

Ingesting Database changes via Sqoop/Hudi

Very simple in just 2 steps.

Step 1: Extract new changes to users table in MySQL, as avro data files on DFS

```
// Command to extract incrementals using sqoop
bin/sqoop import \
  -Dmapreduce.job.user.classpath.first=true \
  --connect jdbc:mysql://localhost/users \
  --username root \
  --password ***** \
  --table users \
  --as-avrodatafile \
  --target-dir \
  s3:///tmp/sqoop/import-1/users
```

Step 2: Use your fav datasource to read extracted data and directly “upsert” the users table on DFS/Hive

```
// Spark Datasource
Import org.apache.hudi.DataSourceWriteOptions._
// Use Spark datasource to read avro
Dataset<Row> inputDataset spark.read.avro('s3:///tmp/sqoop/import-1/users/*');

// save it as a Hudi dataset
inputDataset.write.format("org.apache.hudi")
  .option(HoodieWriteConfig.TABLE_NAME, "hoodie.users")
  .option(RECORDKEY_FIELD_OPT_KEY(), "userID")
  .option(PARTITIONPATH_FIELD_OPT_KEY(), "country")
  .option(PRECOMBINE_FIELD_OPT_KEY(), "last_mod")
  .option(OPERATION_OPT_KEY(), UPSERT_OPERATION_OPT_VAL())
  .mode(SaveMode.Append);
  .save("/path/on/dfs")
```

Alternatively, you can also use the Hudi [DeltaStreamer](#) tool with the DFSSource.