**Spark SQL**

Using the parquet files you created in the last project, [load the data](http://spark.apache.org/docs/latest/sql-programming-guide.html#loading-data-programmatically) in as a table in Spark SQL.

For this lab, we will want to return to Zeppelin. While the same problem that forced us to leave Zeppelin (the trouble with loading JSON into a DataFrame) still exists, it no longer effects us, because the data we want is now in a Parquet table, which does not pose a problem. While Zeppelin leaves much to be desired compared to Jupyter when it comes to Python, it does quite well with SQL, which is what we will be using in this project.

Jupyter Notebook can be run with PySpark but not with Spark SQL, not directly. We can issue Spark SQL commands in PySpark (and therefore in Jupyter) using the [spark.sql](http://spark.apache.org/docs/latest/api/python/pyspark.sql.html#pyspark.sql.SparkSession.sql) method, you'll probably find it more convenient to use the %sql magic in Zeppelin.

(As an aside, you may remember using a %sql magic in Jupyter in 6002. Unfortunately, that will not work for us here because that extension depends on SQLAlchemy which doesn't (yet) know how to interface with Spark SQL.)

Now, [load](http://spark.apache.org/docs/latest/sql-programming-guide.html#loading-data-programmatically) the three tables you created in the last project using:

CREATE TEMPORARY VIEW <table\_name>

USING org.apache.spark.sql.parquet

OPTIONS (

path "s3a://<my\_parquet\_output>"

)

Explore these tables a bit (*e.g.* SELECT \* FROM <table\_name> LIMIT 5 *etc.*) Make sure GROUP BY and ORDER BY work as expected by reproducing the list of top 10 artists.

Now I want you to do something a little more complicated: create a list of the top 10 artists as before, but this time there should be two new columns:

1. the timestamp of the first event that had that artist
2. the first and last name of the user that generated that event

You may find it useful to create one or two temporary views on the way to this final goal.

**Submission:**Submit your Zeppelin notebook as a JSON file.