**Unit Testing**

Test Script to check the number of records matching between the Postgres and Hive tables after the Full Load and the Incremental Load.

To validate the missing or NULL values in any of the columns, duplicate rows, primary key constraint violation in both the Postgres and Hive table both for individual column and for the whole table.

We can display the data types of each column and also the datatypes are matching between the Postgres and the Hive table.

We can pass the data type along with the required column and check whether it is satisfying the condition.

Test Scripts are written both in Pyspark & ScalaSpark.

Pyspark GitHub Path- <https://github.com/Asadkardame/BD_Colab/blob/QA>

1. **FullLoad\_Count\_Check** – To check the count after FullLoad or the Incremental Load
2. **IncrLoad\_Count\_Check** – To match the count only with the incremented record provided the incremental record key
3. **Postgres\_Validation** – Do data validation on the postgres data
4. **Hive\_Validation** – Data Validation on the Hive table
5. **Schema** – Compares the Postgres and the Hive table

**Jenkins Shell Command - spark-submit --master local[\*] \**

**--jars /var/lib/jenkins/workspace/nagaranipysparkdryrun/lib/postgresql-42.5.3.jar \**

**Pyspark\_Incr\_Load.py | tee C:\Users\anusu\Documents\BDpython\Count\_Incr\_Load.log**

==================================================================================

ScalaSpark GitHub Path <https://github.com/AnusuyaKal/SparkScala/blob/master/src/main/scala/scalatest/>

1. **FullLoad\_Count\_Check** – To check the count after FullLoad or the Incremental Load
2. **IncrLoad\_Count\_Check**  – To match the count only with the incremented record provided the incremental record key

**Jenkins Shell Command -**

**mvn package**

**spark-submit --master local --jars /var/lib/jenkins/workspace/nagaranipysparkdryrun/lib/postgresql-42.5.3.jar --class scalatest.FullLoad\_Count\_Check target/FullLoad\_Count\_Check-0.2.jar | tee C:\Users\anusu\Documents\BDpython\FullLoad\_Count\_Check.log**

=================================================================================

**Kafka Testing**

Kafka\_Test – compares the count between the API and the topic

Hbase\_Test – takes the count from the HBase.