

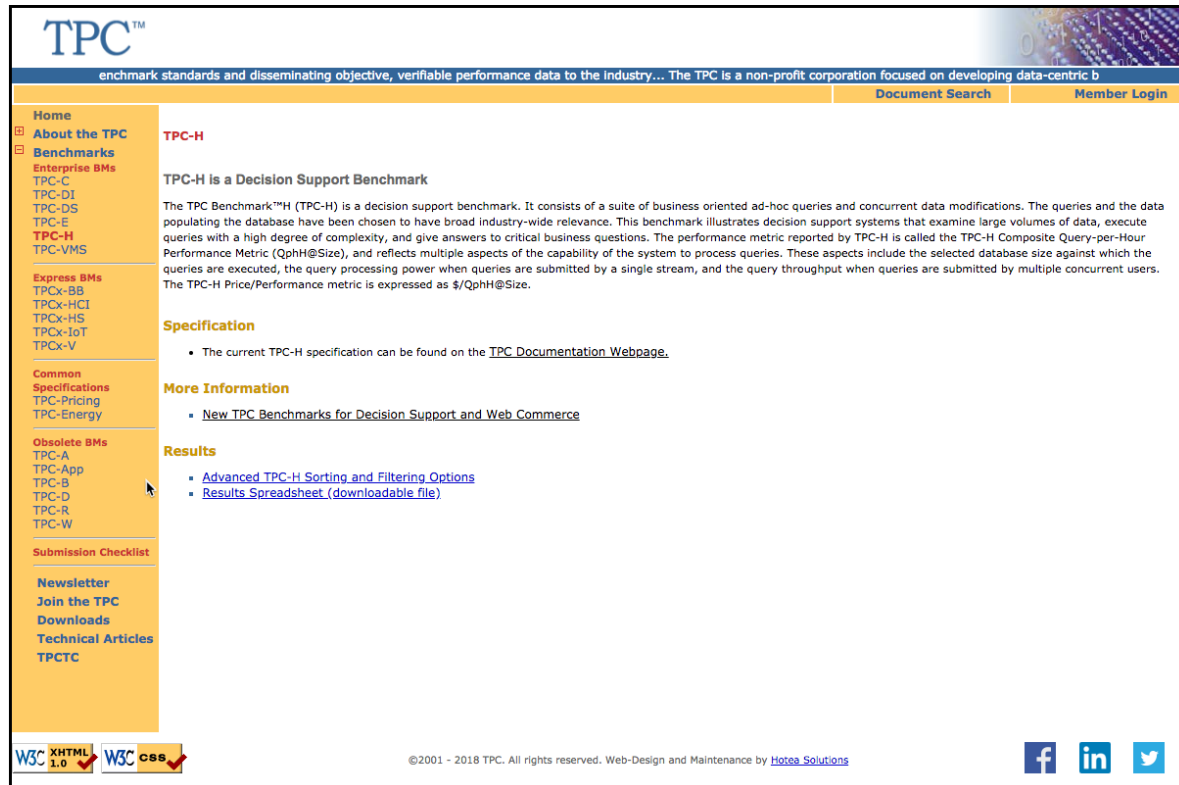
Metatron Test

use tpc-h

sp.hwang@sk.com

2018.05.30

Tmap condition



Data processing engine들을 benchmark 할때 많이 사용하는 데이타라고 함

우리도 이걸로 테스트를 해보는 것이 좋겠다.

electrum / tpch-dbgen

Watch 5 Star 59 Fork 38

Code Issues 0 Pull requests 1 Projects 0 Insights

Join GitHub today

GitHub is home to over 20 million developers working together to host and review code, manage projects, and build software together.

Sign up

TPC-H dbgen <http://www.tpc.org/tpch/>

3 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Find file Clone or download

File	Commit	Time
answers	Import 2.14.0	7 years ago
check_answers	Import 2.14.0	7 years ago
queries	Import 2.14.0	7 years ago
reference	Import 2.14.0	7 years ago
tests	Import 2.14.0	7 years ago
variants	Import 2.14.0	7 years ago
.gitignore	Add .gitignore	7 years ago

데이타는 정해진 고정이지 아니라 필요할때 필요한 만큼 만들어서 사용해야하기 때문에 data generator를 제공한다.

이걸로 데이타를 만들어서 테스트를 한다.

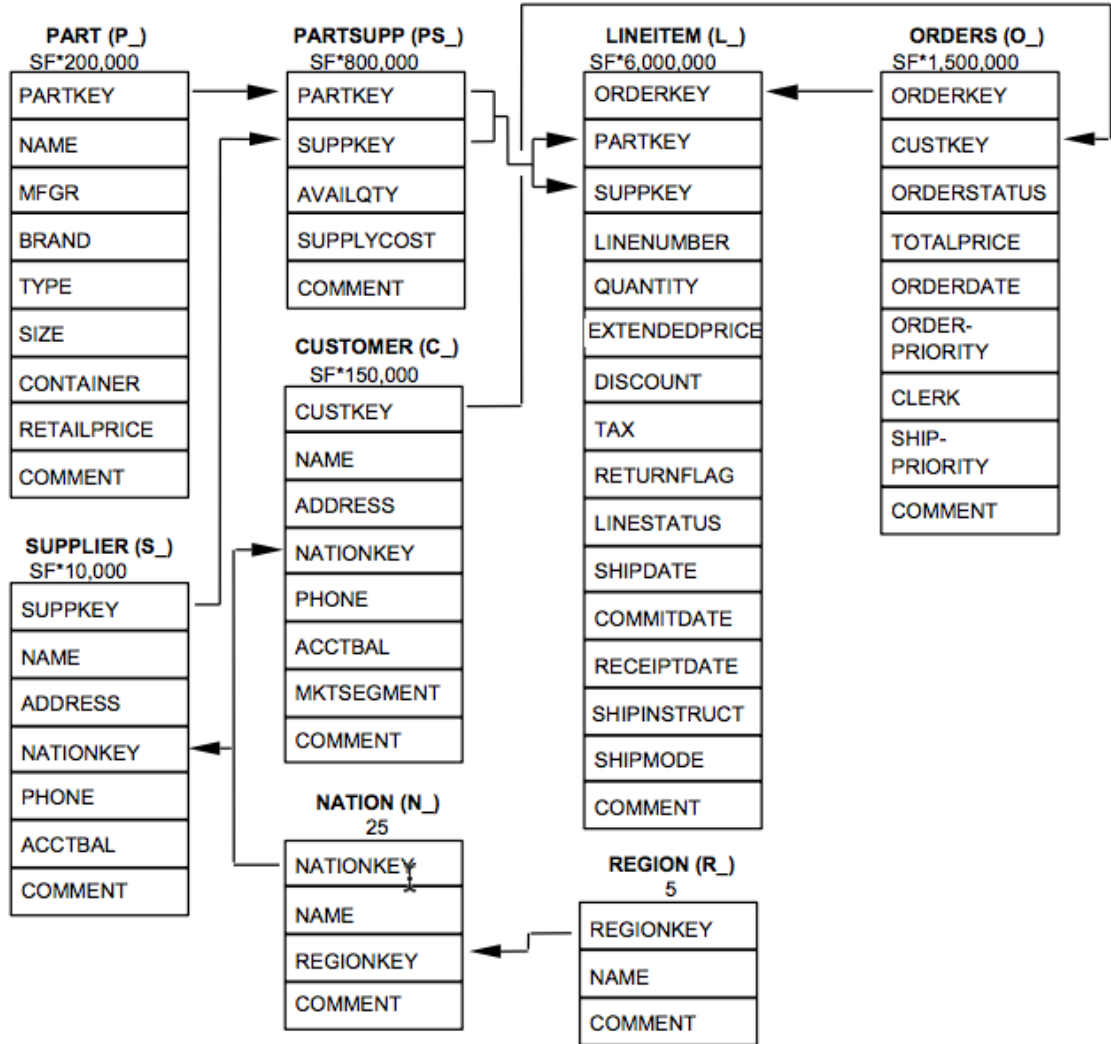
Makefile에서 DATABASE, MACHINE, WORKLOAD 정도 설정하고

make

하면

./dbgen 이 만들어진다.

Figure 2: The TPC-H Schema



옆에 보는 것 처럼 다양한 데이터가 연관되어 생성되는데 우선 가장 큰 lineitem만 가지고 해보자.

다른 녀석들은 차차 추가하자.

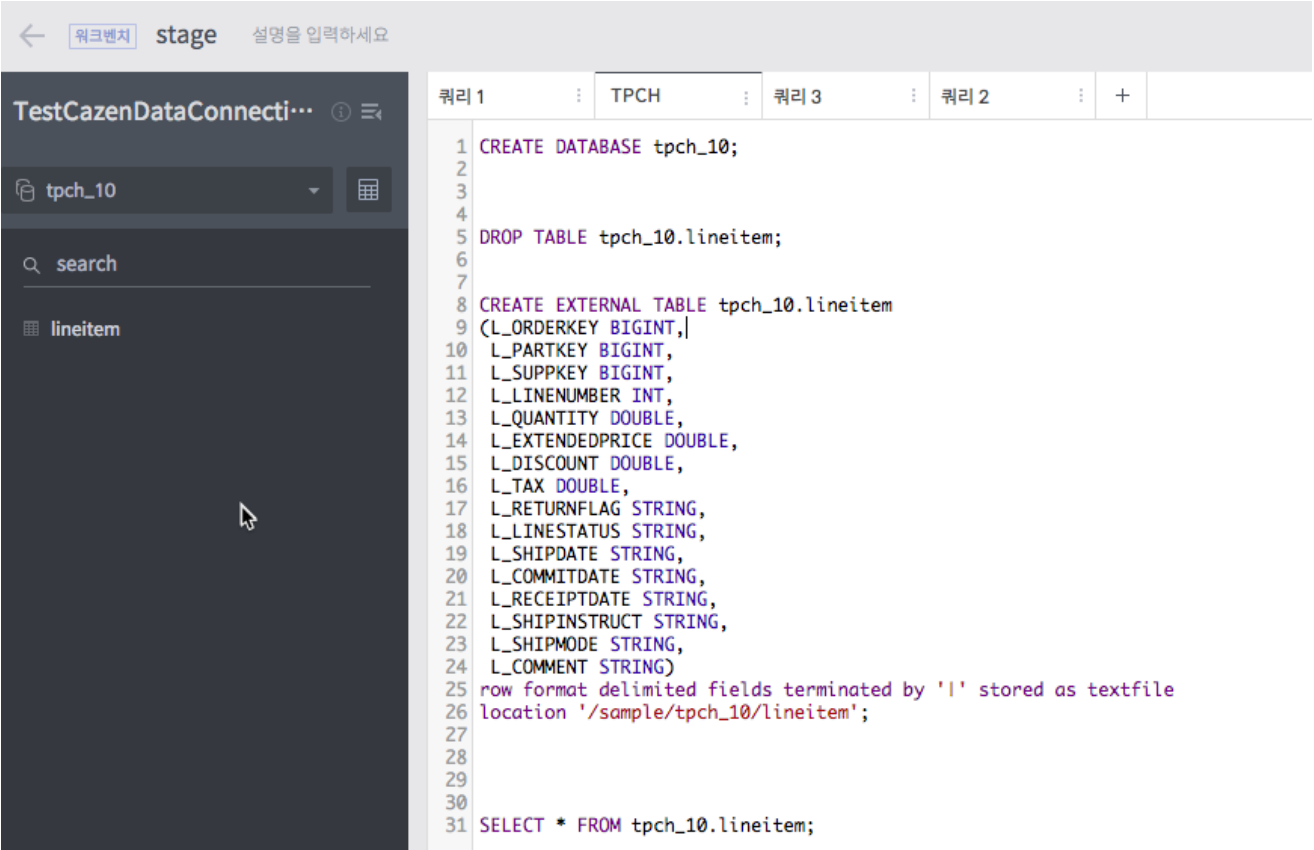
- s : scale factor : 데이터의 크기를 의미
- T : target data : 생성할 데이터 종류
- v : verbose

`./dbgen -s 10 -T L -v`

-> 10의 크기로 lineitem만 과정을 보면서 만들고 싶다.

그러면 lineitem.tbl 이녀석이 만들어진다.

Make hive table

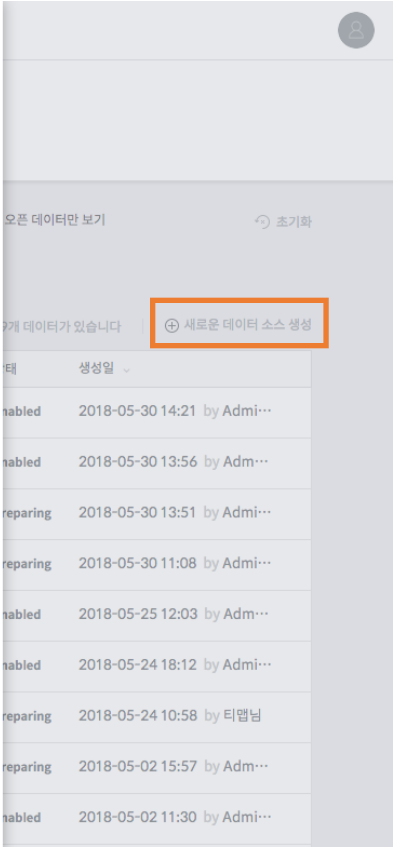
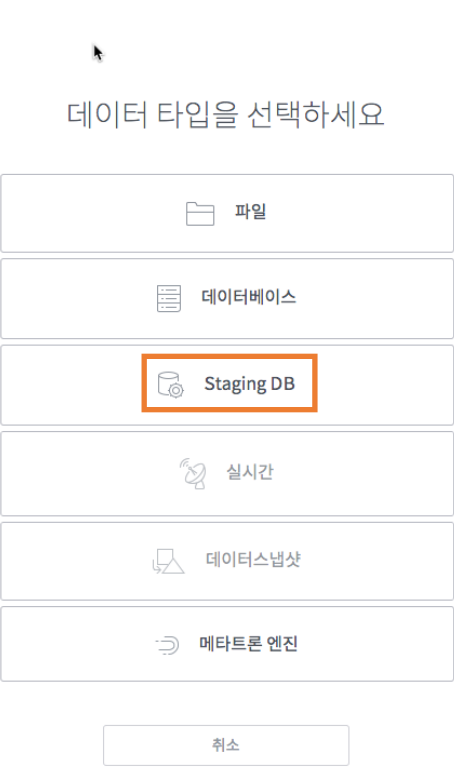
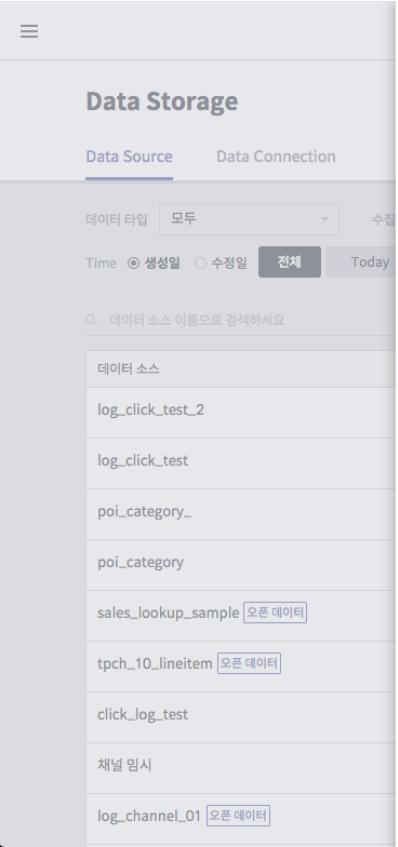


일단 디렉토리를 만들고

```
hadoop dfs -mkdir /sample/tpch_10
hadoop dfs -mkdir /sample/tpch_10/lineitem
hadoop dfs -put lineitem.tbl.1 /sample/tpch_10/lineitem
```

Hive table을 만들자.

```
CREATE DATABASE tpch_10;CREATE EXTERNAL TABLE tpch_10.lineitem
(L_ORDERKEY BIGINT,
L_PARTKEY BIGINT,
L_SUPPKEY BIGINT,
L_LINENUMBER INT,
L_QUANTITY DOUBLE,
L_EXTENDEDPRICE DOUBLE,
L_DISCOUNT DOUBLE,
L_TAX DOUBLE,
L_RETURNFLAG STRING,
L_LINESTATUS STRING,
L_SHIPDATE STRING,
L_COMMITDATE STRING,
L_RECEIPTDATE STRING,
L_SHIPINSTRUCT STRING,
L_SHIPMODE STRING,
L_COMMENT STRING)
row format delimited fields terminated by '|' stored as textfile
location '/sample/tpch_10/lineitem';
```



여기부터는 `datasource_druid_ingestion` 문서를
참고하여 진행하시면됩니다.