

## Instalasi dan konfigurasi apache spark

1. Buat folder baru dengan nama hadoop-spark dan masukkan code untuk konfigurasi hadoop, spark, zookeeper

[illegible]

- ## 2. Jalankan build.sh untuk membuat docker image

```
Users\asini\Documents>chadspark-docker\bigdata-sparkbash build.sh
[+] Building 104.1s (64/64) FINISHED
    [internal] load build definition from Dockerfile                                0.1s
    -> transferring dockerfiles: 3.42kB                                          0.0s
    [internal] load metadata for docker.io/library/ubuntu:24.04                  3.0s
    [auth] library/ubuntu:pull token for registry-1.docker.io                    0.0s
    [internal] load dockerignore                                                  0.1s
    -> transferring context: 28                                                  0.0s
    [internal] load build context                                                13.1s
    -> transferring context: 2.22MB                                              13.0s
    [ 1/58] FROM docker.io/library/ubuntu:24.04@sha256:6d15f6692dd7bfbc53558d7cfff325d43de4249f41ade9e074c3f047ac5
     CACHED [1/58] RUN apt-get --y install openssh-client                        0.0s
     CACHED [11/58] RUN apt-get --y install vim                                  0.0s
     CACHED [3/58] RUN adduser --disabled-password --gecos '' docker            0.0s
     CACHED [4/58] RUN adduser docker sudo                                     0.0s
     CACHED [16/58] RUN sudo apt-get --y install tree                           0.0s
     CACHED [6/58] RUN sudo apt-get --y install software-properties-common       0.0s
     CACHED [2/58] RUN apt-get update && apt-get --y install sudo adduser        0.0s
     CACHED [7/58] RUN sudo add-apt-repository ppa:openjdk-r/ppa                 0.0s
     CACHED [5/58] RUN echo "sudo ALT+(ALT) MODSudo:[ALL]" >> /etc/sudoers      0.0s
     CACHED [10/58] RUN ln -s /usr/lib/jvm/java-1.8.0-openjdk-amd64 /usr/lib/jvm/java-1.8.0
     CACHED [8/58] RUN sudo apt-get update                                      0.0s
     CACHED [12/58] RUN apt-get --y install nano                               0.0s
     CACHED [13/58] RUN apt-get --y install wget tar sudo rsync                0.0s
     CACHED [14/58] RUN sudo apt-get update                                    0.0s
     CACHED [15/58] RUN sudo apt-get --y install apache2                      0.0s
     CACHED [17/58] RUN sudo apt-get install net-tools                         4.1s
    [18/58] RUN apt-get install -y openssl-server                             8.7s
    [19/58] RUN ssh-keygen -t rsa -m sha1 -C "asini" -f /root/.ssh/id_rsa         0.0s
    [20/58] RUN cp /root/.ssh/id_rsa.pub /root/.ssh/authorized_keys             0.4s
    [21/58] RUN chmod 755 /root && chmod 700 /root/.ssh                          0.5s
    [22/58] RUN passwd --unlock root                                           0.5s
    [23/58] COPY hadoop-3.4.1.tar.gz /hadoop-3.4.1-bin.tar.gz                 0.8s
    [24/58] RUN tar -zxvf /hadoop-3.4.1-bin.tar.gz -C /                         7.2s
    [25/58] RUN ln -sf /hadoop-3.4.1/hadoop                                    0.3s
    [26/58] COPY apache-tez-0.10.4-bin.tar.gz /apache-tez-0.10.4-bin.tar.gz    0.2s
    [27/58] RUN tar -zxvf /apache-tez-0.10.4-bin.tar.gz -C /                   0.4s
    [28/58] RUN ln -sf /apache-tez-0.10.4/bin/tez                              0.4s
    [29/58] COPY apache-hive-4.0.1-bin.tar.gz /apache-hive-4.0.1-bin.tar.gz    0.4s
    [30/58] RUN tar -zxvf /apache-hive-4.0.1-bin.tar.gz -C /                   0.4s
    [31/58] RUN ln -sf /apache-hive-4.0.1/bin/hive                             0.4s
    [32/58] COPY hbase-2.5.11-bin.tar.gz /hbase-2.5.11-bin.tar.gz              0.4s
    [33/58] RUN tar -zxvf /hbase-2.5.11-bin.tar.gz -C /                         0.4s
    [34/58] RUN ln -sf /hbase-2.5.11/bin/hbase                                 0.3s
    [35/58] COPY apache-zookeeper-3.8.4-bin.tar.gz /apache-zookeeper-3.8.4-bin.tar.gz
    [36/58] RUN tar -zxvf /apache-zookeeper-3.8.4-bin.tar.gz -C /               0.1s
    [37/58] RUN ln -sf /apache-zookeeper-3.8.4/bin/zookeeper                   0.3s
    [38/58] COPY spark-3.5.5-bin-hadoop3.tgz /spark-3.5.5-bin-hadoop3.tgz      0.5s
    [39/58] RUN tar -zxvf /spark-3.5.5-bin-hadoop3.tgz -C /                   2.4s
    [40/58] RUN ln -sf /spark-3.5.5-bin-hadoop3/spark                          0.2s
    [41/58] RUN apt-get update && apt-get --y install mysql-server mysql-client
    [42/58] COPY mysql-connector-java-8.0.28.jar /hive/lib/                     0.1s
    [43/58] RUN apt-get --y clean all && rm -rf /tmp/* /var/tmp/*                0.3s
    [44/58] RUN apt-get install -y openssl-server                             0.2s
    [45/58] RUN apt-get --y clean all && rm -rf /tmp/* /var/tmp/*                0.3s
    [46/58] RUN apt-get --y clean all && rm -rf /tmp/* /var/tmp/*                0.5s
```

### 3. Jalankan container

```
C:\Users\sains\Documents\tubes-s\hadoop-spark-docker\bigdata-spark\bash start.sh
Secf3bb274c3f5aa442f03c8f7ece052815a203fa768f93c4b96f4db088bc
C:\Users\sains\Documents\tubes-s\hadoop-spark-docker\bigdata-spark
```

### 4. setelah masuk ke container tunggu beberapa menit untuk semua fungsi berjalan, untuk melihat history log nya

```
root@localhost:~# cat /tmp/bootstrap.log
* Starting MySQL database server mysqld
...done.
* Starting OpenBSD Secure Shell server sshd
...done.
* Starting yarn
Starting ResourceManager
Starting NodeManagers
* Starting namenode
Localhost: Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
Localhost: WARNING: YARN_OPTS has been replaced by HADOOP_OPTS. Using value of YARN_OPTS.
Starting namenodes on [0.0.0.0]
0.0.0.0: Warning: Permanently added '0.0.0.0' (ED25519) to the list of known hosts.
0.0.0.0: WARNING: HADOOP_NAMENODE_OPTS has been replaced by HDFS_NAMENODE_OPTS.
Starting datanodes
Starting secondary namenodes [localhost]
248 NodeManager
1536 SecondaryNameNode
1313 DataNode
1221 Jps
1144 NameNode
398 ResourceManager
* Starting up zookeeper...
* Start Hbase
Localhost: running zookeeper, logging to /hbase/logs/hbase-root-zookeeper-localhost.out
Localhost: slf4j: Class path contains multiple SLF4J bindings.
Localhost: slf4j: Found binding in [jar:file:/hbase-2.5.11/lib/client-facing-thirdparty/log4j-slf4j-impl-2.17.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
Localhost: slf4j: Found binding in [jar:file:/hadoop-3.4.1/share/hadoop/common/lib/slf4j-reloadj-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
Localhost: slf4j: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
Localhost: slf4j: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Running master, logging to /hbase/logs/hbase-master-localhost.out
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/hbase-2.5.11/lib/client-facing-thirdparty/log4j-slf4j-impl-2.17.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/hadoop-3.4.1/share/hadoop/common/lib/slf4j-reloadj-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/hadoop-3.4.1/share/hadoop/common/lib/slf4j-reloadj-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Running regionserver, logging to /hbase/logs/hbase-regionserver-localhost.out
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/hbase-2.5.11/lib/client-facing-thirdparty/log4j-slf4j-impl-2.17.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/hadoop-3.4.1/share/hadoop/common/lib/slf4j-reloadj-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/hadoop-3.4.1/share/hadoop/common/lib/slf4j-reloadj-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
* Start of metastore on
Initializing the schema to: 4.0.0
Metastore connection URL: jdbc:mysql://localhost:3306/metastore?createDatabaseIfNotExist=true
Metastore connection Driver: com.mysql.cj.jdbc.Driver
Metastore connection User: hive
Starting metastore schema initialization to 4.0.0
Initialization script hive-schema-4.0.0.mysql.sql
Initialization script completed
* Start HDFS server
* Start HDFS server to run up and running
0825-05-22 09:14:52: Starting Hive Metastore Server
* Start HiveServer2
root@localhost:~#
```

### 5. Cek aktivasi port yang berjalan

```
root@localhost:~# netstat -npt
Active Internet connections (only servers)
tcp        0      0 0.0.0.0:1313->0.0.0.0: LISTEN 1313/java
tcp        0      0 0.0.0.0:13306->0.0.0.0: LISTEN -
tcp        0      0 0.0.0.0:13306->0.0.0.0: LISTEN -
tcp        0      0 0.0.0.0:9870->0.0.0.0: LISTEN 1144/java
tcp        0      0 0.0.0.0:9868->0.0.0.0: LISTEN 1536/java
tcp        0      0 0.0.0.0:9867->0.0.0.0: LISTEN 1313/java
tcp        0      0 0.0.0.0:9866->0.0.0.0: LISTEN 1313/java
tcp        0      0 0.0.0.0:9864->0.0.0.0: LISTEN 1313/java
tcp        0      0 0.0.0.0:22->0.0.0.0: LISTEN 282/sshd :usr/sbin
tcp        0      0 0.0.0.0:9000->0.0.0.0: LISTEN 1144/java
tcp6       0      0 :::8042::: LISTEN 544/java
tcp6       0      0 :::8049::: LISTEN 544/java
tcp6       0      0 :::8033::: LISTEN 398/java
tcp6       0      0 :::8032::: LISTEN 398/java
tcp6       0      0 :::8031::: LISTEN 398/java
tcp6       0      0 :::8030::: LISTEN 398/java
tcp6       0      0 :::8008::: LISTEN 398/java
tcp6       0      0 :::8000::: LISTEN 2257/java
tcp6       0      0 :::16030::: LISTEN 2761/java
tcp6       0      0 :::16020::: LISTEN 2496/java
tcp6       0      0 0.0.0.0:16020->0.0.0.0: LISTEN 2761/java
tcp6       0      0 0.0.0.0:16000->0.0.0.0: LISTEN 2496/java
tcp6       0      0 :::2181::: LISTEN 2257/java
tcp6       0      0 :::10002::: LISTEN 4314/java
tcp6       0      0 :::10001::: LISTEN 4314/java
tcp6       0      0 :::10000::: LISTEN 3966/java
tcp6       0      0 :::22::: LISTEN 282/sshd :usr/sbin
tcp6       0      0 :::45061::: LISTEN 544/java
root@localhost:~#
```

## 6. Login spark

```
root@localhost:~# spark-shell
25/05/22 09:50:40 WARN Utils: Your hostname, localhost resolves to a loopback address: 127.0.0.1; using 172.17.0.3 instead (on interface eth0)
25/05/22 09:50:40 WARN Utils: Set SPARK_LOCAL_ID if you need to bind to another address
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
25/05/22 09:50:43 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Spark context web UI available at http://localhost:4040
Spark context available as 'sc' (master = local[*], app id = local-1747987444189).
Spark session available as 'spark'.
Welcome to

      ____
     / ___/
    / __/   version 3.5.5
   /___/

Using Scala version 2.12.18 (OpenJDK 64-Bit Server VM, Java 1.8.0_452)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

## 7. Akses spark UI

localhost:4040/jobs/

Spark 3.5.5

JobsStagesStorageEnvironmentExecutors

Spark shell application UI

### Spark Jobs (?)

User: root  
Total Uptime: 53 s  
Scheduling Mode: FIFO

Event Timeline

Enable zooming

| Executors | Added     | Removed |         |
|-----------|-----------|---------|---------|
| Jobs      | Succeeded | Failed  | Running |

|  | 00       | 800 | 000      | 200 | 400 | 600 | 800 | 000      | 200 | 400 | 600 | 800 | 000      | 200 | 400 | 600 |
|--|----------|-----|----------|-----|-----|-----|-----|----------|-----|-----|-----|-----|----------|-----|-----|-----|
|  | 09:11:35 |     | 09:11:36 |     |     |     |     | 09:11:37 |     |     |     |     | 09:11:38 |     |     |     |