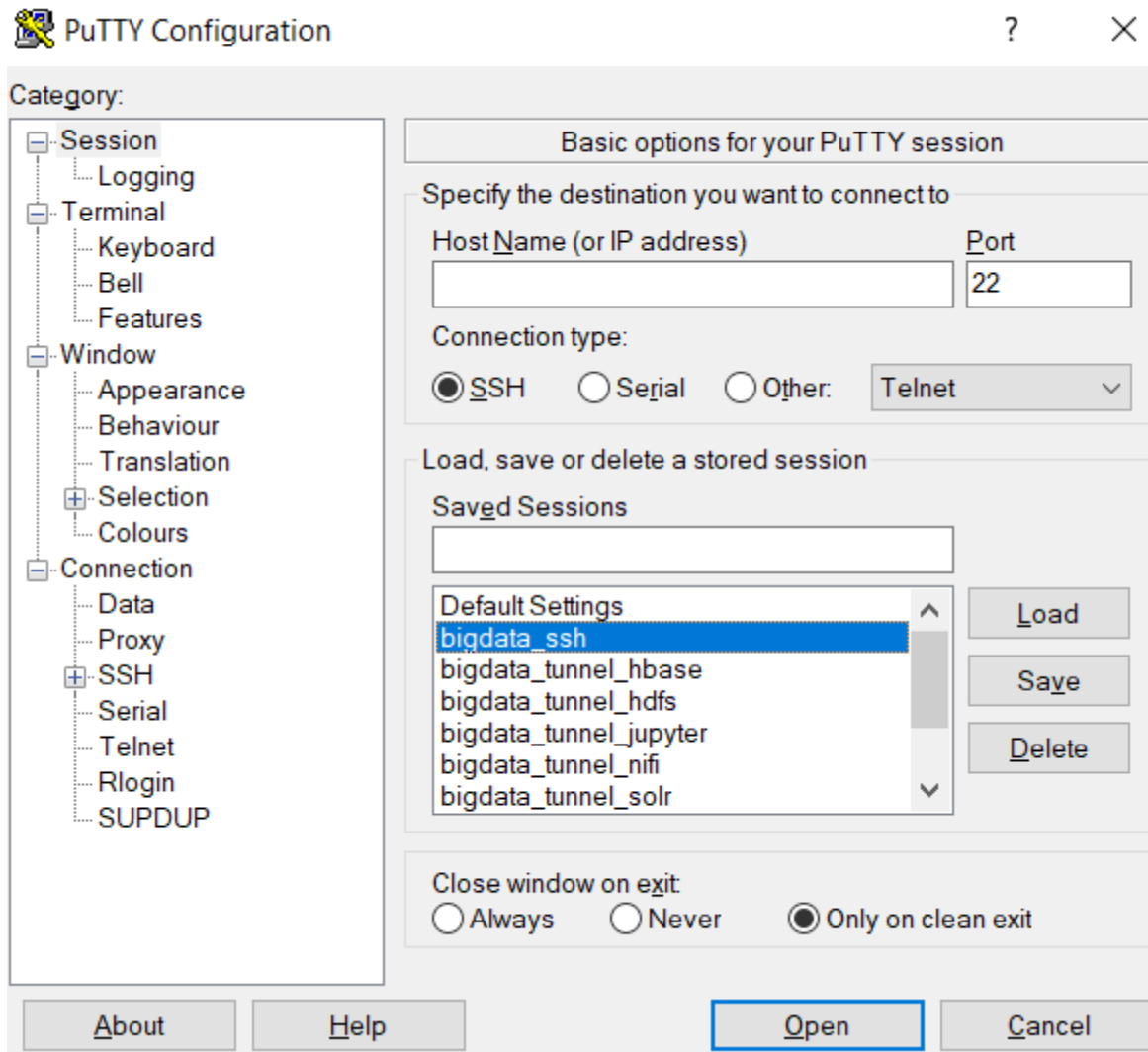


# SparkSQL and Scala with Google Cloud Instance and Putty

- **Putty Configuration**



- **Navigating to Directory and Starting Docker Containers:**

```
louis@bigdata:~/dsc650-infra/bellevue-bigdata$ cd hadoop-hive-spark-hbase
louis@bigdata:~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase$ docker-compose up -d
Creating network "hadoop-hive-spark-hbase_spark_net" with the default driver
Creating hadoop-hive-spark-hbase_zoo2_1 ... done
Creating hadoop-hive-spark-hbase_hivemetastore_1 ... done
Creating hadoop-hive-spark-hbase_zoo1_1 ... done
Creating hadoop-hive-spark-hbase_zoo3_1 ... done
Creating hadoop-hive-spark-hbase_worker2_1 ... done
Creating hadoop-hive-spark-hbase_master_1 ... done
Creating hadoop-hive-spark-hbase_worker1_1 ... done
louis@bigdata:~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase$
```



```
scala> spark.sql("SHOW TABLES").show()
110176 [main] WARN org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.strict.managed.tables does not exist
110176 [main] WARN org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.create.as.insert.only does not exist
110210 [main] WARN org.apache.spark.sql.hive.client.HiveClientImpl - Detected HiveConf hive.execution.engine is 'tez' and will be reset to 'mr' to disable useless hive logic
+-----+-----+-----+
|database|tableName|isTemporary|
+-----+-----+-----+
|         |         |         |
+-----+-----+-----+

scala> spark.sql("SELECT * FROM df WHERE Final > 50").show()
+-----+-----+-----+-----+-----+-----+-----+
|Last name|First name|SSN|Test1|Test2|Test3|Test4|Final|Grade|
+-----+-----+-----+-----+-----+-----+-----+
| Airpump| Andrew|223-45-6789| 49| 1| 90| 100| 83| A|
| Backus| Jim|143-12-1234| 48| 1| 97| 96| 97| A+|
| Elephant| Ima|456-71-9012| 45| 1| 78| 88| 77| B-|
| Franklin| Benny|234-56-2890| 50| 1| 90| 80| 90| B-|
+-----+-----+-----+-----+-----+-----+-----+

scala> spark.sql("SELECT * FROM df").show()
+-----+-----+-----+-----+-----+-----+-----+
|Last name|First name|SSN|Test1|Test2|Test3|Test4|Final|Grade|
+-----+-----+-----+-----+-----+-----+-----+
| Alfalfa| Aloysius|123-45-6789| 40| 90| 100| 83| 49| D-|
| Alfred|University|123-12-1234| 41| 97| 96| 97| 48| D+|
| Gerty| Gramma|567-89-0123| 41| 80| 60| 40| 44| C|
| Android| Electric|087-65-4321| 42| 23| 36| 45| 47| B-|
| Bumpkin| Fred|456-78-9012| 43| 78| 88| 77| 45| A-|
| Rubble| Betty|234-56-7890| 44| 90| 80| 90| 46| C-|
| Noshov| Cecil|345-67-8901| 45| 11| -1| 4| 43| F|
| Buff| Bif|632-79-9939| 46| 20| 30| 40| 50| B+|
| Airpump| Andrew|223-45-6789| 49| 1| 90| 100| 83| A|
| Backus| Jim|143-12-1234| 48| 1| 97| 96| 97| A+|
| Carnivore| Art|565-89-0123| 44| 1| 80| 60| 40| D+|
| Dandy| Jim|087-75-4321| 47| 1| 23| 36| 45| C+|
| Elephant| Ima|456-71-9012| 45| 1| 78| 88| 77| B-|
| Franklin| Benny|234-56-2890| 50| 1| 90| 80| 90| B-|
| George| Boy|345-67-3901| 40| 1| 11| -1| 4| B|
| Heffalump| Harvey|632-79-9439| 30| 1| 20| 30| 40| C|
+-----+-----+-----+-----+-----+-----+-----+
```

```
scala> spark.sql("select * from df where Test3>80").show()
+-----+-----+-----+-----+-----+-----+-----+
|Last name|First name|      SSN|Test1|Test2|Test3|Test4|Final|Grade|
+-----+-----+-----+-----+-----+-----+-----+
|  Alfalfa|  Aloysius|123-45-6789|  40|  90| 100|  83|  49|  D-|
|  Alfred|University|123-12-1234|  41|  97|  96|  97|  48|  D+|
|  Bumpkin|      Fred|456-78-9012|  43|  78|  88|  77|  45|  A-|
|  Airpump|   Andrew|223-45-6789|  49|   1|  90| 100|  83|   A|
|   Backus|      Jim|143-12-1234|  48|   1|  97|  96|  97|  A+|
| Franklin|   Benny|234-56-2890|  50|   1|  90|  80|  90|  B-|
+-----+-----+-----+-----+-----+-----+-----+

scala> spark.sql("select Test4 from df where Test4<90").show()
+-----+
|Test4|
+-----+
|  83|
|  40|
|  45|
|  77|
|   4|
|  40|
|  60|
|  36|
|  88|
|  80|
|  -1|
|  30|
+-----+

scala> spark.sql("select Test1,Test2 from df where (Test1<50 and Test2>90)").show()
+-----+-----+
|Test1|Test2|
+-----+-----+
|  41|  97|
+-----+-----+
```