IBM Journey to Cloud and AI

Analytics Modernization Workshop

Featuring: Cloud Pak for Data (CPD)

Getting Started Guide

Burt Vialpando, Kent Rubin - IBM

October 26, 2020

Experienced User Workshop Setup Checklist

This checklist is for those who have started and set up the workshop clusters before and only need a one-page reminder of what to do. The rest of this guide provides more detail on each of these steps for those unfamiliar with the setup.

☐ Requesting a cluster

Use either the Skytap or TEC HCCX options.

1. Skytap (DTE) Option

    a. To request a cluster for single use (practice, demo) go to: https://dte2.us1a.cirrus.ibm.com/collection/ibm-journey-to-cloud-and-a-i-analytics-modernization-workshop-v-9-0-featuring-cloud-pak-for-data-1qc4iaz8n

    b. To request multiple clusters (up to 8), go to: https://www.ibm.com/demos/workshops/reserve/

2. TEC HCCX Option

    a. To request multiple clusters (up to 15), go to: https://natecs.raleigh.ibm.com/tec-reservations/

    Note: Please see the 'How to Host a Workshop v3.0.1 for more information: https://ibm.box.com/v/Workshop-How-To-Host-v9

☐ Start the virtual machines

For Skytap, use the single power on button for all the machine to start at once.
For those accessing via HCCX ESX servers, this usually has already been done for you.

☐ Wait 15 minutes

Give the OS a chance to render the GUI and for the OpenShift, Docker and NFS services to start.
For those accessing via HCCX or ICC-VE servers, you can skip this step.

☐ Become root in the Unified Desktop Terminal (or PuTTY)

In the terminal ➪ `root` (do this to become root – all terminal commands in this document run as root)

☐ Test out the OpenShift

```
# oc get pods
```

```
# oc login -u ocadmin -p cpdaccess
```
(if you get an error from the above)

☐ Cluster environment health check

```
# checkall
```

☐ Pods check

```
# checkpods          (pods not running)
# checkpodsw         (pods not running – wide list)
# checkpodscount     (count of pods not running – "1" means all are running)
# checkpodswatch     (pods not running every 2 seconds – use another terminal for doing other work)
```

☐ CPD web client check

Check the status of Db2 Warehouse, MongoDB and Data Virtualization, then log out of the web client.

☐ Prime the "Organize" services (optional)

☐ Troubleshoot (when necessary)

Lab 0 Workshop Setup instructions

This is the getting started guide for people doing self-enablement or for instructors and proctors of the CPD (Analytics Modernization) workshop to start the virtual machines and run the labs.

0.1 Important notes:

For the best experience in doing the labs:

1. **No tablets!**

   Attempting the workshop labs on an iPad or similar tablet device will not work. Use a real laptop or even desktop, the bigger the screen the better. Note: a tablet CAN be used as a secondary device to use as a digital workbook. We will print out a workbook for you for the workshop, but if you want, we can give you the workbook in PDF form and you can scroll through it on a tablet digitally. You will still need a nice sized laptop to do the actual lab work on.

2. **Use Chrome or Firefox**

   These are the two browsers you should use, with Chrome being preferred. IE and other browsers do not work well with the Cloud Pak for Data web client – they can even freeze up in Skytap sessions, which is what we use to deliver this workshop. If you don't have Chrome or Firefox on your laptop, please download and install it ahead of time.

3. **Full screen is important**

   Make sure you always have your session in full screen mode, with the browsers maximized to fit their respective screens.

4. **Learn to use zoom in and zoom out**

   Use [Ctrl][+] and [Ctrl][-] or [Ctrl][Mouse-Scroll-Wheel] can adjust the zoom of the browser where the CPD web client runs. Knowing this can be useful in some of the labs.

5. **Use a mouse**

   Using a mouse is strongly recommended; using a keyboard can be difficult with some of the lab exercises. If you don't have a mouse, then it will be assumed you are very familiar with all of the scrolling capabilities of your particular keyboard.

6. **Skytap environments only: keep the CPD cluster active**

   The default suspend time for environments provisioned from Skytap is 300 minutes. If you are doing a group workshop set up by an IBM representative, the suspend time should have been set to 300 minutes as well.

   In either case, if you do not in some way touch or use the CPD cluster within that timeframe, the VMs in the cluster will suspend. *A suspended cluster cannot be resumed and will require a reboot.*

   Running a script from a terminal window that loops continuously will not help! The environment cluster must be actively touched in some way by either clicking in the Cloud Pak for Data web client or opening a terminal and running a command manually.

0.2 Starting the virtual machines

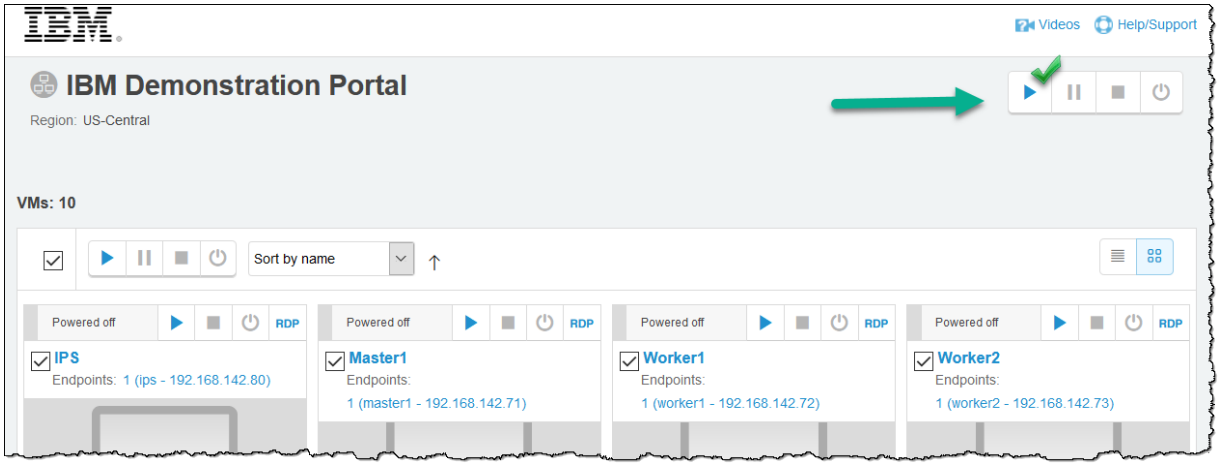
To begin the workshop setup, you must first boot up the VMs in the workshop cluster.

Note: For those using HCCX workshops, you can skip this section and go on to the next.

	Please allow 30+ minutes to bring up the cluster. Much is happening during this time, including OS boot, the time synchronization setup, CRI-O and OpenShift confirmation via cron job at Master1, plus 220+ pods, jobs, etc. including MongoDB, Db2, Data Virtualization, etc.

0.2.1 Starting the VMs in Skytap environments

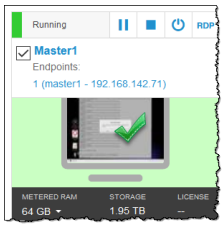
7. Start all the machines with the single “start all” button at the top right, as shown below.



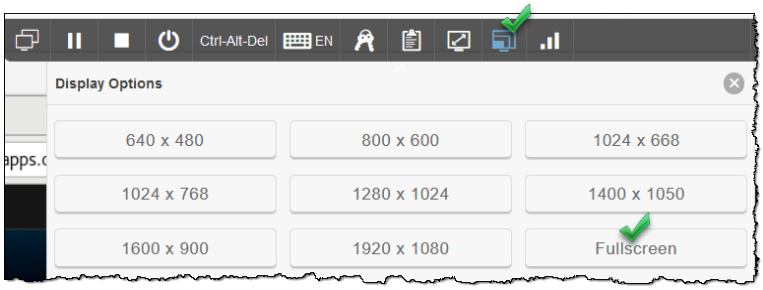
The VMs will start up in an efficient programmed sequence; internet access is part of the networking configuration.

Note: do not be alarmed if you see a message *“The operation is rate limited and will be completed as soon as possible.”* This just means the VMs are booting in a queue and is normal.

8. Wait at least 15 minutes after the management `Master1` comes up before accessing the GUI desktop (from which you will drive the workshop). Click on the desktop icon to launch it.



9. Enter `Fullscreen` mode to maximize the desktop real estate.



0.2.2 Starting VMs in other environments (e.g. HCCX, ICC-VE servers)

If you are performing CPD workshops on HCCX or ICC-VE servers, it is likely that the clusters will be started for you. In addition, you will be given URLs to access the client desktops for each cluster, which is driven by a remote client application like Guacamole or RDP. If this is the case, you will not need to do any of the following and can skip to the next section.

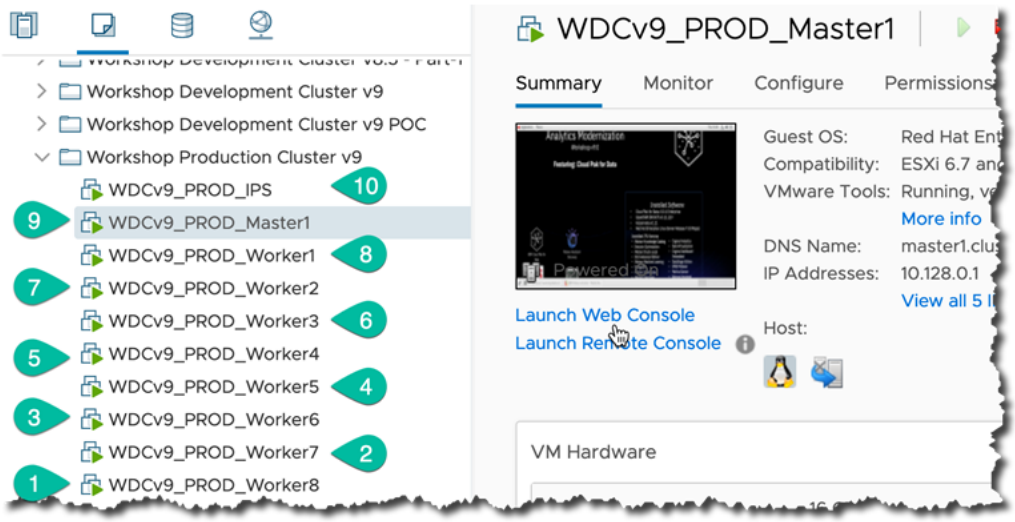
However, if you are asked to start and access your own clusters on the HCCX servers, then this is the process:

10. Start all nodes independently, one after the other as quickly as you can, in this order:

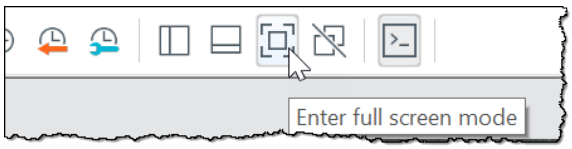
    `Worker8` ⇨ `Worker7` ⇨ `Worker6` ⇨ `Worker5` ⇨ `Worker4` ⇨ `Worker3` ⇨ `Worker2` ⇨ `Worker1` ⇨ `Master1` ⇨ `IPS`

    Start them by right clicking on the node and then `Power` ⇨ `Power On`

    Wait at least 15 minutes after the management `Master1` comes up before accessing it for the GUI desktop (from which you will drive the workshop itself.) Access the GUI by clicking on the `Launch` link for `Master1`. Choose `Launch Remote Console` if you have VMWare installed on your desktop (which has a much clearer screen resolution). Otherwise you can use `Launch Web Console.`



11. Enter full screen mode to maximize your desktop real estate.



0.3 Navigating HCCX or ICC-VE environments

If you are performing CPD workshops on HCCX or ICC-VE servers, you will be sent an email that has your Access URL and Password for your cluster.

0.3.1 Logging into the environment

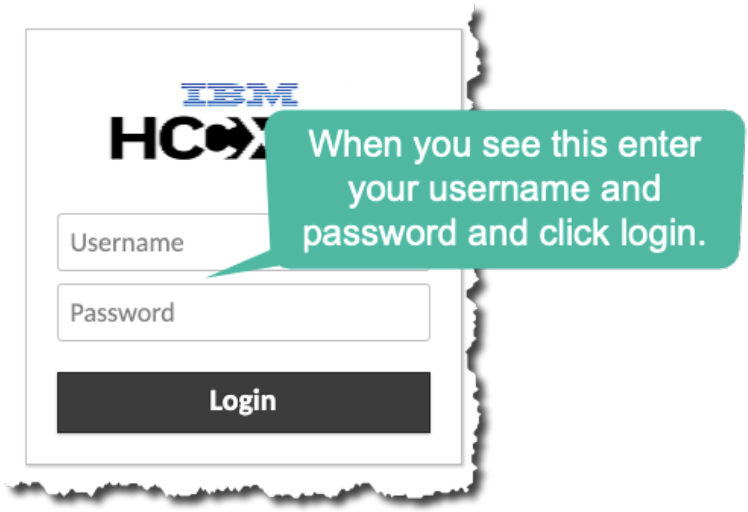
12. Locate the email which will look like the image below.

8 Student IDs for Clusters: (RTP=2 and PNJ=1 and CHI=5)

Student ID	Password	Access URL	Clone Cluster
WDCv81-RTP-Student-06	MODM-BRCRQ5	http://rtp-cloud2.tec.ihost.com:8080/guacamole/#	Raleigh-Clone-1
WDCv81-RTP-Student-07	MODM-BRCRQ5	http://rtp-cloud2.tec.ihost.com:8080/guacamole/#	Raleigh-Clone-2
WDCv81-PNJ-Student-09	MODM-BRCRQ5	http://pal-cloud2.tec.ihost.com:8080/guacamole/#	Paramus-Clone-9
WDCv81-CHI-Student-01	MODM-BRCRQ5	http://chi-cloud2.tec.ihost.com:8080/guacamole/#	Chicago-Clone-1
WDCv81-CHI-Student-02	MODM-BRCRQ5	http://chi-cloud2.tec.ihost.com:8080/guacamole/#	Chicago-Clone-2
WDCv81-CHI-Student-03	MODM-BRCRQ5	http://chi-cloud2.tec.ihost.com:8080/guacamole/#	Chicago-Clone-3
WDCv81-CHI-Student-04	MODM-BRCRQ5	http://chi-cloud2.tec.ihost.com:8080/guacamole/#	Chicago-Clone-4
WDCv81-CHI-Student-05	MODM-BRCRQ5	http://chi-cloud2.tec.ihost.com:8080/guacamole/#	Chicago-Clone-5

13. You will receive a link to your cluster, a username (the Student ID) and a password.

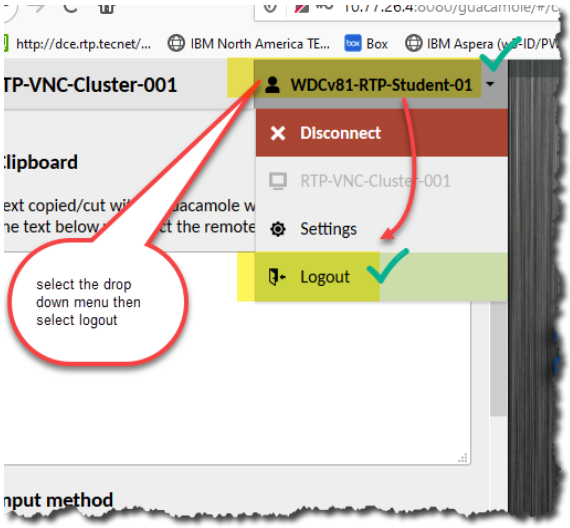
    Click the link and enter your `Username` and `Password` once you reach the Guacamole (remote access) login screen as shown below.



0.3.2 Logging out of the HCCX or ICC-VE Guacamole client

When ending a session, it is best to log out of the client rather than just close the browser.

14. Do this by clicking `CTL`+`Alt`+`SHIFT` to bring up then guacamole menu.
15. Then, click on the `Student ID` (username) you are logged in as ⇨ `Logout`.

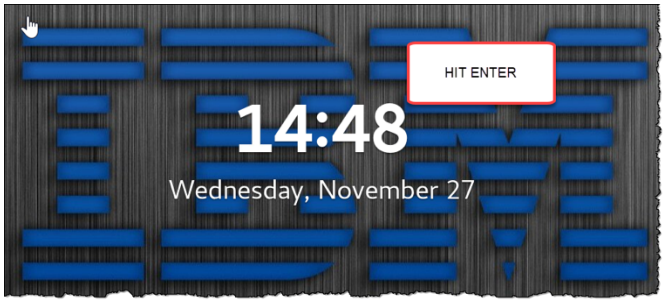


16. To close the Guacamole menu, simply click `CTL`+`Alt`+`SHIFT` again.

0.4 Becoming root in a desktop terminal

Once you have started the cluster (or it has been started for you) continue with these steps to ensure your cluster is ready for the workshop labs.

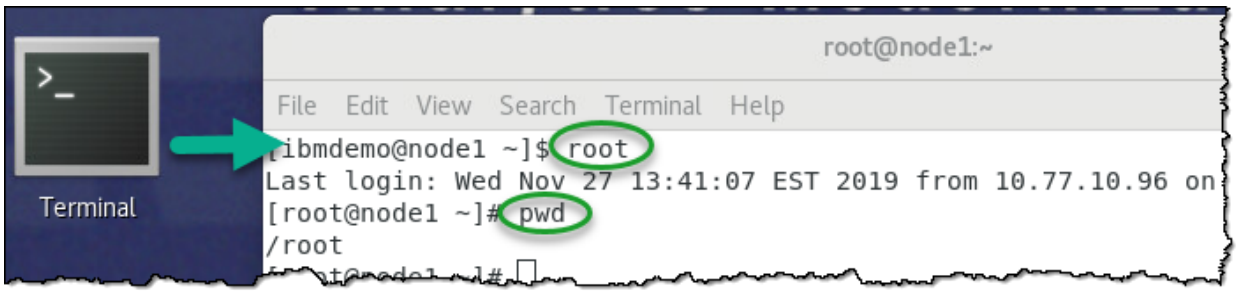
17. If you encounter a screen saver (as shown below) hit `[Enter]` to get past it.



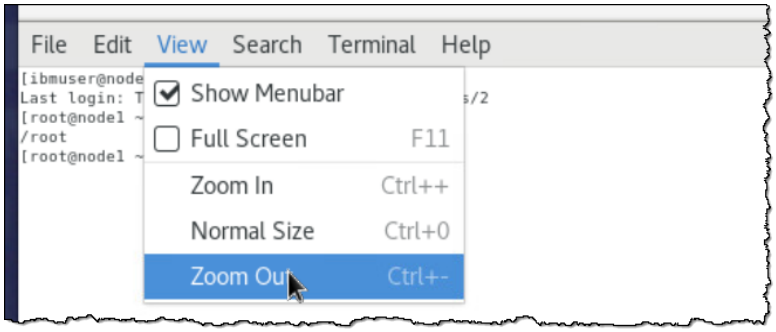
	Note: the Red Hat OS booting username / password is `ibmdemo / Cpdacc3$$` The screensaver lock has been disabled and automatic login has been enabled, so the workshop students are not required to know the login username. This information is provided just in case you should get locked out for an unforeseen reason.

18. On the OS desktop, click the `Terminal` icon ⇨ `root` ⇨ `pwd`
(Note: for those accessing HCCX cluster via browser, you can run `Terminal or Putty`, `ssh to Master1` ⇨ `root` ⇨ `pwd`)

(FYI: the password for `root` is `Cpdacc3$$` – however this alias does use password-less sudo)



19. Tip: If you want to see more output per line, choose `View` ⇨ `Zoom Out`



0.5 Testing the OpenShift Client

Make sure the OpenShift client (oc) is logged in and working properly.

20. In the *Terminal* window, type:

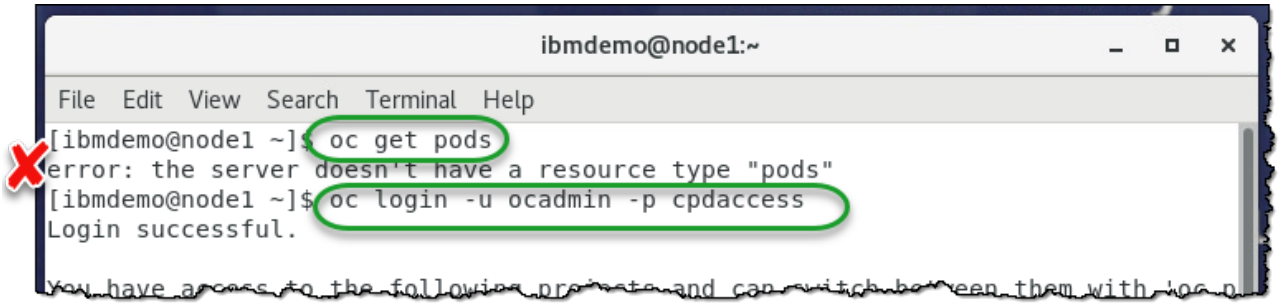
    # `oc get pods`

    This should return a list of all pods running in the zen (CPD) project.

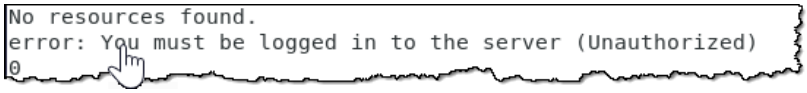
21. If you get an error “the server doesn’t have a resource type ‘pods’ ” or “no resources found”, then you need to log into oc with these commands:

    # `oc login -u ocadmin -p cpdaccess`

    # `oc get pods`

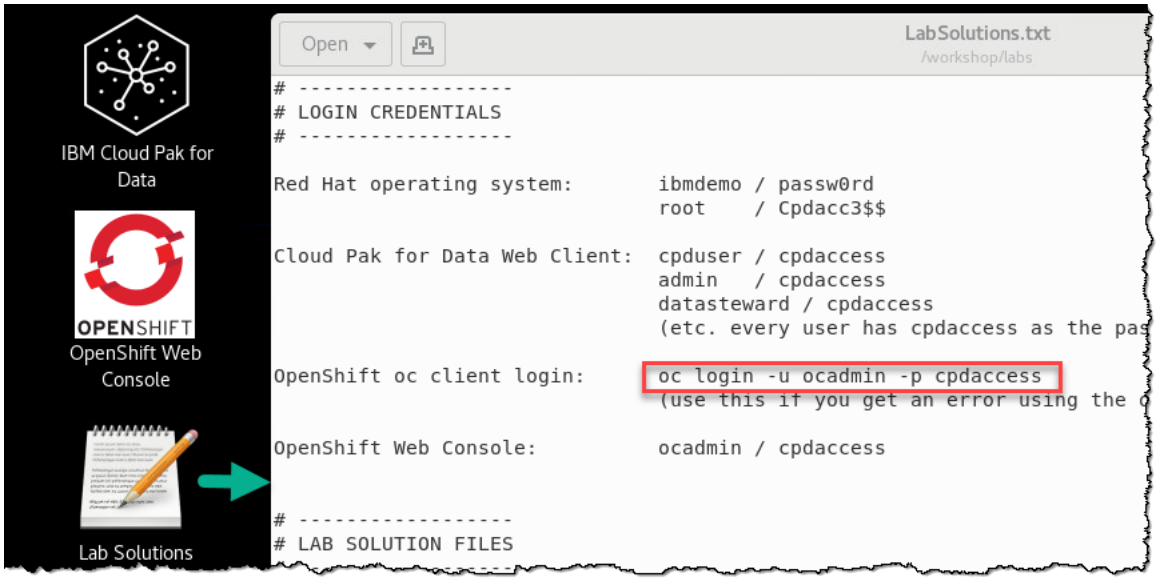


Note: the error returned could have also looked like this.



Occasionally the OpenShift Client (oc) will log out on its own, requiring you to log back in again as shown in the above screen shots.

Note: The desktop icon `Lab Solutions` has various workshop credentials information for convenient reference.



0.6 Cluster environment health check

22. From the terminal, type this alias:

    # `checkall`

    This alias runs a master script that in turn runs other scripts that each check for the health of the various components of the environment. These scripts also outline a potential fix for any problems encountered with each check.

    Note: Each script has an alias which can be run to re-check just that portion of the CPD environment, however there is need to do so if the `checkall` output is healthy.

checkcluster	Status of the OpenShift cluster components
checkntp	Status of the Network Time Protocol (NTP) on the cluster
checkinternet	Status of internet access availability
checkssh	Status of SSH from one node to each other
checkiospeed	Status of Input / Output speed of the disk used by the cluster

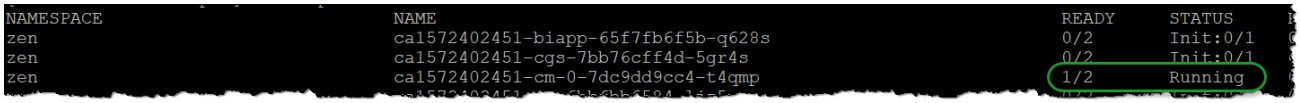
0.7 Pods check

23. From the terminal, type this alias:

    # `checkpods`

    Your CPD cluster is ready when all of the pods are successfully running and ready for all projects (namespaces) in your OpenShift cluster.

    Note: It is not enough that the pods show a STATUS of Running, they must also have a complete READY state of 1/1 or 2/2, etc. meaning all the containers in the pod are Ready.



24. *Optional*: To see the wide list (with node and IP addresses) of the pods not running:

    # `checkpodsw`

25. *Optional*: To see a count of the pods not running (instead of a list):

    # `checkpodscount`

    When this count goes to 1, that means all pods are running. This is because the output counts the header in the list of pods which is all that will show when all pods are successfully running.

26. *Optional*: To run the checkpods script repeatedly every 2 seconds, run this alias:

    # `checkpodswatch`

    [Ctrl]-C to break out of this mode.

    If you do this technique, you can always open another terminal and become root to continue with other work on the cluster while this “watch” continues to run.

0.8 CPD web client check of key data services

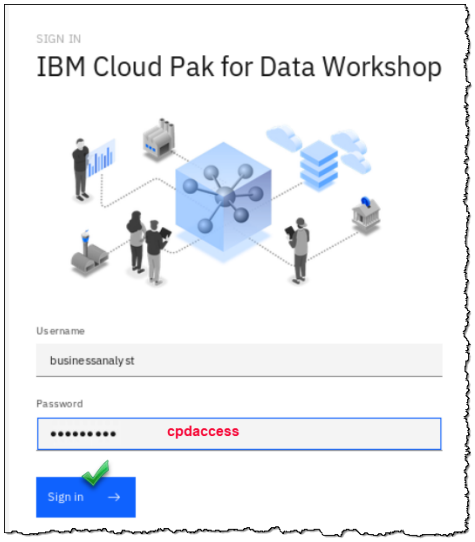
Once all pods are running on the cluster, do a final check within the CPD web client to make sure the key data services are operational:

0.8.1 Launching the web client

27. Click the desktop icon: `Cloud Pak for Data Web Client`

28. The CPD web client GUI displays as shown.

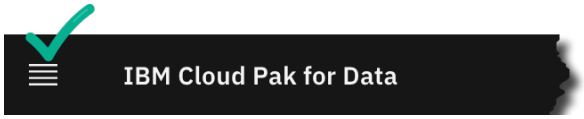
    Use `cpduser` and `cpdaccess` for the *Username* and *Password* and click `Sign in`.



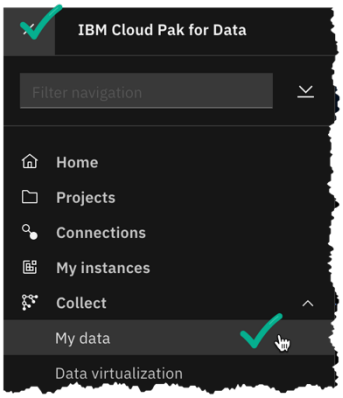
0.8.2 Reviewing Db2 Warehouse and MongoDB

Make sure the Db2 instance is ready.

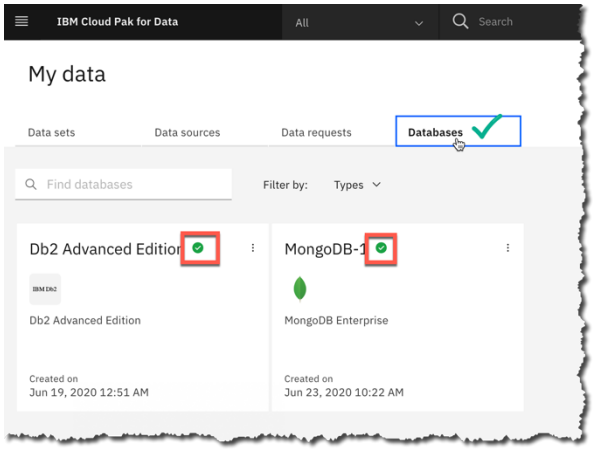
29. Start at the `Navigation Menu`



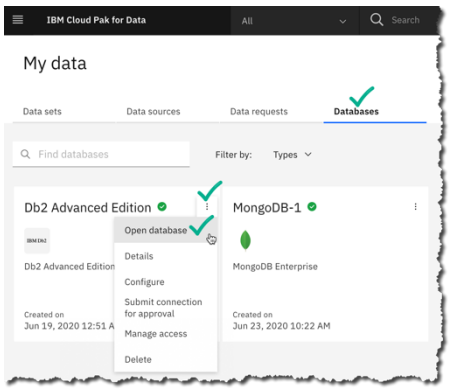
30. Click `Collect` ➪ `My Data`



31. The databases should be up and running



32. If Db2 is not up and running, you can choose the `ellipses` then `Open` to nudge the Db2 pod to start.



33. If MongoDB is not up and running after 25 minutes from the cluster boot up and it is stuck on `Step 5 – Deploying the Mongo replica,` you can "jog" MongoDB to finish deploying by clicking on the `Ellipses` ⇨ `Details` screen.

    (Note: there is no need to select `Open` as you did with the Db2 check, just use `Details`.)

34. Wait a few minutes and the MongoDB tile should render as fully deployed.

    If that did not work, you can force a restart of the deploy by deleting the Mongo replica set pod:

    ```
    # oc get pods -n cp4d | grep mongo
    ```

    ```
    # oc delete pod -n cp4d [mongo replica set pod name]
    ```

    Wait a few minutes and the Mongo DB tile should render as deployed.

35. If that does not work, the last resort is to force the MongoDB deployment is to delete of the mongo pods. Use these commands:

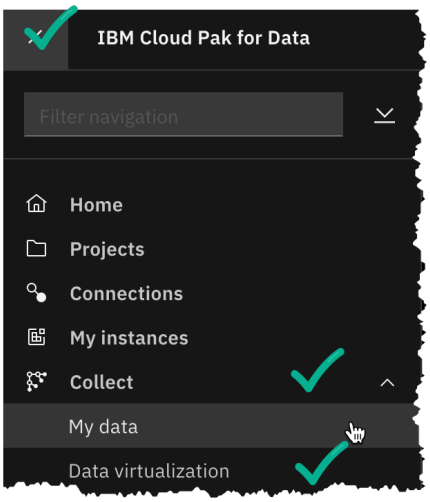
    ```
    # oc get -n cp4d pods | grep mongo
    ```

    ```
    # oc delete pod -n cp4d [podname1] [podname2] [podname3] [podname4] etc.
    ```

0.8.3 Reviewing Data Virtualization

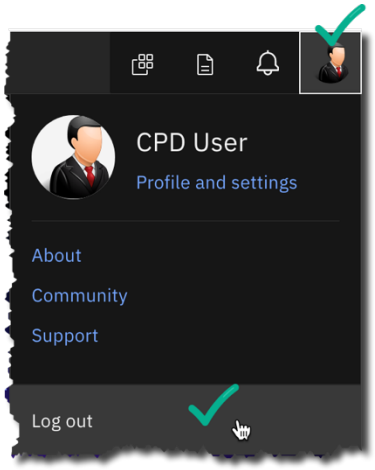
Make sure the Data Virtualization instance is ready.

36. Start at the `Navigation Menu` ⇨ `Collect` ⇨ `Data Virtualization`



37. The DV screen should open up without a provisioning screen. It will not have any virtualized assets, but it will be available and ready to be used.
38. Log out of the CPD web client

    At the top right corner of the screen, click icon for `cpduser` ⇨ `Log out`



0.9 Priming the “Organize” services (optional)

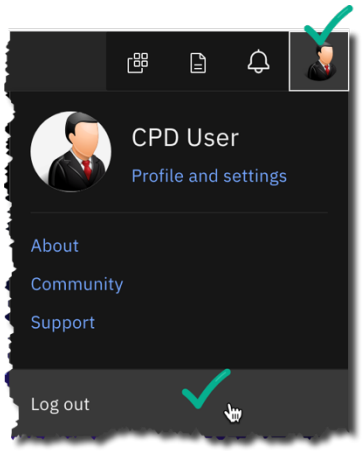
We have experienced that the very first invocation of any of the services under the `Organize` dropdown from the `Navigation Menu` can be sluggish and take a few minutes to render.

If you are starting this cluster for someone else to use and want to give them a better experience, you can launch these services up front in the CPD web client.

Taking the time to do this yourself up front will eliminate the slow response for them when they get to that part of the lab exercises.

39. Log into the `Cloud Pak for Data Web Client`
40. Click `Organize` ⇨ `Data and AI governance` ⇨ `Categories`
41. Click `Organize` ⇨ `Data and AI governance` ⇨ `Business terms`
42. Click `Organize` ⇨ `Data and AI governance` ⇨ `Policies`
43. Click `Organize` ⇨ `Data and AI governance` ⇨ `Rules`
44. Click `Organize` ⇨ `Data and AI governance` ⇨ `Classifications`
45. Click `Organize` ⇨ `Data and AI governance` ⇨ `Data classes`
46. Make sure to log out of the CPD web client after that:

    At the top right corner of the screen, click icon for `cpduser` ⇨ `Log out`



0.10 Troubleshooting

This section is only if you are encountering problems with the cluster booting up properly.

0.10.1 Problem with: checkcluster

If `checkall` shows that one of the OpenShift, Docker or NFS services is not running on one of the nodes, you can try to start that service manually.

47. First ssh to the node having the problem (If it is Master1, no need. You are already at Master1.)

    Example: # `ssh Worker1`

    Then run the start command for the service that is not running:

    Example: # `systemctl start atomic-openshift-node`
    Example: # `systemctl start docker`
    Example: # `systemctl start nfs-server`

48. Then exit out of the ssh session

    Example: # `exit`

49. Check the cluster health again by issuing the following alias:

    # `checkcluster`

50. If these actions do not fix the problem, shut down the cluster and reboot again.

    See: **Shutting down the cluster**

    If a reboot does not fix the problem, the cluster has been provisioned badly and should be discarded.

    Request and/or provision a replacement cluster.

0.10.2 Problem with: a few pods not starting

If a few pods (< 10) are not starting after 45 minutes, you can always delete them and OpenShift will bring them up again, usually successfully.

The pods below are the ones that most often start the slowest and if left alone will eventually start on their own. They are not needed in the first few labs of the workshop anyway.

```
File  Edit  View  Search  Terminal  Help
Every 2.0s: /workshop/scripts/checkpods.sh

NAMESPACE                       NAME                                                READY     STATUS
zen                             wdp-profiling-598f877d48-rf4vz                      0/1       Running
zen                             wkc-glossary-service-5975d7bd5f-n24tm               0/1       Running
```

```
NAMESPACE                       NAME                                                READY     STATUS
zen                             wkc-workflow-service-bbb687d96-p7wnz                0/1       Running
```

Note: for a pod to be truly “running,” it is not enough that a pod show a STATUS of Running. It must also show that all the containers in the pod are Ready. Examples:

```
NAME                                                READY     STATUS    RESTARTS   AGE
dataconn-engine-opdiscovery-5cbbb876b8-rjdv5        1/1       Running   2          14h
dv-0                                                3/3       Running   18         12d
dv-1-1574875471738-ibm-unified-console-6d65d68c68-xrzrz   1/1       Running   2          14h
dv-3-1574875471738-ibm-dv-api-65dbdc68f7-wtsq6      1/1       Running   7          14h
dv-addon-6b46cb89cd-w2jd9                           1/1       Running   2          14h
dv-caching-9c7cbc97f-67fnp                          1/1       Running   2          14h
dv-service-provider-7d9488754d-pkgwp                1/1       Running   7          14h
[root@node1 scripts]#  oc get pod | grep elas
elasticsearch-master-0                              2/2       Running   6          1d
elasticsearch-master-1                              2/2       Running   14         14d
elasticsearch-master-2                              2/2       Running   12         14d
[root@node1 scripts]#
```

51. To delete a pod, first refer to the `checkpods` output for the NAMESPACE and NAME of the pod. Then delete the pod:

    `# oc -n [NAMESPACE] delete pod [NAME]`

    The non-running pod will be deleted and immediately restarted.

    Note: you can delete more than one pod at a time in the same namespace, like this:

    `# oc -n [NAMESPACE] delete pod [NAME1] [NAME2] [NAME3]…`

52. If it has fallen off the `checkpods` list, you can double check that the pod has successfully started:

    `# oc -n [NAMESPACE] get pod [NAME]`

53. *Optional:* If you wish to see the pods in any namespace (also called a “project” in OpenShift), use:

    `# oc get project`

    `# oc get pod -n [project]`

	Cloud Pak for Data pods are contained in the cp4d project (namespace) with the exception of the Watson APIs, which run in zen. However, OpenShift itself runs in a number of others. You may have to occasionally delete non zen pods to “nudge” them to start.

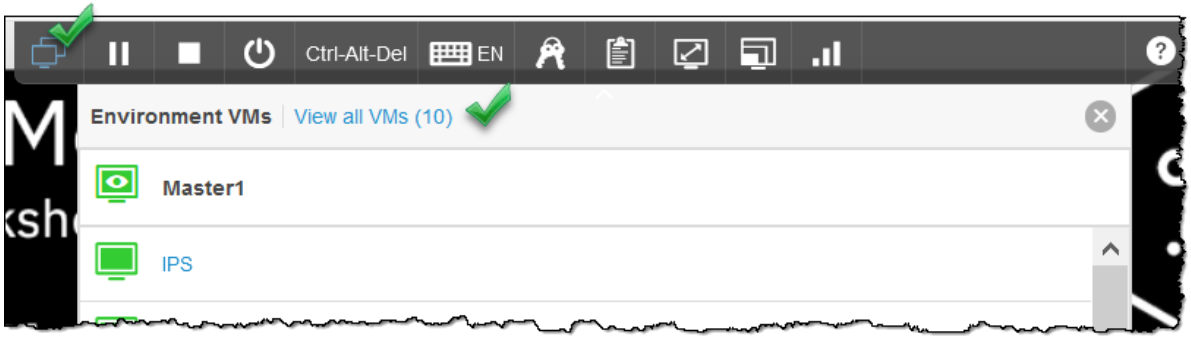
0.11 Shutting down the cluster

0.11.1 Shutting down gracefully (for a reboot)

54. To shut down the cluster gracefully, use the following alias:

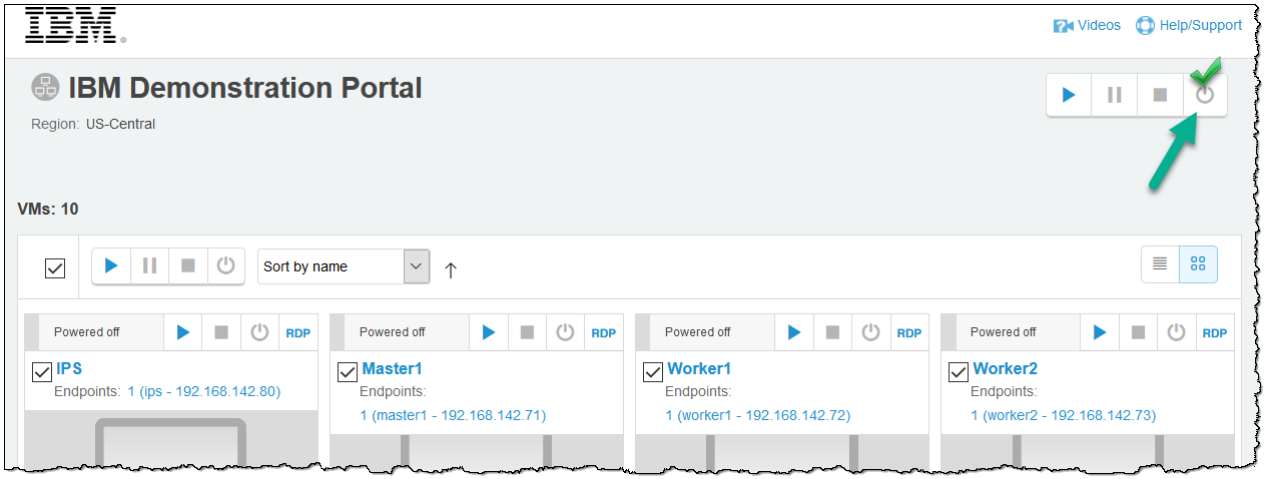
    # `shutdowncluster`

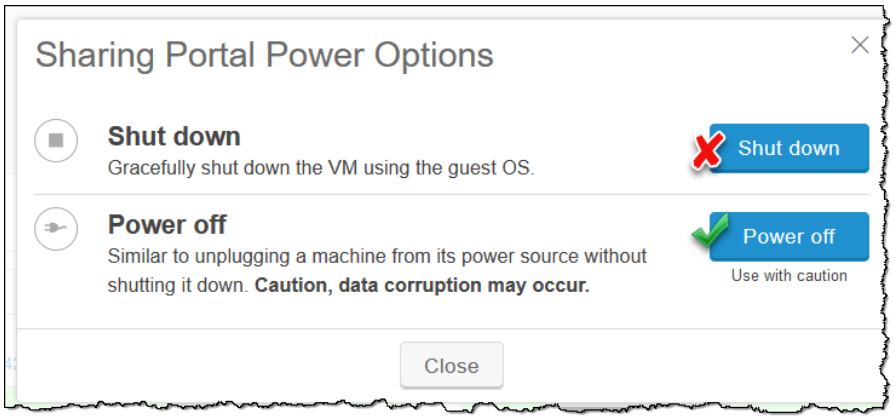
55. In Skytap environments, find the VMs icon in the drop down menu ⇨ `View all VMs`



56. Stop the virtual machine images by clicking the `Power options` ⇨ `Power off`

    Note: Do NOT use `Shutdown`; see next section for details on a forceful shutdown.



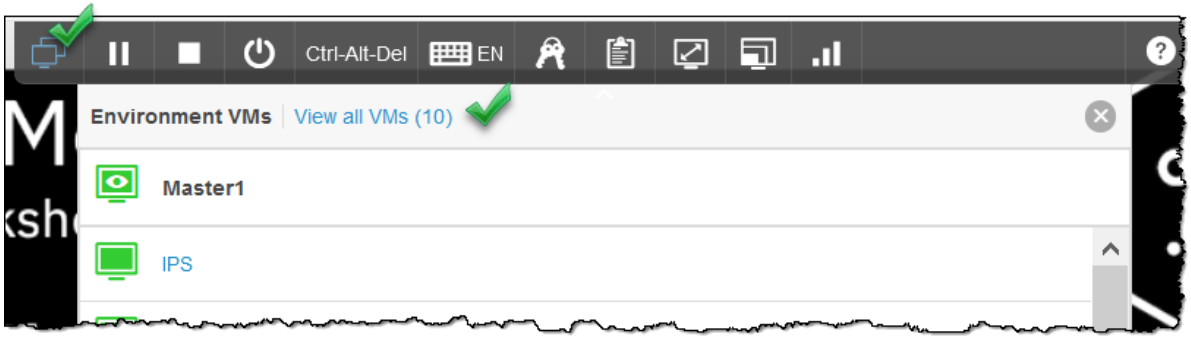


57. If any node does not power down properly, the Power Off that individual node (VM) that did not process properly.

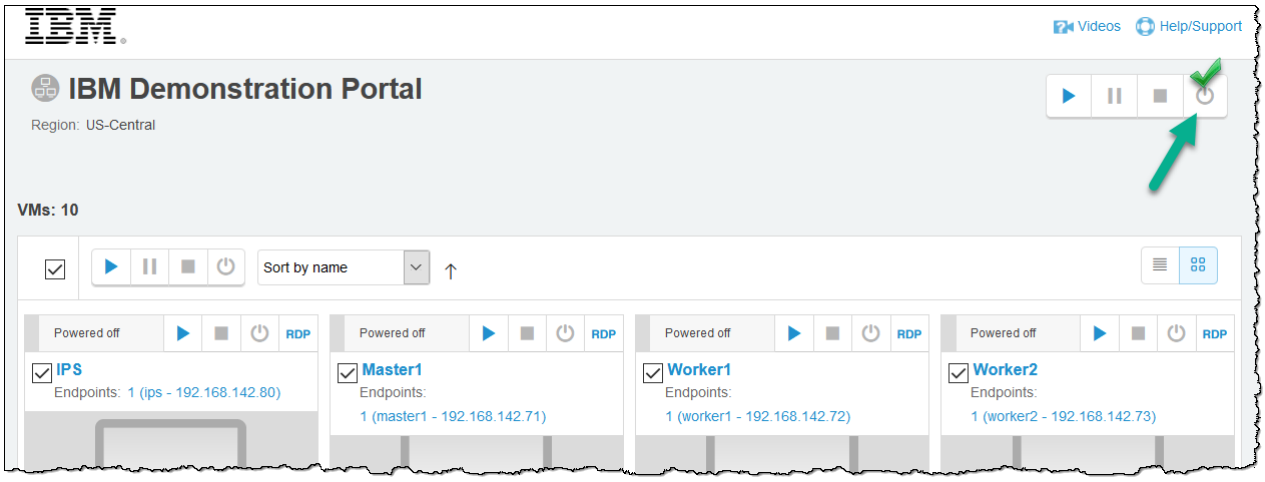
0.11.2 Shutting down forcefully (if graceful option does not work)

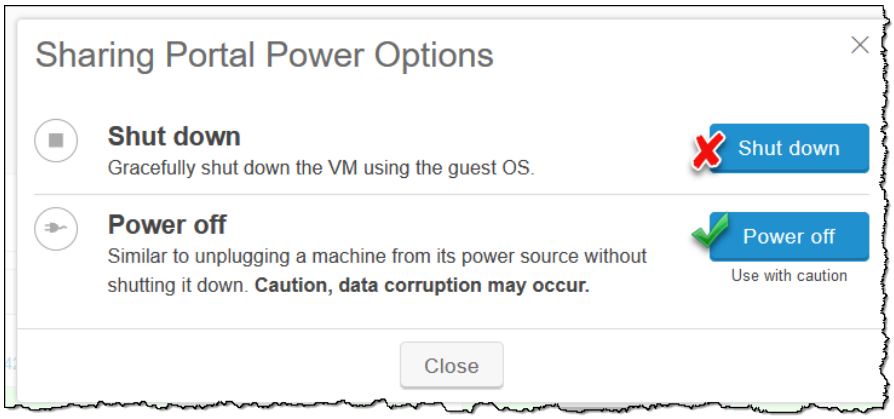
If a graceful shutdown is not an option (the cluster suspended, or crashed) you can force a Power Off on all clusters without using the alias shutdowncluser.

58. In Skytap environments, find the VMs icon in the drop down menu ⇨ `View all VMs`



59. Stop the virtual machine images by clicking the `Power options` ⇨ `Power off`





60. If any node does not power down properly, the Power Off that individual node (VM) that did not process properly.

0.12 Fixing a failing data profiling (and data class list)

If you are having problems with the profiling exercise in the Organize lab, this is how you can fix this problem:

61. Refresh your browser until the data class list shows up.
62. Open a terminal window and log into the oc cluster (or open the OpenShift console and navigate to pods.)
63. Find the pod that's name starts with "wdp-activities"

    ```
    # oc get pod -n cp4d | grep wdp-activities
    ```

64. Delete that pod and let it regen

    ```
    # oc delete pod -n cp4d [podname]
    ```

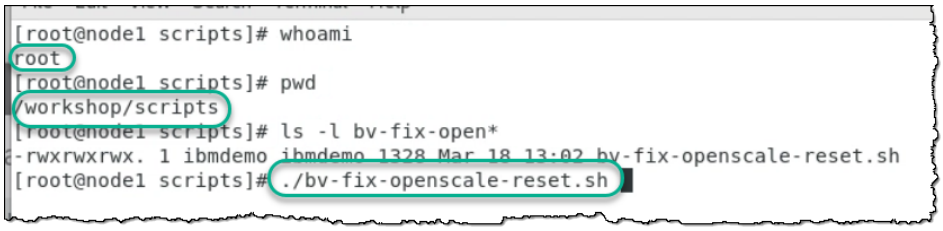
65. Wait for the pod to be fully operational. The profiling will now work as expected.

0.13 OpenScale lab configuration step (when required)

Your workshop cluster has been configured so that you can run the OpenScale Auto Setup without requiring any configuration steps.

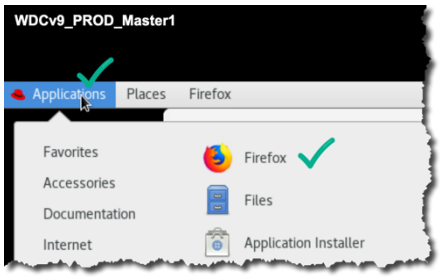
If for any reason you find the need to reset OpenScale to do the Auto Setup again, follow these steps:

66. First, open a Terminal become `root`, change directories to `/workshop/scripts` and run script `bv-fix-openscale-reset.sh` (This cleans up the Db2 Warehouse OpenScale schemas)



67. Second, in the CPD web client go to the `Analyze` ⇨ `Analytic deployments.` First delete the deployment, then the model. (If these are not in the deployment space, there is nothing to do.)

68. Finally, open a Firefox browser: `Applications` ⇨ `Firefox`



69. Type in the browser this URL: http://ibm.biz/OpenScaleLabConfiguration

70. `Download` the file to your working cluster.

    Note: the file will be stored in **`/home/ibmdemo/Downloads/openScaleLabConfigurtion.txt`**

71. Open the text file and follow the instructions. (Use the desktop file browser, gedit or vi.)

    This file facilitates a copy / paste into a Terminal window for the configuration work of resetting the OpenScale warehouse in the Db2 Warehouse database.

    Once all the above commands are run in the terminal you can now do the OpenScale lab Auto Setup again.

0.14 Other tips and tricks

0.14.1 CPD Web Client Information Server credentials pop-up

Occasionally an Information Server credentials window will pop up in the CPD web client. You can resolve this problem in two ways:

1) Respond to the pop-up with username/password: `isadmin/isadmin`
2) Clear the browser cache

Note: don't confuse this credentials request with one for user cpduser, which may occasionally pop up as well. In that case answer that credentials request with `cpduser/cpdaccess`

0.14.2 "Firefox already running" message

If you cannot start up the CPD web client in Firefox because it is giving you a message saying it is already started, do this:

```
# pgrep firefox
```

This will return the PID of the Firefox service. Kill that PID with this command:

```
# kill -9 [PID]
```

Now you can start the CPD web client as usual.

0.14.3 Aliases

You can review all the aliases used to develop the workshop:

```
# more /workshop/scripts/alias.sh
```

To see them resolved by the OS, use:

```
# alias [alias name]
```

0.14.4 Workshop directories

For further exploration and customization of the workshop cluster to understand how it was set up, review the directory `/workshop`

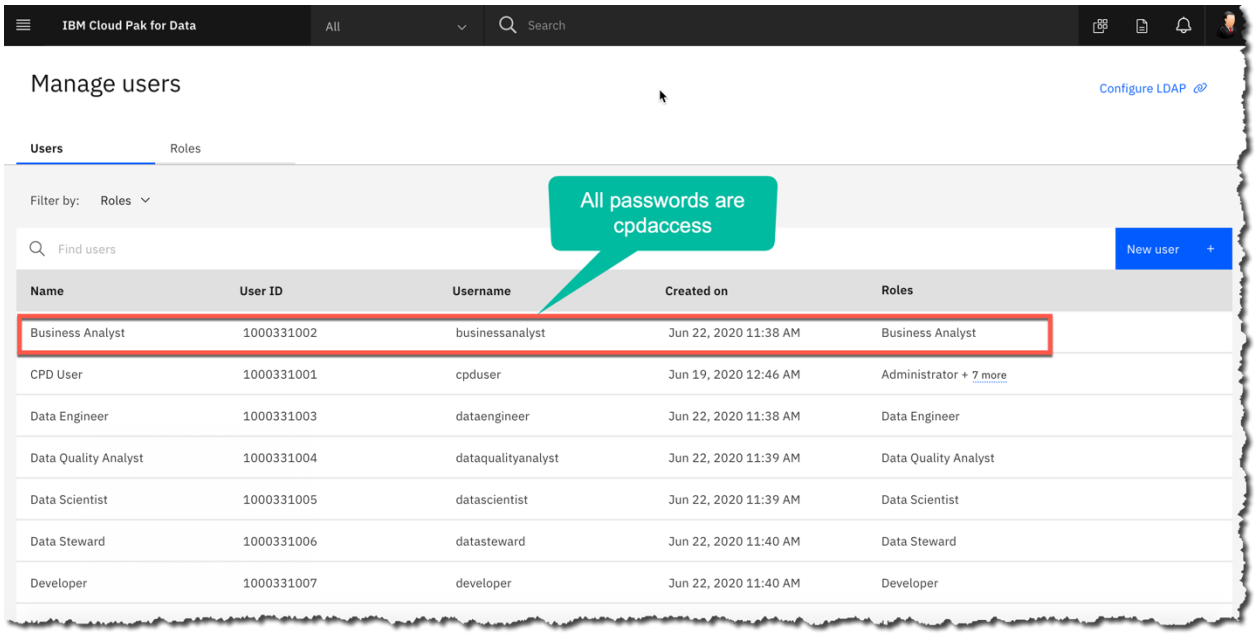
`/desktop`	Source for the workshop launchers and images
`/labs`	Source for individual lab assets
`/scripts`	Alias, cron, check, and fix scripts as well as the OpenScale setup directory (/fastpath) and scripts.
`/setup/db2`	Db2 setup scripts and data
`/setup/mongo`	MongoDb setup scripts and data
`/stocktrader`	Not used in the 8.x versions of the workshop but kept for backwards compatibility.

0.15 CPD web client users and roles

Every CPD role is represented in the workshop cluster with a matching user, using `cpdaccess` as the password .

Utilize these users for demoing purposes as well as your own experimentation on how permissions are treated on the CPD cluster between various users and roles.

To review them: `Navigation Menu` ⇨ `Administer` ⇨ `Manage users`



Log out of the CPD Web client and log in with the one you wish to use (for your own exploration):

Please Recycle