**MAVEN REPO :** <https://mvnrepository.com/>

**HADOOP** - Noted of configuration to test locally

- reference : <https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz>

- jdk 1.8 : <https://enos.itcollege.ee/~jpoial/allalaadimised/jdk8/>

- winutils : <https://github.com/cdarlint/winutils>

copy and replace this to : \hadoop-3.2.1\share\hadoop\hdfs

- fix jar hadoop 3.2.1 : <https://github.com/FahaoTang/big-data/blob/master/hadoop-hdfs-3.2.1.jar>

* Namenode web : <http://localhost:9870/dfshealth.html#tab-overview>
* Resource Manager web : <http://localhost:8088/cluster>

| **file** | **core-site.xml** |
| --- | --- |
| <configuration>  <property>  <name>fs.defaultFS</name>  <value>hdfs://localhost:9000</value>  </property>  </configuration> | |

| **file** | **mapred-site.xml** |
| --- | --- |
| <configuration>  <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  <property>  <name>mapreduce.application.classpath</name>  <value>%HADOOP\_HOME%/share/hadoop/mapreduce/\*,%HADOOP\_HOME%/share/hadoop/mapreduce/lib/\*,%HADOOP\_HOME%/share/hadoop/common/\*,%HADOOP\_HOME%/share/hadoop/common/lib/\*,%HADOOP\_HOME%/share/hadoop/yarn/\*,%HADOOP\_HOME%/share/hadoop/yarn/lib/\*,%HADOOP\_HOME%/share/hadoop/hdfs/\*,%HADOOP\_HOME%/share/hadoop/hdfs/lib/\*</value>  </property>  </configuration> | |

| **file** | **yarn-site.xml** |
| --- | --- |
| <configuration>  <property>  <name>yarn.resourcemanager.hostname</name>  <value>localhost</value>  </property>  <property>  <name>yarn.nodemanager.aux-services</name>  <value>mapreduce\_shuffle</value>  </property>  <property>  <name>yarn.nodemanager.env-whitelist</name>  <value>JAVA\_HOME,HADOOP\_COMMON\_HOME,HADOOP\_HDFS\_HOME,HADOOP\_CONF\_DIR,CLASSPATH\_PREPEND\_DISTCACHE,HADOOP\_YARN\_HOME,HADOOP\_MAPRED\_HOME</value>  </property>  </configuration> | |

| **file** | **hdfs-site.xml** |
| --- | --- |
| <configuration>  <property>  <name>dfs.replication</name>  <value>1</value>  </property>  <property>  <name>dfs.namenode.name.dir</name>  <value>/C:/hadoop/hadoop-3.2.1/data/namenode</value>  </property>  <property>  <name>dfs.datanode.name.dir</name>  <value>/C:/hadoop/hadoop-3.2.1/data/datanode</value>  </property>  </configuration> | |

CONFIG TO TRY **LOCAL DOCKER** BY OFFICIAL

| **file** | **config** |
| --- | --- |
| CORE-SITE.XML\_fs.default.name=hdfs://namenode  CORE-SITE.XML\_fs.defaultFS=hdfs://namenode  HDFS-SITE.XML\_dfs.namenode.rpc-address=namenode:8020  HDFS-SITE.XML\_dfs.replication=1  HDFS-SITE.XML\_dfs.permissions.enabled=false  HDFS-SITE.XML\_dfs.datanode.data.dir=/tmp/hadoop-hadoop/dfs/data  MAPRED-SITE.XML\_mapreduce.framework.name=yarn  MAPRED-SITE.XML\_yarn.app.mapreduce.am.env=HADOOP\_MAPRED\_HOME=$HADOOP\_HOME  MAPRED-SITE.XML\_mapreduce.map.env=HADOOP\_MAPRED\_HOME=$HADOOP\_HOME  MAPRED-SITE.XML\_mapreduce.reduce.env=HADOOP\_MAPRED\_HOME=$HADOOP\_HOME  YARN-SITE.XML\_yarn.resourcemanager.hostname=resourcemanager  YARN-SITE.XML\_yarn.nodemanager.pmem-check-enabled=false  YARN-SITE.XML\_yarn.nodemanager.delete.debug-delay-sec=600  YARN-SITE.XML\_yarn.nodemanager.vmem-check-enabled=false  YARN-SITE.XML\_yarn.nodemanager.aux-services=mapreduce\_shuffle  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.maximum-applications=10000  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.maximum-am-resource-percent=0.1  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.resource-calculator=org.apache.hadoop.yarn.util.resource.DefaultResourceCalculator  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.root.queues=default  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.root.default.capacity=100  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.root.default.user-limit-factor=1  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.root.default.maximum-capacity=100  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.root.default.state=RUNNING  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.root.default.acl\_submit\_applications=\*  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.root.default.acl\_administer\_queue=\*  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.node-locality-delay=40  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.queue-mappings=  CAPACITY-SCHEDULER.XML\_yarn.scheduler.capacity.queue-mappings-override.enable=false | |

| **file** | **docker-compose.yaml** |
| --- | --- |
| version: "2"  services:  namenode:  image: apache/hadoop:3  hostname: namenode  command: ["hdfs", "namenode"]  ports:  - 9870:9870  - 8020:8020  env\_file:  - ./config  environment:  ENSURE\_NAMENODE\_DIR: "/tmp/hadoop-root/dfs/name"  datanode:  image: apache/hadoop:3  command: ["hdfs", "datanode"]  env\_file:  - ./config  ports:  - 50010:50010  - 9866:9866  - 9864:9864  resourcemanager:  image: apache/hadoop:3  hostname: resourcemanager  command: ["yarn", "resourcemanager"]  ports:  - 8088:8088  env\_file:  - ./config  volumes:  - ./test.sh:/opt/test.sh  nodemanager:  image: apache/hadoop:3  command: ["yarn", "nodemanager"]  env\_file:  - ./config | |