#Impala Table creation syntax to point to Parquet file which was created by PySpark code.

#Since this cluster is created locally with just Impala configured, I had moved the file from S3 bucket to Linux and then from there to HDFS

#Command used: hdfs dfs -put will take the file from Linux file system and put over to HDFS file system.

create table recipe\_dataframe\_parquet3(cooktime string, datePublished string, description string, image string, ingredients string, name string, preptime string, recipeYield string, url string, newCol string, ts timestamp) STORED AS PARQUET

LOCATION '/tmp/fin\_parquet/fin\_parquet/';

#Post this operation, we will have Impala table created and we are good to have sql queries executed on top of it. For ex:

Highlight query will return 2 records that the Parquet file holds:

