



# HW#2: load TPC-DS database and execute its queries

## ● HW#2 requires following tasks:

- download & install MonetDB
- download & compile TPC-DS benchmark tool
- generate SF=1 (1GB) TPC-DS database by using TPC-DS tool
- create TPC-DS tables in monetdb
- load generated 1GB TPC-DS database to MonetDB
- execute a series of test queries
- execute three TPC-DS queries
- verify test results

● In this slide, we consider that your OS is CentOS / RHEL / Ubuntu / Mac OSX

# Install MonetDB

- **Recommended OS: Mac OSX, Ubuntu, CentOS**
  - if you use Windows, then please consider installing Centos as virtual OS by using “VirtualBox” or “VMWare”
- **Go to <https://www.monetdb.org> and click “Download Now”**
- **Download and install the latest version of MonetDB depending on your OS**
- **If there is any issue during installation, please google it using a keyword such as “monetdb 설치”, “install monetDB” and so on**

# Install TPC-DS

- **Go to TPC-DS homepage and download TPC-DS tool**
  - [http://www.tpc.org/tpc\\_documents\\_current\\_versions/current\\_specifications.asp](http://www.tpc.org/tpc_documents_current_versions/current_specifications.asp)
  - click “Download TPC-DS\_Tools\_v2.10.1.zip” at Source Code tab of Active Benchmarks
- **Unzip the downloaded tool and install on your machine**
  - extract the file to specific directory and move to the directory
  - type “**make**” to compile TPC-DS tool
  - after the compilation is finished, type “**./dsdgen –help**” to verify the correct installation
  - if there is any issue, then please google it with an error message

# Generate SF=1 TPC-DS

## ● VERY EASY TO DO!

- go to the directory where TPC-DS tool is compiled
- run “`./dsdgen -sc 1`”
- after the program is finished, then please type “`ls -l *.dat`”
- if the following files are listed, then you successfully generates TPC-DS database

```
[[sun@SUN00 tools]$ ls -l *.dat
-rw-r--r--. 1 root root      1891 Mar 12 22:13 call_center.dat
-rw-r--r--. 1 root root    1631792 Mar 12 22:13 catalog_page.dat
-rw-r--r--. 1 root root    21378371 Mar 12 22:13 catalog_returns.dat
-rw-r--r--. 1 root root   295910384 Mar 12 22:13 catalog_sales.dat
-rw-r--r--. 1 root root    5502165 Mar 12 22:13 customer_address.dat
-rw-r--r--. 1 root root   13209372 Mar 12 22:13 customer.dat
-rw-r--r--. 1 root root   80660096 Mar 12 22:13 customer_demographics.dat
-rw-r--r--. 1 root root   10317438 Mar 12 22:13 date_dim.dat
-rw-r--r--. 1 root root         28 Mar 12 22:15 dbgen_version.dat
-rw-r--r--. 1 root root    151653 Mar 12 22:13 household_demographics.dat
-rw-r--r--. 1 root root         328 Mar 12 22:13 income_band.dat
-rw-r--r--. 1 root root  236420139 Mar 12 22:14 inventory.dat
-rw-r--r--. 1 root root    5051899 Mar 12 22:14 item.dat
-rw-r--r--. 1 root root    37233 Mar 12 22:14 promotion.dat
-rw-r--r--. 1 root root    1339 Mar 12 22:14 reason.dat
-rw-r--r--. 1 root root    1113 Mar 12 22:14 ship_mode.dat
-rw-r--r--. 1 root root    3155 Mar 12 22:14 store.dat
-rw-r--r--. 1 root root   32710005 Mar 12 22:15 store_returns.dat
-rw-r--r--. 1 root root  388445409 Mar 12 22:15 store_sales.dat
-rw-r--r--. 1 root root    5107780 Mar 12 22:15 time_dim.dat
-rw-r--r--. 1 root root         585 Mar 12 22:15 warehouse.dat
-rw-r--r--. 1 root root     5776 Mar 12 22:15 web_page.dat
-rw-r--r--. 1 root root    9806236 Mar 12 22:15 web_returns.dat
-rw-r--r--. 1 root root  146877674 Mar 12 22:15 web_sales.dat
-rw-r--r--. 1 root root     8771 Mar 12 22:15 web_site.dat
[[sun@SUN00 tools]$
```

# Run MonetDB and create TPC-DS tables

## ● Follow the guide of MonetDB tutorial

- <https://www.monetdb.org/Documentation/UserGuide/Tutorial>
- /path/to/yourfarm will be a storage for TPC-DS database
- monetdbd create → create database catalog to the directory
- monetdbd start → start database catalog
- after then, type “**monetdb create voc & monetdb release voc**”
- type “**mclient -u monetdb -d voc**” and type a password “**monetdb**”

```
shell> monetdbd create /path/to/mydbfarm
shell> monetdbd start /path/to/mydbfarm
shell> monetdb create voc
shell> monetdb release voc
shell> mclient -u monetdb -d voc
password:<monetdb>
Welcome to mclient, the MonetDB/SQL interactive terminal (Oct2014-SP1)
Database: MonetDB v11.19.7 (Oct2014-SP1), 'mapi:monetdb://localhost:50000/voc'
Type \q to quit, \? for a list of available commands
auto commit mode: on
sql>SELECT 'hello world';
+-----+
| single_value |
+-----+
| hello world  |
+-----+
1 tuple (0.530ms)

sql>\q
```

# Create TPC-DS tables on MonetDB

- Use data definition language (DDL) of SQL
  - e.g., create table ...
- For DDL for TPC-DS, please refer the following website
  - <https://github.com/cwida/tpcds-result-reproduction/blob/master/tools/tpcds.sql>
  - please type a DDL for each table (e.g., create table item (...);) in MonetDB shell
- If you successfully creates TPC-DS tables, then please verify a list of 24 TPC-DS tables
  - e.g., store\_sales, store\_returns, date\_dim, ...

# Load TPC-DS database to MonetDB

- a TPC-DS table generated by TPC-DS tool (e.g., dsdgen) is a kind of **CSV** file having “|” as its field (column) **delimiter**
  - e.g., TPCDS|Database|Is|Complex|But|Fun|To|Play|
- A bulk CSV insert SQL command (e.g., DML SQL language) supported by MonetDB allows tuples stored in TPC-DS table to store corresponding table in MonetDB
  - e.g., COPY INTO item FROM '/database/tpcds/SF1/item.dat' USING DELIMITERS '|', '\n', '' NULL AS '';
- By using the above DML sentence, please attempt to load all TPC-DS tables to your MonetDB TPC-DS database

# Test!

- Please check the number of tuples for the following tables
  - store\_sales (2,880,404 tuples), store\_returns (287,514 tuples)
- Please execute following three TPC-DS queries
  - query 3, query 7, query 52
  - for SQL query, please refer the following website:  
<https://github.com/Agirish/tpcds/>

# Submit a report to [dgist.info.courses@gmail.com](mailto:dgist.info.courses@gmail.com)

- p.1: describe your process for each step to do this HW
  - OS selected, trouble shooting, ...
- p.2: report the number of tuples per each TPC-DS table
  - e.g., `select count(*) from store_sales;`
  - a total of **24 lines**
- p.3: report the query results of five TPC-DS queries
  - select five TPC-DS queries randomly from 99 TPC-DS queries
  - describe the **query number** and the **number of resulting tuples** of the corresponding query
  - a total of **five lines**
- Due date: **4/11 (Thu) 24:00**