Software Installation Frequently Asked Questions (FAQ)

- **Is it possible to install the VM on a tablet computer, e.g., iPad?**

Although installation of VMs on tablets might be possible through specialized applications, we recommend using a computer with the following capabilities for the hands-on exercises.

Hardware Requirements: Quad Core Processor (VT-x or AMD-V support recommended), 64-bit; 8 GB RAM; 20 GB disk free. Most computers with 8 GB RAM purchased in the last 3 years will meet the minimum requirements.You will need a high speed internet connection because you will be downloading files up to 4 Gb in size.

- **How do I find my hardware information?**

On Windows: Open System by clicking the Start button, right-clicking Computer, and then clicking Properties;

On Mac: Open Overview by clicking on the Apple menu and clicking "About This Mac."

- **If I have problems with Step 3 (Querying Relational Data), where should I look for help troubleshooting with Hands On/Pyspark?**

The instructions for Step 3 assume you have saved and unzipped the big-data-3.zip file in the Downloads directory. If you have not done this, then in order to fix the problem, you can examine the setup.sh and execute again some part of the script manually (**changing** the provided directory to your current work directory).

You have to execute the commands from the **/.../big-data-3** directory

1

2

3

set permissions so postgres can read csv files

chmod 755 /home/cloudera/Downloads/big-data-3

You don't need to create users or structure of tables (according to your screenshots, they are correctly created)

Check that all .CSV files have been unzipped and have the correct permissions. If not sure, run the code again (you need to be in the **/.../big-data-3** directory)

1

2

3

4

5

6

decompress csv files

gzip -d *.csv.gz

set permissions so postgres can read csv files

chmod 644 *.csv

Then you need to edit the **init-postgres.sql** file, that is located in the **/.../big-data-3/setup** directory. You can use **gedit** or any other file editor.

At the end of the file, you will see that it is trying to access the files in the Downloads directory. **Change** the path to your current path and save the file.

1

2

3

COPY buyclicks FROM '/home/cloudera/Downloads/big-data-3/buy-clicks.csv' DELIMITER ',' CSV HEADER;

COPY gameclicks FROM '/home/cloudera/Downloads/big-data-3/game-clicks.csv' DELIMITER ',' CSV HEADER;

COPY adclicks FROM '/home/cloudera/Downloads/big-data-3/ad-clicks.csv' DELIMITER ',' CSV HEADER;

Once the file is saved, go back one directory level to **/.../big-data-3** again and execute the following command:

1

psql -f setup/init-postgres.sql

Don't worry if you see some errors when creating the tables, it is normal, as tables were already created so it doesn't allow to create them again. The important thing here is that you can see that the data is copied successfully.

1

2

3

4

5

6

7

8

9

10

[cloudera@quickstart big-data-3]$ psql -f setup/init-postgres.sql

psql:setup/init-postgres.sql:9: ERROR: relation "buyclicks" already exists

psql:setup/init-postgres.sql:19: ERROR: relation "gameclicks" already exists

psql:setup/init-postgres.sql:29: ERROR: relation "adclicks" already exists

DELETE 0

DELETE 0

DELETE 0

COPY 2947

COPY 755806

COPY 16323

Run again the psql and when you do the select to any of the tables, you should see the data has been loaded.

Pyspark (No Jupyter Found):

At the end of Step 5 in the Reading Instructions for Downloading Hands On Datasets (Week 1), Anaconda asks to modify $HOME/.bashrc. If you enter yes, you should not have to manually add the export PATH= line to your .bashrc.

If that does not work, try the following:

Add the line in the file $HOME/.bashrc: export PATH=$PATH:/home/cloudera/anaconda3/bin

- Error: ***psql: FATAL: Ident authentication failed for user "cloudera"***

Run the following commands in the terminal window:

1

2

3

4

5

echo "host   cloudera   cloudera   127.0.0.1/32   trust" > /tmp/pg

sudo cat /var/lib/pgsql/data/pg_hba.conf >> /tmp/pg

sudo cp /var/lib/pgsql/data/pg_hba.conf /var/lib/pgsql/data/pg_hba.conf.DIST

sudo mv /tmp/pg /var/lib/pgsql/data/pg_hba.conf

sudo service postgresql start

Students have also found the following command to help:

1

local all all trust

You may also find additional assistance in the Discussion Forums for the corresponding week.