**Cassandra Join operation using Apache Spark – Cassandra connector**

**Installation**

1. Java 1.8+
   1. JAVA\_HOME
2. Cassandra 3.11.4
3. spark-2.4.3-bin-hadoop2.7
4. Hadoop Winutils
   1. HADOOP\_HOME

**Step1:** Up Cassandra process

**Step2:** Open command prompt at **spark-2.4.3-bin-hadoop2.7\bin**

(It will open scala prompt)

**Step4:** Enter following command to load spark shell and all necessary libs

**spark-shell --packages com.datastax.spark:spark-cassandra-connector\_2.11:2.4.1 --conf spark.cassandra.connection.host=localhost –verbose**

**Step 5:** Check your tables at Cassandra to load it into spark

Ex. I have **student** table

**Step 6:** Enter following scrip to load needed table to spark

**val student = spark.**

**read.**

**format("org.apache.spark.sql.cassandra").**

**options(Map("table"->"student","keyspace"->"javafrm")).**

**load()**

**Step7:**Check table using below command

**student.show()**

**Step 8:** Enter following scrip to load another table to spark

Ex. I have **school** table

**val school= spark.**

**read.**

**format("org.apache.spark.sql.cassandra").**

**options(Map("table"->"school","keyspace"->"javafrm")).**

**load()**

**school.show()**

**Step 9:** Temporary register both table in spark

**student .registerTempTable("student")**

**school.registerTempTable("school")**

**Step 10:** Write Join query to registered table

**sql("select t1.id,t1.name from student t1 left join school t2 on t1.id = t2.id").show()**

**Step 10:** Write Join query to registerd table

**sql("select t1.emp\_id, t1.emp\_name from employee\_table t1 left join employee\_table t2 on t1.emp\_id = t2.emp\_id").show()**