Store & Access Data in Non-Text Format

Exercise Path: ~/workspace/data-format

MySQL database: lanier

MySQL table: accounts

**And in this lab, you will use import data in Avro format and create a**

**Hive table to access it.**

1. Change directories to the Exercise Path.

$ cd workspace/data-format

1. Import the accounts table to an Avro data format.

$ sqoop import --connect jdbc:mysql://localhost/lanier

--username [username] --password [password] --table accounts

--target-dir lanier/accounts\_avro --null-non-string '\\N'

--as-avrodatafile

1. View the files imported by Sqoop into HDFS using the hdfs command. What do you see when you try to view the content of the data files?
2. Sqoop generates a schema (in the file accounts.avsc) in the current directory. Review this file and copy it to HDFS below lanier/.

hdfs dfs -put accounts.avsc lanier/

1. In Hive, create a table using this schema:

CREATE EXTERNAL TABLE

accounts\_avro STORED AS AVRO TBLPROPERTIES

('avro.schema.url'= 'hdfs:/user/[username]/lanier/ accounts.avsc')

1. Confirm correct creation of the table by issuing a query, such as:

SELECT \* FROM accounts\_avro LIMIT 10

1. Use the DESCRIBE or DESCRIBE FORMATTED command to list the columns and data types of the accounts\_avro table created from the Avro schema.
2. Then use terminal to locate your data.

**END**