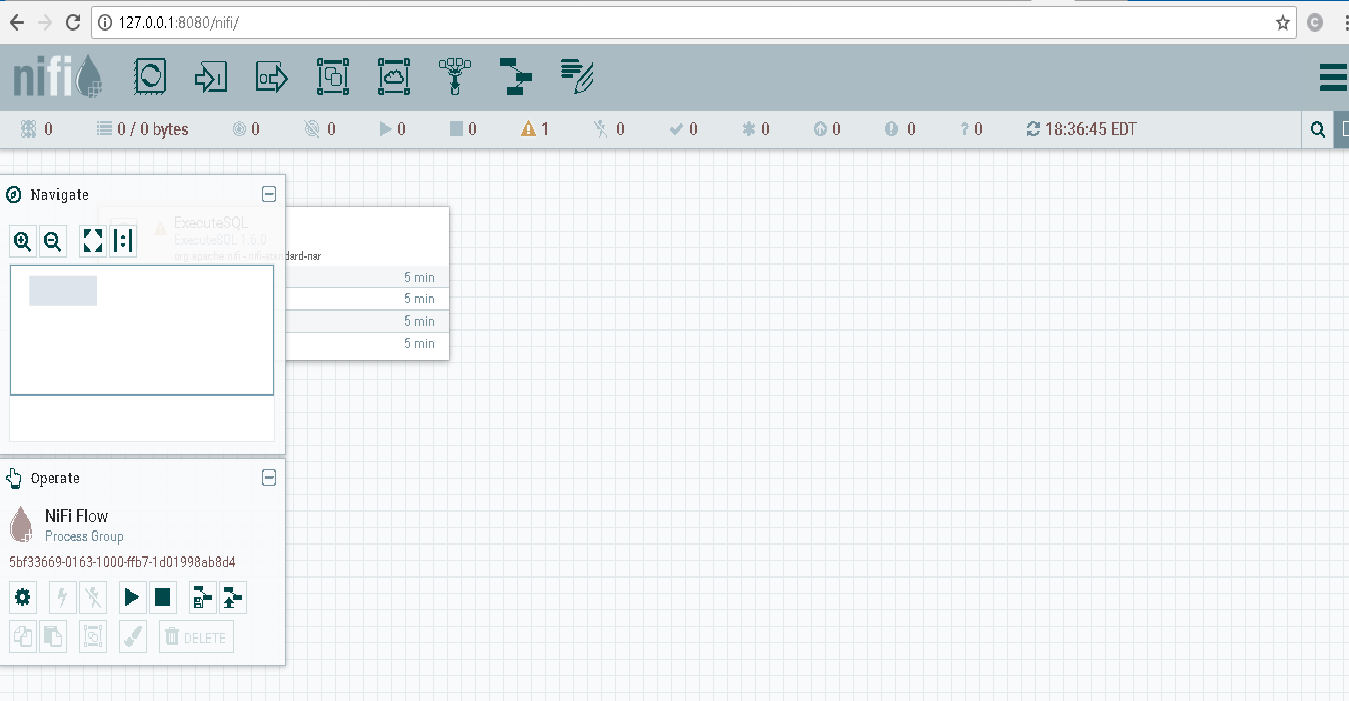
**Converting CSV Files to Apache Hive Tables With Apache ORC Files**

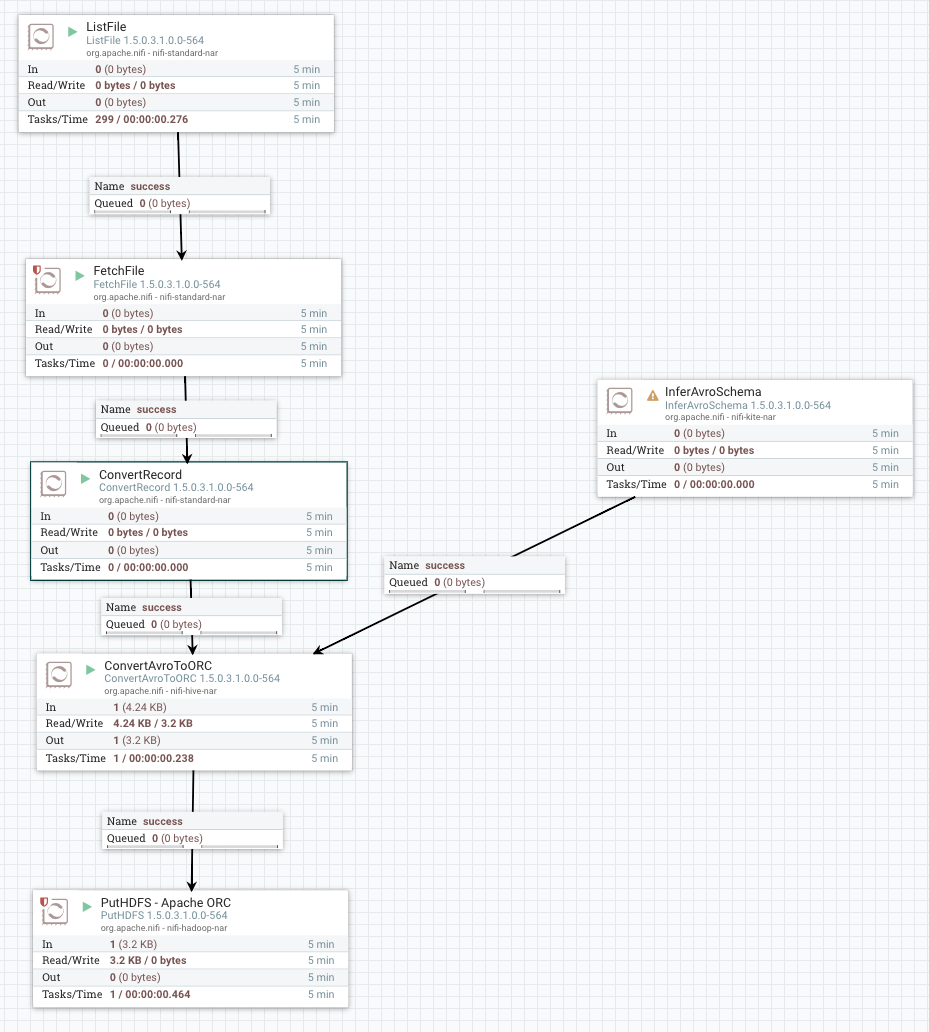
Once you have an Apache Avro file, it's easy to convert it to Apache ORC and then store it in HDFS. Here's how to do it.

**by** [**Tim Spann**](https://dzone.com/users/297029/bunkertor.html)

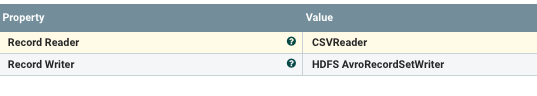
[Hortonworks Sandbox](https://dzone.com/go?i=285437&u=https%3A%2F%2Fhortonworks.com%2Fproducts%2Fsandbox%2F%3Futm_campaign%3Ddzonepre%2Fpostroll%26utm_medium%3Ddisplay%26apos%3B%26utm_source%3Ddzone%26utm_id%3D2216633) for HDP and HDF is your chance to get started on learning, developing, testing and trying out new features. Each [download](https://dzone.com/go?i=285437&u=https%3A%2F%2Fhortonworks.com%2Fproducts%2Fsandbox%2F%3Futm_campaign%3Ddzonepre%2Fpostroll%26utm_medium%3Ddisplay%26apos%3B%26utm_source%3Ddzone%26utm_id%3D2216633) comes preconfigured with interactive tutorials, sample data and developments from the Apache community.

I received some CSV files of data to load into Apache Hive. There are many ways to do this, but I wanted to see how easy it was to do in Apache NiFi with zero code.

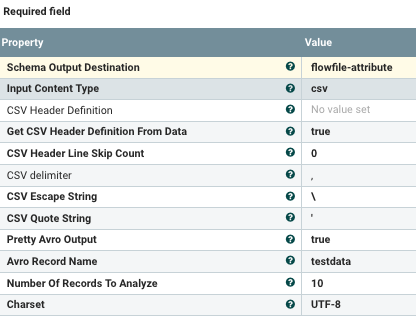




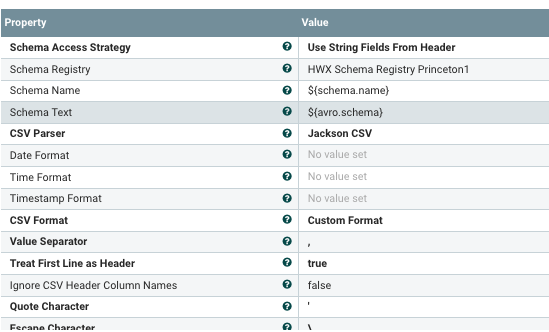
I read CSV files from a directory of files. Then, I can convert the **CSV** to **AVRO** directly with **ConvertRecord.**



I will need a schema, so I use the below settings for **InferAvroSchema**. If every file is different, you will need to do this every time.



CSV reader:



I use the Jackson CSV parser, which works very well. The first line of the CSV is a header. It can figure out the fields from the header.

Once I have an Apache Avro file, it's easy to convert to Apache ORC and then store in HDFS.

[Hortonworks Sandbox](https://dzone.com/go?i=285441&u=https%3A%2F%2Fhortonworks.com%2Fproducts%2Fsandbox%2F%3Futm_campaign%3Ddzonepre%2Fpostroll%26utm_medium%3Ddisplay%26apos%3B%26utm_source%3Ddzone%26utm_id%3D2216633) for HDP and HDF is your chance to get started on learning, developing, testing and trying out new features. Each [download](https://dzone.com/go?i=285441&u=https%3A%2F%2Fhortonworks.com%2Fproducts%2Fsandbox%2F%3Futm_campaign%3Ddzonepre%2Fpostroll%26utm_medium%3Ddisplay%26apos%3B%26utm_source%3Ddzone%26utm_id%3D2216633) comes preconfigured with interactive tutorials, sample data and developments from the Apache community.