

(1)

PUBLIC | AUTOMATED BUILD

sequenceiq (/u/sequenceiq/)/hadoop-docker (/r/sequenceiq/hadoop-docker/) ☆

Last pushed: 5 months ago

Repo Info (/r/sequenceiq/hadoop-docker/)

Short Description

An easy way to try Hadoop

Full Description

#Apache Hadoop 2.7.0 Docker image

Note: this is the master branch - for a particular Hadoop version always check the related branch

A few weeks ago we released an Apache Hadoop 2.3 Docker image - this quickly become the most popular (<https://registry.hub.docker.com/search?q=hadoop&s=downloads>) Hadoop image in the Docker [registry](https://registry.hub.docker.com/) (<https://registry.hub.docker.com/>).

Following the success of our previous Hadoop Docker [images](https://registry.hub.docker.com/u/sequenceiq/hadoop-docker/) (<https://registry.hub.docker.com/u/sequenceiq/hadoop-docker/>), the feedback and feature requests we received aligned with the Hadoop release cycle, so we have released an Apache Hadoop 2.7.0 Docker image - same as the previous version, it's available as a trusted and automated build on the official Docker [registry](https://registry.hub.docker.com/) (<https://registry.hub.docker.com/>).

FYI: All the former Hadoop releases (2.3, 2.4.0, 2.4.1, 2.5.0, 2.5.1, 2.5.2, 2.6.0) are available in the GitHub branches or our [Docker Registry](https://registry.hub.docker.com/u/sequenceiq/hadoop-docker/) (<https://registry.hub.docker.com/u/sequenceiq/hadoop-docker/>) - check the tags.

Build the image

If you'd like to try directly from the Dockerfile you can build the image as:

```
docker build -t sequenceiq/hadoop-docker:2.7.0 .
```

Pull the image

The image is also released as an official Docker image from Docker's automated build repository - you can always pull or refer the image when launching containers.

```
docker pull sequenceiq/hadoop-docker:2.7.0
```

Start a container

In order to use the Docker image you have just build or pulled use:

Make sure that SELinux is disabled on the host. If you are using boot2docker you don't need to do anything.

```
docker run -it sequenceiq/hadoop-docker:2.7.0 /etc/bootstrap.sh
```

Testing

You can run one of the stock examples:

```
cd $HADOOP_PREFIX
# run the mapreduce
bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-

# check the output
bin/hdfs dfs -cat output/*
```

Hadoop native libraries, build, Bintray, etc

The Hadoop build process is no easy task - requires lots of libraries and their right version, protobuf, etc and takes some time - we have simplified all these, made the build and released a 64b version of Hadoop nativelibs on this [Bintray repo](https://bintray.com/sequenceiq/sequenceiq-bin/hadoop-native-64bit/2.7.0/view/files) (<https://bintray.com/sequenceiq/sequenceiq-bin/hadoop-native-64bit/2.7.0/view/files>). Enjoy.

Automate everything


As we have mentioned previously, a Docker file was created and released in the official [Docker repository](https://registry.hub.docker.com/u/sequenceiq/hadoop-docker/) (<https://registry.hub.docker.com/u/sequenceiq/hadoop-docker/>)

Docker Pull Command



```
docker pull sequenceiq/hadoop-docker
```

Owner

 sequenceiq

Source Repository

 [sequenceiq/hadoop-docker](https://github.com/sequenceiq/hadoop-docker) (<https://github.com/sequenceiq/hadoop-docker>)

Comments (21)



clarkliu20083736

a month ago

nice
thank you very much



taylorchua

a month ago

Hii !

I was just curious is there any tutorial on building multi-node hadoop cluster . As i very keen in learn . i'm trying out an experience with 1 master 2 slave node using centos . hope to heard from you soon ! Thanks ! :)



mamun001

3 months ago

Wow!!!! Thanks!!



gnoloahs

4 months ago

宿主机是ubuntu x86_64 , docker run --net=host -it sequenceiq/hadoop-docker:2.7.0
/etc/bootstrap.sh -bash
为什么datanode没有启动呢?



vschiavoni

5 months ago

After submitting the example, the job remains in PREP state. Is this expected ?

`./hadoop job -list`

DEPRECATED: Use of this script to execute mapred command is deprecated.
Instead use the mapred command for it.

16/01/07 04:44:52 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

Total jobs:1

JobId	State	StartTime	UserName	Queue	Priority	UsedContainers	RsvdContainers	UsedMem	RsvdMem	NeededMem	AM info
-------	-------	-----------	----------	-------	----------	----------------	----------------	---------	---------	-----------	---------

job_1452158980050_0001	PREP	1452159025481	root	default	NORMAL	0	0	0M	0M	0M	0M
------------------------	------	---------------	------	---------	--------	---	---	----	----	----	----

http://4b8dae616769:8088/proxy/application_1452158980050_0001/
(http://4b8dae616769:8088/proxy/application_1452158980050_0001/)



afrido

7 months ago

Hi,

I am using this image to create a local hadoop cluster, however when i try to ssh in it always prompt me for root password which is unknown. I was expecting ssh to be a password less as mentioned in the Dockerfile but it doesn't work.

Any comment/feedback would be really appreciated.

Thanks.



leonfl

9 months ago

I think 50030 should be added in EXPOSE



naelsondouglas

10 months ago

Thanks for the project. It will help-me a lot.



goace

a year ago

Never mind..I reinstall the docker and pull the image again, everything goes fine.



goace

a year ago

Hi, I followed your instructions but get an error:

```
$ docker run -i -t sequenceiq/hadoop-docker:2.5.2 /etc/bootstrap
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-example
```

error: java.net.ConnectException: Call From d9c011376e94/172.17.0.12 to d9c011376e94:9000 failed on connection exception.

Seems nothing is listing on 9000, why?

