Installing IBM Open Platform with Apache Hadoop Non-Production Environment Docker image

Installing the free IBM BigInsights Quick Start Editions, non-production software

IBM Open Platform with Apache Hadoop, v4.1: Docker image README

Welcome to the IBM® Open Platform with Apache Hadoop Docker image readme.

This Docker image provides the quickest way to set up IBM Open Platform with Apache Hadoop so that you can try it with real data and run real applications in a non-production environment. There is no data limit on the cluster, and there is no time limit on the license. This image uses a non-warranted program license, and is not for production use.

To read more about the IBM Open Platform with Apache Hadoop and the features that it includes, see the BigInsights® documentation, http://www.ibm.com/support/knowledgecenter/SSPT3X_4.1.0/com.ibm.swg.im.infosphere.biginsights.welcome.doc/doc/welcome.html.

System requirements

Before you download, ensure that your system meets the minimum requirements:

Operating system

One of the following supported host operating systems and docker versions:

- Apple MacOS X 10.8+ 64-bit (Docker-Toolbox-1.8.1c)
- Red Hat Enterprise Linux (RHEL) 6.5 or higher - 64-bit (docker 1.6.2)
- Red Hat Enterprise Linux (RHEL) 7.x - 64-bit (docker 1.8.1)
- Windows 7+ -64-bit (Docker-Toolbox-1.8.1c)

Hardware requirements

- A minimum of 4 core processors for the Docker image and the IBM Open Platform with Apache Hadoop software.

  To get the best performance, it is recommended that your system have a minimum of 8 processor cores. The product will work with fewer processors, but you will experience a significant difference in performance.
- At least 4 GB RAM for IBM Open Platform with Apache Hadoop and the Docker image. Therefore, your host machine should have RAM that exceeds these memory levels.
- A minimum of 30 GB of free space on the host hard disk.

  The Docker container image size is 3.5 GB compressed.

Getting Started

Procedure

1. Download the Docker image.

   Download the IBM Open Platform with Apache Hadoop 4.1 Docker image from the following site: http://www.ibm.com/software/data/infosphere/hadoop/trials.html.

   The ZIP file contains a TAR file for the Docker image and the loadBIDocker.sh to load the image.

2. Extract the contents of the ZIP file to the following location:

   Windows environments
   Extract the contents of the downloaded ZIP file to the current default user profile directory: `c:/Users/<current_windows_user_profile>` directory. To determine the default user profile, type **`echo %USERPROFILE%`** in the command line.

   MAC environments
   Extract the contents of the downloaded ZIP file to any local directory.

   The default Archive Utility tool might not work. A suggested tool is the Unarchiver Utility: https://itunes.apple.com/us/app/the-unarchiver/id425424353?mt=12

3. Set up your environment for loading the Docker image.

   For Windows operating systems, you might need to verify and update the BIOS to enable virtualization.

   For Windows and Mac you must use a docker-machine to load the Docker image.

   Attention: If you have previously installed boot2docker, which is now deprecated, it is highly recommended that you migrate to a docker-machine. If you have both boot2docker and docker-machine in your system, you will have issues in loading the Docker image. Before installing a newer docker-machine application, make sure you uninstall, delete, and clean up any traces of the previous boot2docker application.

   Setting up a docker-machine environment (or simply **`Docker Toolbox`** which contains a docker-machine):

   On Windows systems
   Install the docker-machine from the following link:
   `https://docs.docker.com/installation/windows/`

   Accept the defaults and click **Finish** when complete.

   On MAC systems
   Install the docker-machine from the following link:
   `https://docs.docker.com/installation/mac/`

   Accept the defaults and click **Finish** when complete.

   On RHEL systems
   Verify that **`docker-io.x86_64`** or **`docker-engine.x86_64`** is installed with the following command:
   `yum list installed | grep docker`

   If it is not installed, install it by using the following command:

   RHEL6.5 or higher
   `yum install docker-io`

   RHEL7.x
   follow the directions at `https://docs.docker.com/installation/rhel/`

   Tip: If you use an Ubuntu Linux distribution, do the following steps to make sure that you use the correct location for storing Docker related data:

   a. From the Linux terminal window, edit the `/lib/systemd/system/docker.service` file to add `-g /mnt/docker` to the end of the **`ExecStart`** option.
   b. Add the following content to the `/lib/systemd/system/docker.service` file:

```
[Unit]
Description=Docker Application Container Engine
Documentation=http://docs.docker.com
```

```
After=network.target docker.socket
Requires=docker.socket

[Service] EnvironmentFile=-/etc/default/docker
ExecStart=/usr/bin/docker -d $DOCKER_OPTS -H fd://
MountFlags=slave LimitNOFILE=1048576
LimitNPROC=1048576
LimitCORE=infinity

[Install]
WantedBy=multi-user.target
```

c. Run the following command:

```
systemctl daemon-reload
```

d. Start the docker service.

```
service docker start
```

To verify that your Docker machine is running correctly, double-click the **Docker Quickstart Terminal** icon on your desktop and type the following command:

```
docker run hello-world
```

Processing occurs, and some error-like messages might appear, but you should see the following if your environment is set up correctly:

```
...
Status: Downloaded newer image for hello-world:latest
Hello from Docker.
This message shows that your installation appears to be working correctly.
...
```

4. Load the Docker image. The load action creates a new **bidocker** Docker machine with the required configurations and starts the **bidocker** Docker machine with all the environment settings.

Option	Description

Loading the Docker image on RHEL	1. Change to the directory where the image is downloaded, verify the docker status, and run the script as follows:
	`[root@localhost ~]# service docker status`
	`docker (pid 2079) is running...`
	`[root@localhost ~]# ls`
	`bi_v41_iop.tar loadBIDocker.sh`
	`[root@localhost ~]#./loadBIDocker.sh –load`

Option	Description

Loading the Docker image on Windows 7/MAC	1. Open a **Docker Quickstart Terminal** (double-click on the icon). This action creates a default VM.
	Tip: If you are running the Docker machine in a Windows environment, the first time that you open the Docker Terminal, the operations might seem to be stuck. Progress is actually being made, but the terminal might not show any activity for some time.
	2. Load the image.
	For Windows environments
	a. When you open the **Docker Quickstart Terminal**, if you are prompted for a password, enter the Windows administrator password and allow firewall access.
	b. Confirm the VM is up and running by opening the **VirtualBox** application.
	c. Change to the directory where you extracted the Quick Start Edition image:
	`cd /c/Users/my_windows_user_profile_folder/BigInsights_QSE_v4100_xxxxxx`
	d. Run the **load** utility.
	`docker load --input bi_v41_iop.tar`
	e. Create the docker container by typing the following command on one line:
	`echo "" | docker run -i --name iop_cont -e TERM=xterm --privileged=true -P -p 8080:8080 -p 51000:51000 -p 8443:8443 -p 10000:10000 -p 50070:50070 -p 19888:19888 -p 8088:8088 -p 18080:18080 -h rvm.svl.ibm.com -m 4g bi_v41_iop:latest //bin//bash`
	f. Start the container:
	`docker start iop_cont`
	g. With the container running, you can specify some useful configuration parameters, such as starting the bash shell:
	`docker exec -i -t iop_cont //etc//start-all.sh -bash`
	The **-i** keeps an interactive display. The **-t** allocates a pseudo TTY.
	h. Open a new Windows terminal to run a utility to determine the proper IP address:
	`docker-machine ip default`
	Make a note of the IP address.
	i. Update the `C:\Windows\System32\drivers\etc\hosts` file by appending the following line:
	`<ip_address> rvm.svl.ibm.com`
	For Mac environments
	In the **Docker Quickstart Terminal**, change to the directory where the image is downloaded and run the **loadBIDocker.sh** script:
	`$cd <installed_dir>`
	`$ls`
	`bi_v41_iop.tar`
	`loadBIDocker.sh`
	`$./loadBIDocker.sh --load`

5. Load the docker image manually (if needed) Follow these steps if you cannot use the script to load the image. Otherwise, skip to the next step.
   a. Change to the directory where the image is downloaded, and run the following commands:
      1) `docker load --input bi_v41_iop.tar`

         Note: Loading the image might take a long time on the **docker-machine** environment.
      2)
```
echo "" | docker run -i --name iop_cont -e TERM=xterm
  --privileged=true -P -p 8080:8080 -p 51000:51000 -p 8443:8443 -p 10000:10000
  -p 50070:50070 -p 19888:19888 -p 8088:8088 -p 18080:18080
  -h rvm.svl.ibm.com -m 4g bi_v41_iop:latest //bin//bash
```

3) `docker start iop_cont`
4) `docker exec -i -t iop_cont //etc//start-all.sh -bash`
5) If you come out of the bash prompt and want to go back, do the following commands:
   a) Check the status of your container:
      `docker ps -a`
   b) If the container is stopped, start the container:
      ```
      docker start iop_cont
      docker exec -i -t iop_cont //etc//start-all.sh -bash
      ```
6) Run the following command to get the bash prompt:
   `docker exec -it iop_cont bash`

b. Map the host IP.
   1) Determine the IP address of the container.

      Windows/MAC
      `docker-machine environment` - Open a separate docker-machine window and run the following command to determine the IP
      `docker-machine ip bidocker`

      Linux Open a terminal window and run this command to determine the IP address:
      `docker inspect biginsights_cont | grep \"IPAddress\"`
   2) Update the `hosts` file by appending the following line:
      `<ip_address> rvm.svl.ibm.com`
      - Host file location on Windows – `C:\Windows\System32\drivers\etc\hosts`
      - Host file location on Mac - `/private/etc/hosts`
      - Host file location on Linux - `/etc/hosts`

6. Explore the sample data and tutorials.

   The IBM Open Platform with Apache Hadoop is loaded with the Ambari stack and the associated add-on services such as Hive, YARN, Spark, and HDFS. The necessary repository file is already configured. You will find tutorials and data that relate to Spark and Hadoop in the root Docker image in the `/sampleData` folder.

7. Open the Ambari web interface.

   Open a browser window and type the following address: `http://rvm.svl.ibm.com:8080`. Use **`admin`** for the user ID and **`admin`** for the password.

   The Ambari dashboard opens, and after a few minutes, many of the services are started. The Ambari user interface might show some operations in red when you first see the services. Click the blinking ***n*** **Ops** icon to show the start status of each of the services. The *n* represents the number of operations that are still in the process of completing. It might take 10-15 minutes for the services to start.

8. Optional: If needed, update the Windows BIOS.

   For Windows operating systems, you might need to update the BIOS to enable virtualization. Read the BIOS configuration documentation for your particular operating system. The following steps provide general direction:

   a. Shut down, then reboot your system. Be sure to:
      1) Access the BIOS of your system **before** it completely reboots.
      2) In the Configuration section, click **CPU**.
      3) Enable both of the virtualization settings: **Virtualization Technology** and **VT-d**.
      4) Save the setting.
      5) In the Configuration section, click **Intel AMT** and make sure that it is also enabled. The Console Type should be **VT100+**.
      6) Save and exit. The system continues to reboot.
   b. Then, shut down, and restart the system.

Suspending and resuming the Docker image

If you need to shut down your system, or if you need to temporarily stop using the image, then follow these steps to suspend and then later resume the Docker image.

Procedure

1. From the Ambari web interface, stop all of the services that are running by clicking **Actions** > **Stop All**.
2. From the **Docker QuickStart Terminal**, stop the container.
   ```
   docker stop iop_cont
   ```
3. From the **Docker QuickStart Terminal**, stop the virtual machine.

   If you loaded the image by using the `loadBIDocker.sh` script:
   ```
   docker-machine stop bidocker
   ```

   If you loaded the image by using the manual steps:
   ```
   docker-machine stop default
   ```
4. When you are ready to begin working again with the Docker image, run the **`loadBIDocker.sh`** script or follow the manual load steps.

Known Issues and troubleshooting tips for the IBM Open Platform with Apache Hadoop,v4.1 Docker image

Try these hints and tips if you have trouble with the installation or using the IBM Big Insights Quick Start Edition image.

- When you open multiple **Docker QuickStart Terminal** instances in a Windows environment, you might see errors that result from the host operating system losing a connection to the virtual machine *default*. Do the following steps to ensure that you shut down the Docker properly and allow the system to re-establish a connection between the host and guest operating system.
  1. Open the Virtual Box application and shut down the default VM.
  2. Double-click the **Docker QuickStart Terminal** icon.
  3. Run the following command:
     ```
     docker-machine ssh default
     ```
  4. Change to the directory where you extracted the Quick Start Edition files:
     ```
     cd /c/Users/my_windows_user_profile_folder/BigInsights_QSE_v4100_xxxxxx
     ```
  5. Run the load command.
     ```
     docker load --input bi_v41_valueadd.tar
     ```
  6. If you see messages that pertain to connection issues, repeat step "IBM Open Platform with Apache Hadoop, v4.1: Docker image README" on page 1.

October 2015