

Troubleshooting the VM

As you work with the VM, you might from time to time come up against some odd issues. Please consult this document as needed to help you troubleshoot and resolve the issues.

If you have worked through all the suggestions here and still have trouble, please reach out to your fellow students or the instructors through the forums for this course. Help your fellow students as you can, but your instructors will do their best to help you as soon as possible.

Please try to resolve the issues on your own first. We understand how frustrating it can be, but you'll learn more if you try it on your own!

Note: Many issues can be solved by restarting the VM, so this should be the first thing you try. Restart the machine by clicking System > Shut Down from the menu bar, then click Restart. (Do not just pause or suspend the VM or quit the software running it.) Many times the issue is a service (such as Hive, Impala, Hue, HDFS, or one of the underlying services those rely on) going down, and restarting the machine will restart all these services.

VM CPU and RAM Requirements

The VM for this class is designed to use one processor core (CPU core) and 4GB RAM. Reducing the amount of RAM to below 4GB is not recommended and is likely to cause failures when running some queries on large tables. Increasing the amount of RAM to some amount greater than 4GB is unnecessary, but it will not cause problems so long as your computer has sufficient RAM to allow it. You should always leave at least about 4GB available for the operating system outside the VM to use. For example, if your computer has 8GB total ram, you should never configure the VM to use more than about 4GB. If your computer has 16GB total RAM, you should never configure the VM to use more than about 12GB.

However, you should *not* increase the number of processor cores (CPU cores) used by the VM. If you do increase the number of processor cores used by the VM, then it is absolutely necessary to also increase the amount of RAM. For example, with two processor cores, you should use at least 6GB RAM. Increasing the number of processor cores without also increasing the amount of RAM is likely to cause failures.

VM Is Slow

If the VM is running slowly, it might be that you are using too many resources for the memory available to the VM. If you're using Hue, first try closing the browser and reopening it. This sometimes clears out the resources.

If that doesn't help, then in the VM, go to System> About this Computer> Resources to see how much CPU and memory (RAM) is being used.

If your RAM usage is high, close all applications and browser windows or tabs except the one you're using. Avoid having Hue open in multiple browser windows or tabs, because this can use a lot of RAM.

Services Not Available

Occasionally you might find that a service or process on the VM has failed and needs restarting. The simplest way to do this is to restart the VM. See the note before the "VM CPU and RAM Requirements" section, above.

Errors in Hue

Clicking around in certain parts of Hue that are not part of the exercises might result in error messages. The Job Browser is one such example. Clicking it might show a red popup layer in the browser with an error message similar to this:

```
HTTPConnectionPool(host='localhost', port=11000): Max retries exceeded
```

```
with url: /oozie/v1/jobs?len=100&doAs=training&filter=user%3Dtraining%3Bstartcreatedtime%3D-7d&user.name=hue&offset=1&timezone=America%2FLos_Angeles&jobtype=wf (Caused by  
NewConnectionError('<requests.packages.urllib3.connection.HTTPConnection object at 0x7f4f1c53f090>': Failed to establish a  
new connection: [Errno 113] No route to host',))
```

This occurs because the Job Browser depends on a component called Oozie, which we do not include in the VM. Oozie is not used in any exercises for this course; installing it would make the VM larger and require more memory to run, which would reduce performance.

Note that the job browser is not the only place where errors like this might occur. Other areas of Hue might yield errors related to other components that have not been installed. In general, we have tested that the exercises related to Hue work without error. If you deviate from the exercise instructions, then you might encounter errors such as the one described above.

Impala Database Not Loading in Hue

Some people report that Hue does not show the databases in Impala, but they do show with Hive. Hue can be finicky sometimes. Try clicking the refresh button in the left pane when you have Impala selected. (That's the one that looks like two curved arrows.) "Clear cache" is sufficient. If you try that a couple of times and it doesn't work, you can try refreshing the browser window, or logging out and back in again.

If none of that helps, then maybe Impala is down. Open a terminal window (the computer monitor icon in the menubar or on the desktop) and use this command:

```
sudo service --status-all
```

There should be three with the word "Impala" in it; if any of them are not [OK], use this command, replacing [service]—including the square brackets—with either catalog, state-store, or server (depending on which came back not OK).

```
sudo service impala-[service] restart
```

For example, the catalog command would be

```
sudo service impala-catalog restart
```

Difficulty Connecting to Beeline

Your first step when a command doesn't work as expected should always be to check *carefully* for typos! Be sure the command you are using to start Beeline is

```
beeline -u jdbc:hive2://localhost:10000
```

It's very easy to type jdbc instead of jdbc, for example, and easy to overlook that typo.

It's also easy to use the wrong number of 0s at the end. (There should be four 0s.)

Difficulty Connecting to Impala Shell

Your first step when a command doesn't work as expected should always be to check *carefully* for typos! Be sure the command you are using to start Impala Shell is

impala-shell

It's very easy to forget to use a dash and instead use a space, for example.

Difficulty Connecting for S3 or Other Internet Services

Some commands (such as any commands that interact with S3, the cloud service we're using for these courses) require that the VM itself be connected to the internet. In the upper right corner, find one of these icons to determine the connection state and what you should do:



The VM is connected; if there is a problem, check your computer's connection rather than the VM's connection.



(Animated) The VM is trying to connect; give it a moment until it resolves to one of the other two icons.



The VM is disconnected; click the icon and choose Auto Ethernet to reconnect. If the problem persists even after you reconnect the network, then restart the VM.