

```
//initialize docker
```

```
sdc-setup
```

```
eval "$(triton env --docker env)"
```

```
//download docker image
```

```
docker pull oonlyo/spark:2.2
```

```
//run spark master
```

```
docker run -d --network zones --label com.joyent.package=large --name spark1 -e  
SPARK_TYPE="master" oonlyo/spark:2.2
```

```
//Check Spark master ip
```

```
master=spark://$(docker inspect -f '{{ .NetworkSettings.IPAddress }}'  
spark1):7077
```

```
echo $master
```

```
//run spark slave
```

```
docker run -d --network zones --label com.joyent.package=huge --name spark2 -e  
SPARK_TYPE=slave -e SPARK_MASTERS=$master -e SPARK_WORKER_CORES=2  
oonlyo/spark:2.2
```

```
docker run -d --network zones --label com.joyent.package=huge --name spark3 -e  
SPARK_TYPE=slave -e SPARK_MASTERS=$master -e SPARK_WORKER_CORES=2  
oonlyo/spark:2.2
```

```
docker ps
```

```
[[sdc:s4522613] s4522613@mos:~$ docker run -d --network zones --label com.joyent.package=large --name spark1 -e SPARK_TYPE="master" oonlyo/spark:2.2  
98b9bc46b4aad34e396c9360eedf74c54d3ec8136f1e74f852aa117b84c1d1c  
ERROR[0021] error getting events from daemon: Error response from daemon: (NotImplemented) events is not implemented (436a3983-0031-4246-abe2-de29a5de2409)
```

```
[sdc:s4522613] s4522613@moss:~$ master=spark://$(docker inspect -f '{{.NetworkSettings.IPAddress}}' spark1):7077
[sdc:s4522613] s4522613@moss:~$ echo $master
spark://172.23.99.155:7077
[sdc:s4522613] s4522613@moss:~$ docker run -d --network zones --label com.joyent.package=huge --name spark2 -e SPARK_TYPE=slave -e SPARK_MASTERS=$master -e SPARK_WORKER_CORES=2 oonlyo/spark:2.2
835d5c4e7984c457cc2e8686bda2b697d168d36d23b4f3fa5d3f4d2c4680a4e
ERROR[0021] error getting events from daemon: Error response from daemon: (NotImplemented) events is not implemented (d89bd3aa-a45c-4787-a909-43dd0e37d190)
[sdc:s4522613] s4522613@moss:~$ docker run -d --network zones --label com.joyent.package=huge --name spark3 -e SPARK_TYPE=slave -e SPARK_MASTERS=$master -e SPARK_WORKER_CORES=2 oonlyo/spark:2.2
9e3b735785636453809eb38c49715f4994961127f13cea65e83ddbc6a67dddc5
ERROR[0021] error getting events from daemon: Error response from daemon: (NotImplemented) events is not implemented (50e1ef4d-fca6-47f2-8635-8ab65d1602f5)
[sdc:s4522613] s4522613@moss:~$
```

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NAMES
9e3b73578563	oonlyo/spark:2.2	"/entrypoint.sh"	23 minutes ago	Up 23 minutes		spark3
835d5c4e7984	oonlyo/spark:2.2	"/entrypoint.sh"	36 minutes ago	Up 36 minutes		spark2
98b9bc46b4aa	oonlyo/spark:2.2	"/entrypoint.sh"	38 minutes ago	Up 38 minutes		spark1
4497e2e5c14e	oonlyo/phpcass	"docker-php-entryp..."	8 days ago	Up 8 days	80/tcp, 443/tcp, 9000/tcp	phpcass
9c88b166bf40	oonlyo/zeppelin:0.8	"/entrypoint.sh"	2 weeks ago	Up 46 hours		s4522613

```
//access Cassandra keyspark.table
val keySpace = "book"
val table = "orders"
val df = sqlContext.read.format("org.apache.spark.sql.cassandra").options(Map("keyspace" -> keySpace, "table" -> table)).load()
df.createOrReplaceTempView(table)
```

```
keySpace: String = book
table: String = orders
df: org.apache.spark.sql.DataFrame = [orderid: int, bookid: int ... 3 more fields]
```

Took 7 sec. Last updated by anonymous at October 12 2018, 3:45:00 PM.

```
%sql
/*retrieve all orders of a table*/
select * from orders
```

SPARK JOBS FINISHED

settings

orderid	bookid	orderdate	rating	userid
1023	88752	2016-02-28	5	5
1021	37858	2017-03-26	4	10
1000	88752	2018-05-10	4	1
1010	37858	2017-10-28	4	10
1001	75510	2018-10-02	3	4
1018	27723	2017-06-18	4	9
1012	71046	2018-07-19	2	10
1022	48501	2017-07-14	3	4

Took 11 sec. Last updated by anonymous at October 12 2018, 3:45:11 PM.

```
//Import Spark ML library
import org.apache.spark.mllib.recommendation.ALS
import org.apache.spark.mllib.recommendation.MatrixFactorizationModel
import org.apache.spark.mllib.recommendation.Rating
//get data from cassandra
val test = df.select($"userid", $"bookid", $"rating")
//Each row in the score matrix RDD is converted to a Rating type
val records = test.as[(Int,Int,Double)].map{
  case (userid, bookid, rating) => Rating(userid.toInt, bookid.toInt, rating.toDouble)
}
//convert spark row type to spark rdd type
val ratings = records.rdd
//ALS
val model = ALS.train(ratings, 10, 10, 0.01)
//input user id
val userId = 10
//recommend 4 books to user 10
val K = 4
//training
val topKRecs = model.recommendProducts(userId, K)
// the first col is user id, second col is bookid recommend to user, third col is rating
println(topKRecs.mkString("\n"))

import org.apache.spark.mllib.recommendation.ALS
import org.apache.spark.mllib.recommendation.MatrixFactorizationModel
import org.apache.spark.mllib.recommendation.Rating
test: org.apache.spark.sql.DataFrame = [userid: int, bookid: int ... 1 more field]
records: org.apache.spark.sql.Dataset[org.apache.spark.mllib.recommendation.Rating] = [user: int, product: int ... 1 more field]
ratings: org.apache.spark.rdd.RDD[org.apache.spark.mllib.recommendation.Rating] = MapPartitionsRDD[450] at rdd at <console>:46
model: org.apache.spark.mllib.recommendation.MatrixFactorizationModel = org.apache.spark.mllib.recommendation.MatrixFactorizationModel@40472766
userid: Int = 10
K: Int = 4
topKRecs: Array[org.apache.spark.mllib.recommendation.Rating] = Array(Rating(10,37858,3.9958023165146295), Rating(10,20506,2.996851746188752), Rating(10,85598,2.24470838833372), Rating(10,71046,1.9979011582573147))
Rating(10,37858,3.9958023165146295)
Rating(10,20506,2.996851746188752)
Rating(10,85598,2.24470838833372)
Rating(10,71046,1.9979011582573147)
```

SPARK JOBS FINISHED

Took 20 sec. Last updated by anonymous at October 12 2018, 3:45:32 PM.