

## 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://mirror.csclub.uwaterloo.ca> 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 cygrunsrv: NT/W2K service initiator
  - o libxml2: GNOME XML library
  - o libopenldap2\_3\_0: Lightweight Directory Access Protocol suite (runtime)
  - o postgresql: PostgreSQL Data Base Management System
  - o postgresql-client: Front-end programs for PostgreSQL
  - o postgresql-dev: Development files for PostgreSQL
  - o python: Python language interpreter
  - o sqlite3: An embeddable SQL database engine
  - 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 your Cygwin shell. Hit the Windows key and type **mintty** [enter]. A linux-style command shell window should open. Type [ctrl]+d or **exit** to close it.

Step 4: Set up postgresql

- open mintty as administrator
- export PATH="\$PATH:/usr/sbin"
- execute /usr/bin/cygservice-config

- Follow the “easy” instructions “for the impatient” directions at the bottom of /usr/share/doc/Cygwin/postgresql-8.2.11.README.
- NOTE: After you invoke `pg_ctl start ...`, the postgres server will continue running (and occupying the terminal) until you shut it down with `pg_ctl stop ...` or kill it. Just open another terminal window to connect to the server and continue the loading process.
- From the psql prompt:
  - o create database tpch;
  - o create database imdb;
- NOTE: if you created a database and need to delete it, use syntax like:
  - o echo "drop database tpch;" | psql
  - o (you can't drop a database you're currently connected to)

#### Step 5: Prepare datasets

- Download the imdb and tpch datasets from the dataset page of the course website, and untar them (the -a flag is quite useful for auto-identifying the compression type of an archive)
- Execute the following commands in your shell to create useful helper functions:

```
function table-names {
  cat $1 | tr [:upper:] [:lower:] | grep 'create table' | sed 's:create table[ ]*\([^ ]*\).*\;$\;1;'
}
function load-sqlite {
  sqlite3 $1 ".import $2.tbl $2"
}
function load-pgsql {
  echo "copy $2 from '$(pwd)/$2.tbl' delimiter '|' CSV;" | psql $1
}
}
```

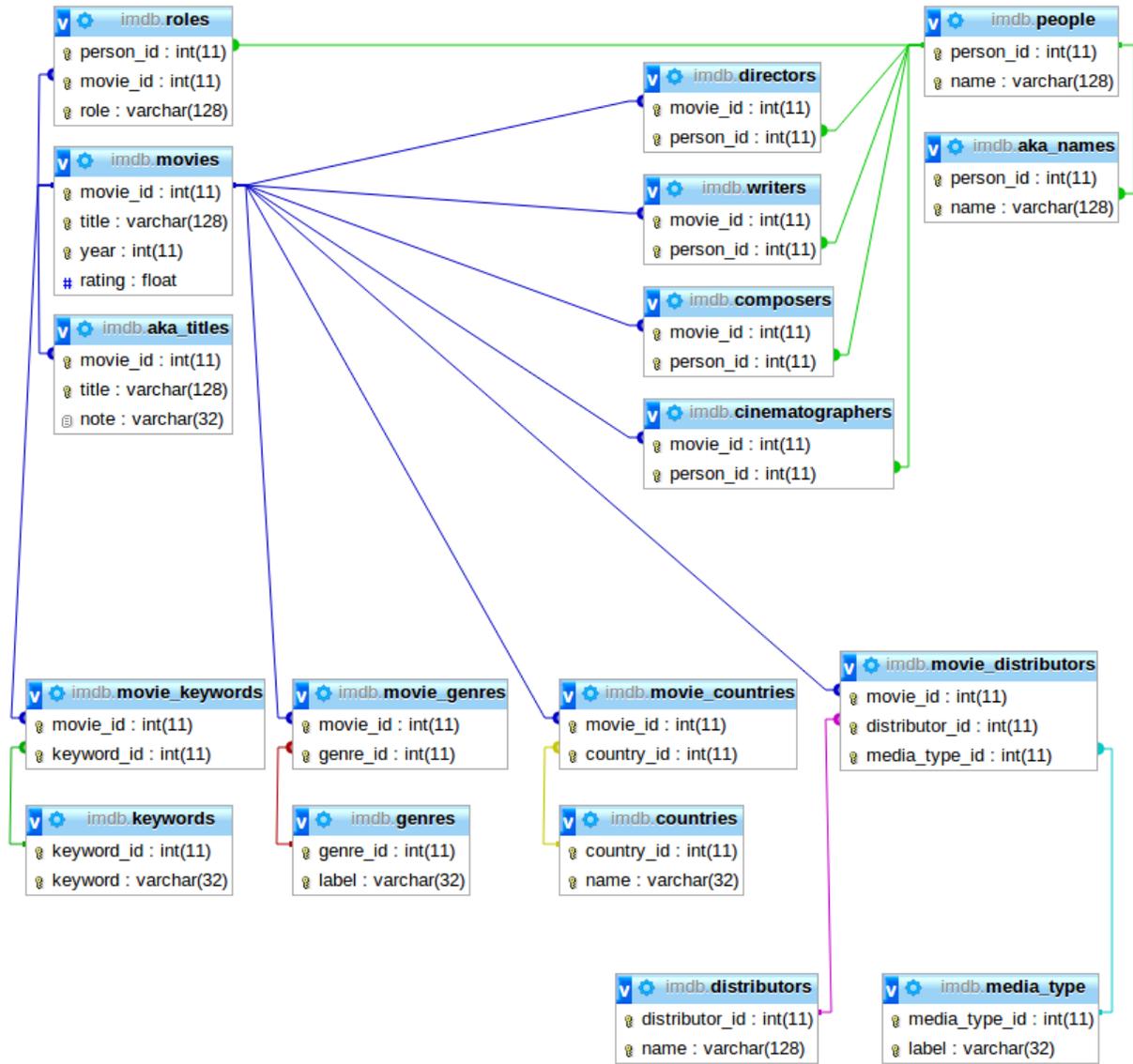
#### Step 6: Load the TPC-H dataset

- cd tpch
- cat tpch.ddl | sqlite3 tpch.db
- for f in \$(table-names tpch.ddl); do load-sqlite tpch.db \$f; done
- cat tpch.ddl | psql tpch
- for f in \$(table-names tpch.ddl); do load-pgsql tpch \$f; done
- cd -

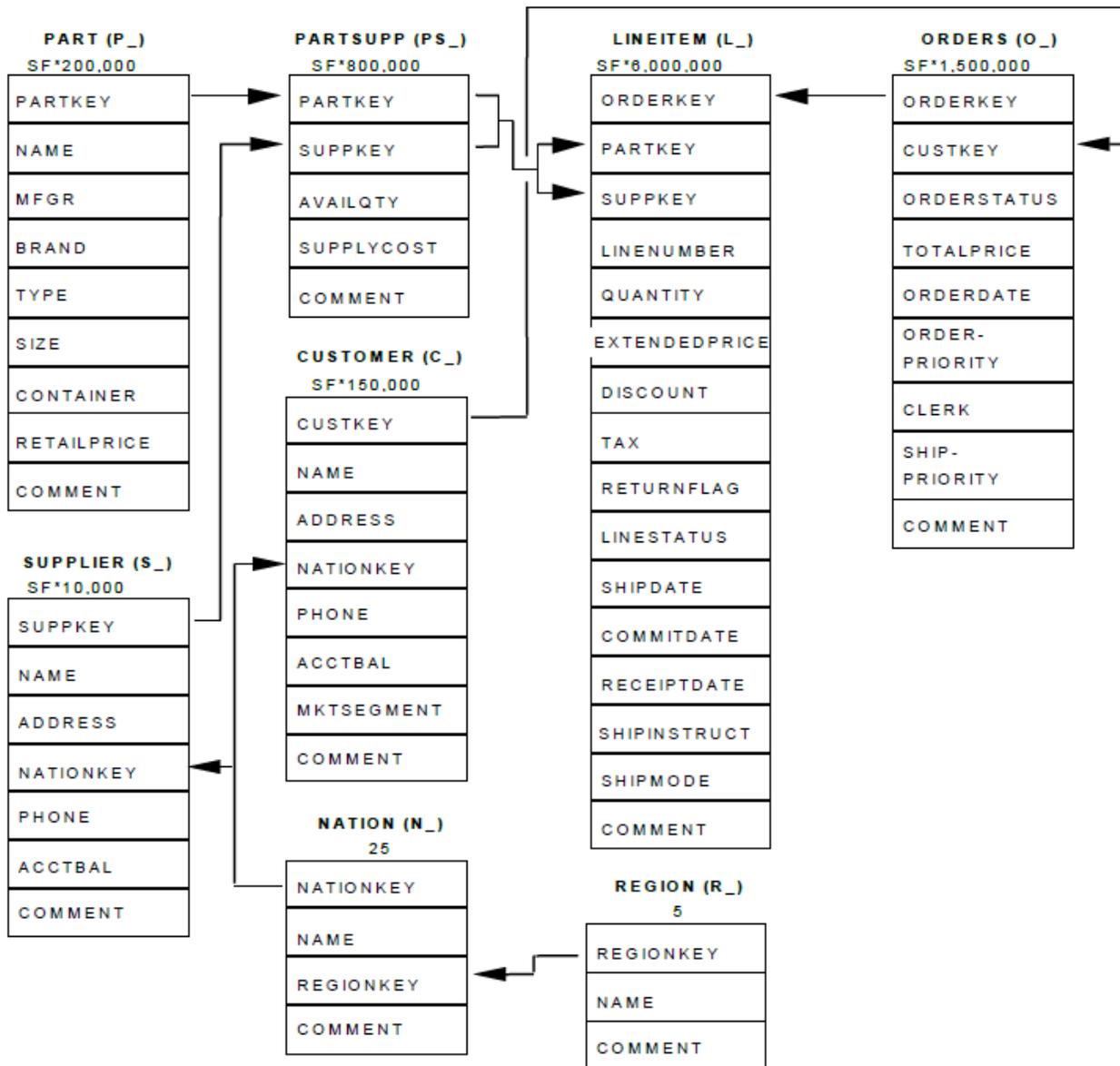
#### Step 7: Follow the example from Step 6 to load the IMDB dataset

For your convenience, the IMDB and TPC-H database designs are depicted below:

IMDB:



TPCH:



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.