

# Installing SQLite and a sample TPC-H dataset on a Windows machine

Step 1: Download the cygwin installer. Go to [www.cygwin.com](http://www.cygwin.com) and download the setup program



Step 2: Install cygwin. Run setup.exe; per-screen instructions below (hit 'next' to advance)

- Welcome (nothing to do)
- Keep default 'Install from Internet'
- Select 'Just Me' if you don't own your computer (otherwise 'All Users' is fine). Also change the root directory to something you own (like a cygwin folder in 'My Documents') if C:\cygwin isn't accessible.
- Make sure the package directory is somewhere you have rights to
- Keep default 'Direct connection'
- Choose an http mirror from the list (<http://mirrors.xmission.com> works well)
- Click the 'View' button until it says 'Full', then make sure the following packages are selected from the list (use the 'Search' bar to narrow down the list if you want)
  - o gcc-core: C compiler
  - o make: The GNU version of the 'make' utility
  - o mintty: Terminal emulator with native Windows look and feel
  - o sed: The GNU sed stream editor
  - o sqlite3: An embeddable SQL database engine (3.x apps and docs)
  - o tar: A GNU file archiving program
  - o wget: Utility to retrieve files from the WWW via FTP and HTTP
- Ignore any post-install errors
- Decide if you want icons on your desktop (make sure to add them to the start menu!)

Step 3: Test shell. Go to Start -> All Programs -> Cygwin -> mintty (or, on Vista/Windows7 just hit the Windows and type 'mintty [enter]'). A linux-y looking shell should open up

Step 4: Prepare sqlite3. Run the following commands:

- mkdir c43-tpch
- cd c43-tpch
- wget <http://www.cs.utoronto.ca/~ryanjohn/teaching/csc43-s11/ddl-schema.sql>
- sqlite3 -init ddl-schema.sql tpch.db (a "sqlite>" prompt should appear)
- .schema (a list of tables and indexes should appear)
- select 'hello world'; (what happens?)

- To get help on sqlite-specific commands type `‘.help [enter]’`
- To exit type `‘.exit’` or `‘[Ctrl]-d’`.
- To interrupt a query that’s taking too long, type `‘[Ctrl]-c’` two or three times quickly.
- To close a mintty window, type `‘exit’` or `‘[Ctrl]-d’` or close it with the `‘x’` button like any windows program.

Step 5: Generating and importing the data. Open a mintty and run the following commands, which will download the data generator, but modify it to produce only 1% of its normal data volume. When you’ve finished, `tpch.db` should take up about 15MB. Note that the indented lines in the middle below are a single multi-line command.

- `cd c43-tpch`
- `wget http://www.tpc.org/tpch/spec/tpch_2_13_0.tar.gz`
- `tar xzf tpch_2_13_0.tar.gz`
- `sed --in-place= -e 's/00,$/,/' driver.c`
- `cat makefile.suite |`  
`sed -e 's/CC[ ]*=/CC = gcc/' |`  
`sed -e 's/DATABASE[ ]*=/DATABASE = SQLSERVER/' |`  
`sed -e 's/MACHINE[ ]*=/MACHINE = LINUX/' |`  
`sed -e 's/WORKLOAD[ ]*=/WORKLOAD = TPCH/' > Makefile`
- `make`
- `./dbgen -vf`
- `for f in *.tbl; do sed --in-place= -e 's/|$//' $f; done`
- `for f in *.tbl; do sqlite3 tpch.db ".import $f $(basename $f .tbl)"; done`

Step 6: Open a mintty, `‘cd c43-tpch’`, run `‘sqlite3 tpch.db’`, and start playing with the data! For your reference, here are the three queries we worked on in class together as well as the schema from the TPC-H specification (available at <http://www.tpc.org/tpch/spec/tpch2.13.0.pdf>):

```
select S.name, count(*) parts
from Supplier S, Nation N, PartSupp PS
where S.nationkey = N.nationkey and S.supkey = PS.supkey
      and N.name = 'CANADA'
group by S.supkey
order by parts desc
limit 5;
```

```
select C.name, O.orderDate, O.totalPrice
from Orders O, Customer C, Nation N
where O.orderDate between '1998-01-01' and '1998-12-31'
      and C.custkey = O.custkey and C.nationkey = N.nationkey
      and N.name = 'CANADA'
order by O.totalPrice DESC
limit 5;
```

```
select P.name, sum(L.quantity) Qty
from Part P, Orders O, LineItem L
where P.partkey = L.partkey and L.orderkey = O.orderkey
group by P.partkey
order by Qty desc
limit 5;
```

Note: In our hacked version of the database all SF\*<N> numbers below have N/100. For example, there are 2,000 parts in the database, not 200,000.

