.apache.spark.sql.SQLContext
```

```
scala> val ssq = new SQLContext(sc)
```

```
# spark.read.csv("hdfs://localhost:9000/Dataset  
of-product") create Or Replace Temp  
View("products")
```

```
# ssq.sql("select * from products limit 5").  
show()
```

```
# ssq.sql("select _C0 as product_id, _C2 as  
product-price from product")
```


• create TempView ("Products1")

```
# ssq.sql("select * from products1 limit 5")  
show()
```

→ Adding Filter & Saving it in another table "products2"

```
# ssq.sql("select * from products1 where  
product_price < 1000 and product_name  
like '%vector series%'").createTempView  
("products2")
```

```
# ssq.sql("select * from products2 limit 5")  
show()
```

* Saving the o/p.

```
# ssq.sql("select * from products2").  
write.option("compression", "deflate").  
format("com.databricks.spark.avro").  
save("hdfs://localhost:9000/Exam/q1")
```

Q2.

classmate

Date
Page

```
val q2 = sc.textFile ("home/dbda/Desktop/part-00000")
```

```
val q3 = q2.map(x => (x.substring(0,1),  
x.substring(2)))
```

```
q3.saveAsSequenceFile ("hdfs://localhost:  
9000/Exam/Q2/seq")
```

```
q3.SavesaveAsTextFile ("hdfs://localhost:9000/  
Exam/Q2/Q2_text_format")
```