HiveTestBench

- About Hive TestBench
- Compilation and Run of Hive TestBench
  - Create a local clone
  - Compile the benchmark
  - Prepare data for the scale
  - Run queries
    - TPC-DS
    - TPC-H

About Hive TestBench

The information available on [Hive TestBench](https://github.com/hortonworks/hive-testbench) web page.

Compilation and Run of Hive TestBench

Create a local clone

Clone the Hive TestBench

```
$ git clone https://github.com/hortonworks/hive-testbench.git
```

Compile the benchmark

Compile the benchmark using the scripts

```
$ ./tpcds-build.sh
$ ./tpch-build.sh
```

Choose the scale of data on which to run the tests Under 4 nodes, less than 100-250GB, 4-10 nodes should handle 1TB, Large clusters can scale up to higher values

Prepare data for the scale

As an example

```
$ ./tpcds-setup.sh 1000
$ ./tpch-setup.sh 1000
```

By default this creates tables in ORC format, you can choose a different format by setting

```
FORMAT=<SEQUENCEFILE, TEXTFILE, RCFILE, ORC, PARQUET, or AVRO>
For example; FORMAT=rcfile ./tpcds-setup.sh 30000
```

Run queries

You can use hive, beeline or any SQL tool
TPC-DS

cd sample-queries-tpcds
hive -i testbench.settings
hive> use tpcds_bin_partitioned_orc_1000;
hive> source query55.sql;

TPC-H

cd sample-queries-tpch
hive -i testbench.settings
hive> use tpch_flat_orc_1000;
hive> source tpch_query1.sql;