Readme for reproducibility submission of paper ID 294

(Handling Environments in a Nested Relational Algebra with Combinators and an Implementation in a Verified Query Compiler)

JOSHUA S. AUERBACH, IBM Research
MARTIN HIRZEL, IBM Research
LOUIS MANDEL, IBM Research
AVRAHAM SHINNAR, IBM Research
JÉRÔME SIMÉON, IBM Research

The reproducibility artifact accompanying the SIGMOD 2017 paper entitled *Handling Environments in a Nested Relational Algebra with Combinators and an Implementation in a Verified Query Compiler* is distributed on github as a branch of the main Q*cert code-base on which the paper is based.

In order to automate the artifact, it comes with a Docker script which: (i) sets up a Linux virtual machine, (ii) downloads the software dependencies, (iii) compile Q*cert, (iv) runs the experiments and (v) re-compile the paper for comparison with the one in the SIGMOD 2017 proceedings.

The rest of this README provides the steps through the complete process.

Getting the code

The reproducibility artifact can either be obtained by cloning the Q*cert repository using git:

```
git clone https://github.com/querycert/qcert.git
cd qcert
git checkout sigmod-repro
```

or by downloading it as a zip file at:

```
https://github.com/querycert/qcert/archive/sigmod-repro.zip
```

and unzipping it.

Getting the TPC-H Benchmark

The licensing for the TPC-H benchmarks that we use for evaluation does not allow redistribution as part of this artifact, so it must be obtained separately.

Go to the TPC-H download page:

```
http://www.tpc.org/tpc_documents_current_versions/current_specifications.asp
```

and click `Download TPC-H_Tools_v2.17.2.zip` under Source Code/TPC-H.

Fill out the form and agree to the license. You will then receive an email with a link to a individualized page hosting the TPC-H benchmarks. Download the benchmarks (you will only be able to download them once; if you need to do this again, you must fill out the license form again).

Please accept the suggested filename (which will look something like `...-tpc-h-tool.zip`, where the leading part is some individualized string), and save it to the `qcert` directory (the top level directory of the code) with the following name: `tpc-h-tool.zip` (all lowercases).

Getting Docker

In order to automatically fetch the dependencies and build the code, you will need to install Docker, which can be downloaded for your platform at:

```
https://www.docker.com/
```

Once docker is installed you will need to launch it (as any other application for your system, which will start the Docker daemon). It will need to be running for the remainder of the process.

Reproduce the paper

You can now run the full `reproduce-paper.sh` script from the `qcert` directory:

```
bash-3.2$ cd qcert
bash-3.2$ ./reproduce-paper.sh
```

This will This will create a new docker image (i.e., a virtual machine) called `qcert:repro` with all required dependencies, and use it to build the code, run the experiments, and rebuild the submitted paper.

It will then copy the rebuilt paper out of the image and put back on your local file system as:

```
qcert/qcert-sigmod-2017-paper-reproduction.pdf
```

Note that the version of the paper published in sigmod 2017 is provided for convenience in the Q*cert source as:

```
qcert/qcert-sigmod-2017-paper-original.pdf
```

Explore

After building the docker image with the `qcert/reproduce-paper.sh` script, you can get a shall to that image by running:

```
docker run -i -t qcert:repro bash
```

This can be used to poke around and run the compiler against other examples, including new examples.

Note that `nano` is provided in the image for basic editing needs and `sudo apt-get` can be used to install other editors as desired.