Cassandra + Spark + Python

We will use the [spark-cassandra-connector](https://github.com/datastax/spark-cassandra-connector) to bring Spark and Cassandra together. When you run a Spark job using this library, you need to include the corresponding [Spark Package](https://spark-packages.org/):

spark-submit --packages datastax:spark-cassandra-connector:2.3.1-s\_2.11 …

You need to configure the SparkSession object to connect correctly to our cluster. Create your spark variable like this:

cluster\_seeds = ['199.60.17.188', '199.60.17.216']

spark = SparkSession.builder.appName('Spark Cassandra example') \

.config('spark.cassandra.connection.host', ','.join(cluster\_seeds)).getOrCreate()

With this done, you should be able to read DataFrames from Cassandra or write DataFrames tables to Cassandra:

df = spark.read.format("org.apache.spark.sql.cassandra") \

.options(table=table, keyspace=keyspace).load()

df.write.format("org.apache.spark.sql.cassandra") \

.options(table=table, keyspace=keyspace).save()