

**My System: Windows 7**

**Hadoop Version: 2.7.3**

**Java version: 1.8.0\_221**

**Hive version: 2.1.0**

**Derby version: 10.12.1.1**

**ID: WQD180093**

**NAME: CHONG MING KEAT**

**Step 1:** Verifying JAVA Installation. Type java -version in cmd to check installed java version. To install java, go to this link to download, <https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>.

```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\User>java -version
The system cannot find the file C:\ProgramData\Oracle\Java\javapath\java.exe.

C:\Users\User>cd C:\Java\jdk1.8.0_221\bin









C:\Java\jdk1.8.0_221\bin>java -version
java version "1.8.0_221"
Java(TM) SE Runtime Environment (build 1.8.0_221-b11)
Java HotSpot(TM) 64-Bit Server VM (build 25.221-b11, mixed mode)

C:\Java\jdk1.8.0_221\bin>
```

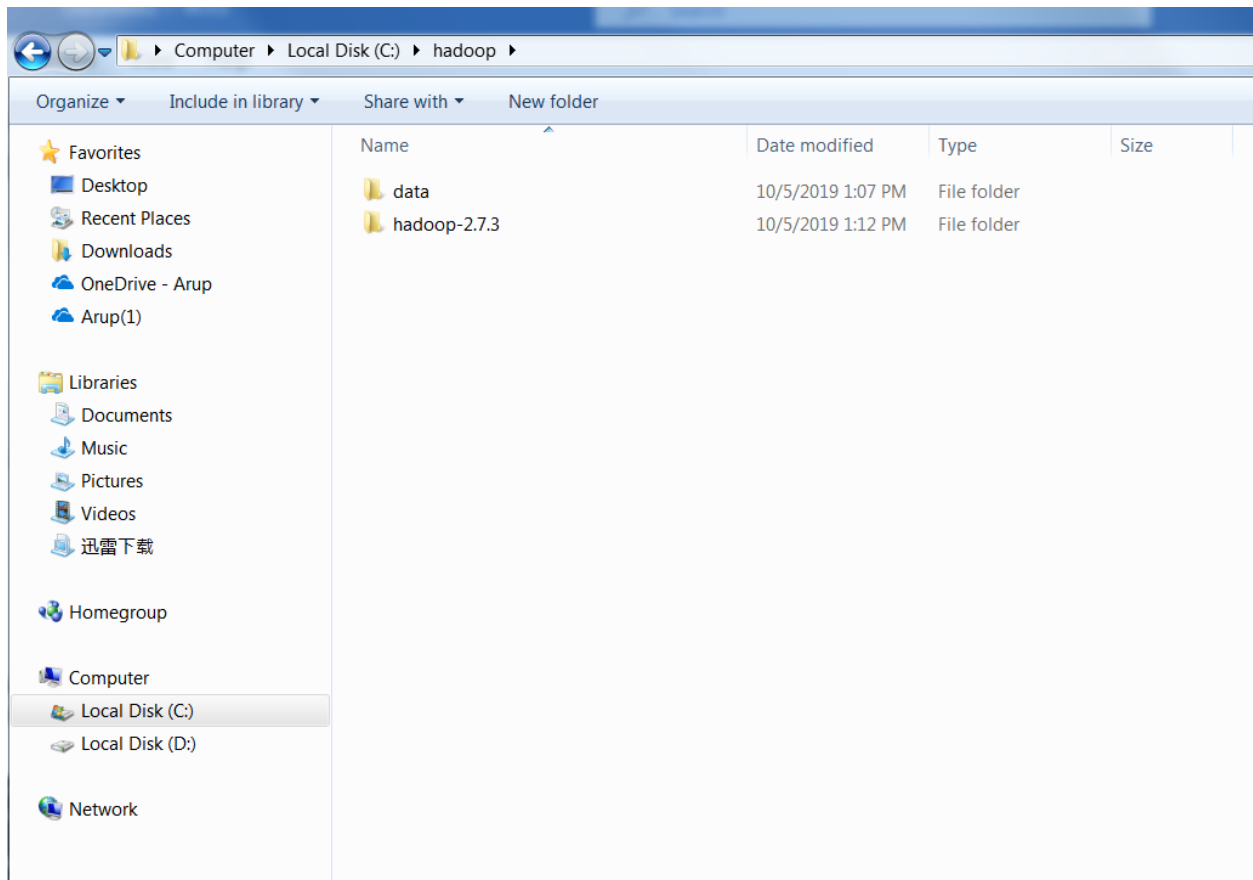
**Step 2:** Verifying Hadoop Installation

**2.1** Go to this link and download Hadoop, <https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/>

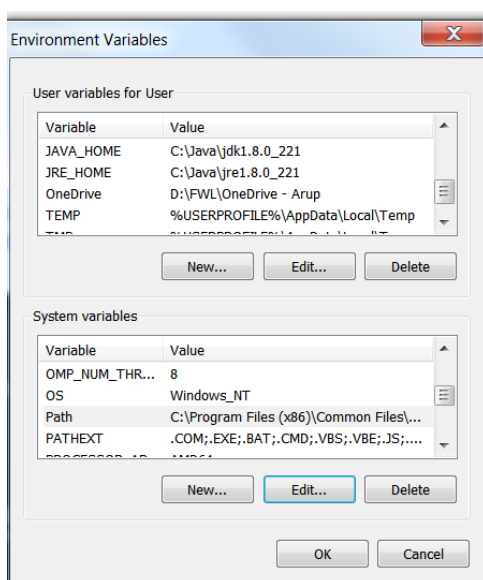
## Index of /dist/hadoop/core/hadoop-2.7.3

Name	Last modified	Size	Description
 <a href="#">Parent Directory</a>		-	
 <a href="#">hadoop-2.7.3-src.tar.gz</a>	2016-08-25 19:25	17M	
 <a href="#">hadoop-2.7.3-src.tar.gz.asc</a>	2016-08-25 19:25	521	
 <a href="#">hadoop-2.7.3-src.tar.gz.mds</a>	2016-08-25 19:25	1.1K	
 <a href="#">hadoop-2.7.3.tar.gz</a>	2016-08-25 19:25	204M	
 <a href="#">hadoop-2.7.3.tar.gz.asc</a>	2016-08-25 19:25	521	
 <a href="#">hadoop-2.7.3.tar.gz.mds</a>	2016-08-25 19:25	958	

## 2.2 Unzip the file and save inside your drive.



## 2.3 Set JAVA\_HOME, HADOOP\_HOME and append the path of bin for java and Hadoop at system variables.

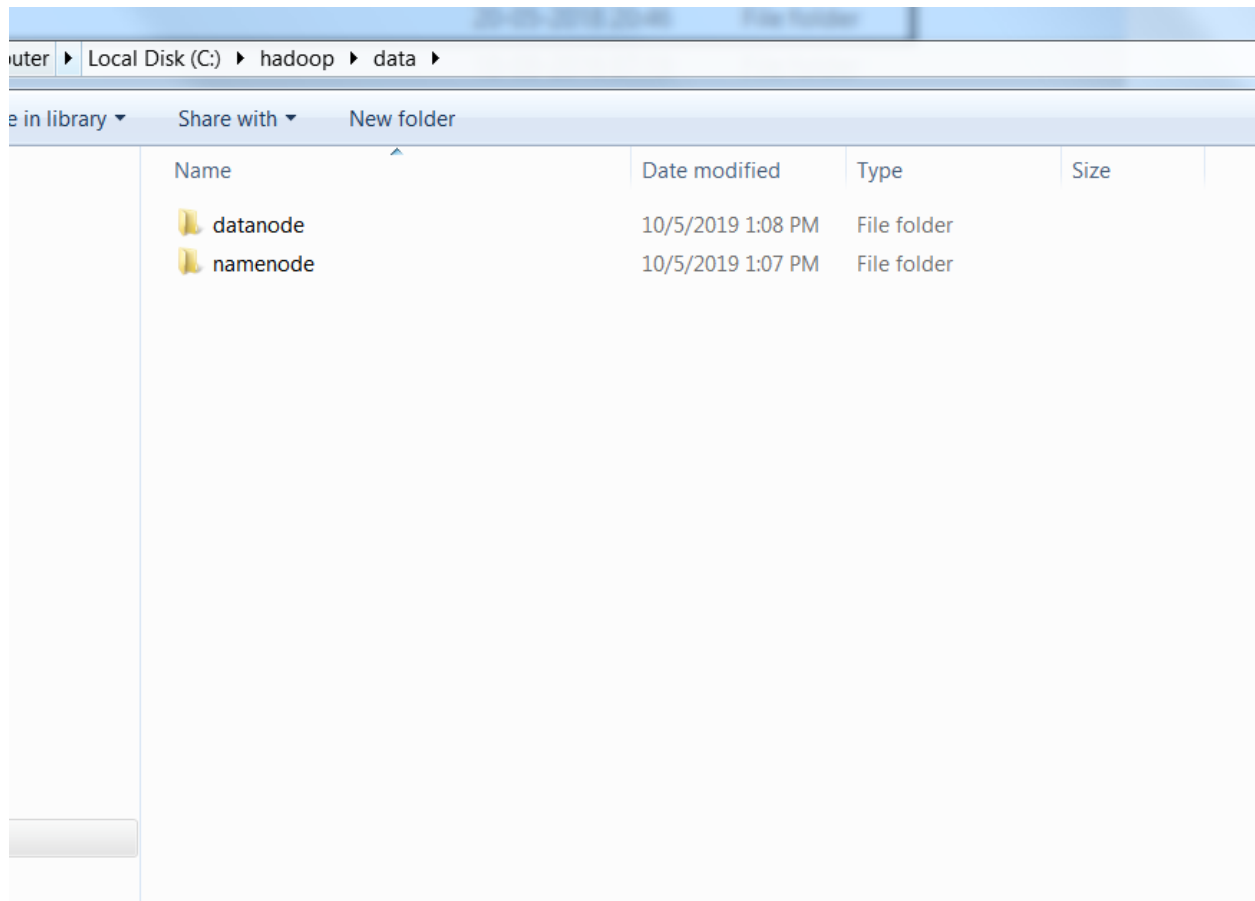


## 2.4 configure Hadoop files, Hadoop-env.cmd, core-site.xml, Mapred-site.xml, Hdfs-site.xml and Yarn-site.xml

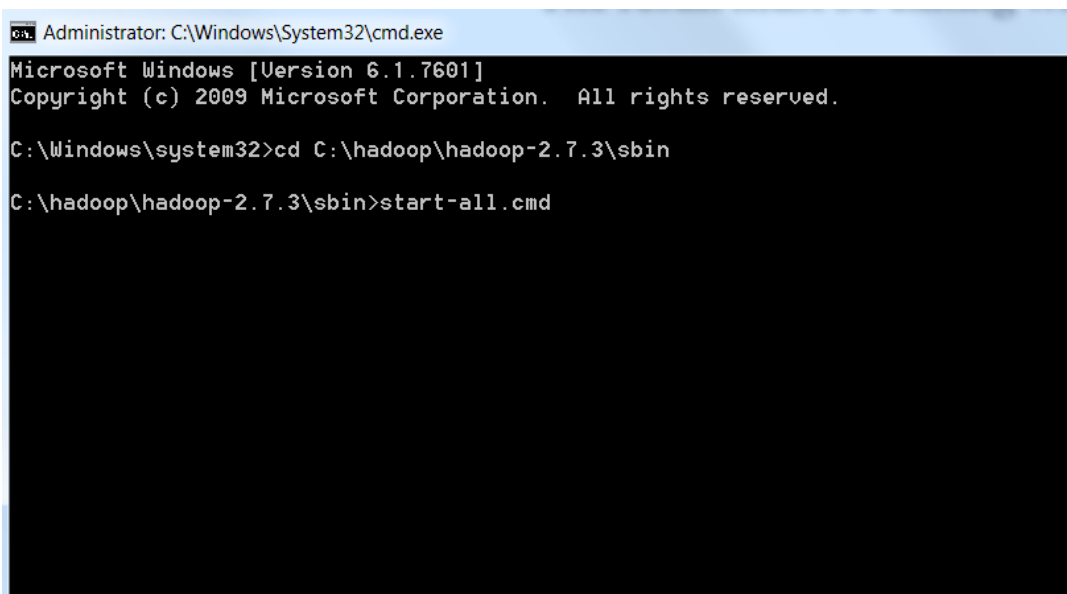
his PC > OS (C:) > hadoop > hadoop-2.7.3 > etc > hadoop

Name	Date modified	Type	Size
core-site.xml	17-05-2018 14:07	XML Document	2 KB
hadoop-env.cmd	17-05-2018 13:57	Windows Comma...	4 KB
hadoop-env.sh	17-05-2018 13:12	Shell Script	5 KB
hadoop-metrics.properties	17-05-2018 13:12	PROPERTIES File	3 KB
hadoop-metrics2.properties	17-05-2018 13:12	PROPERTIES File	3 KB
hadoop-policy.xml	17-05-2018 13:12	XML Document	10 KB
hdfs-site.xml	20-05-2018 21:03	XML Document	2 KB
httpfs-env.sh	17-05-2018 13:12	Shell Script	2 KB
httpfs-log4j.properties	17-05-2018 13:12	PROPERTIES File	2 KB
httpfs-signature.secret	17-05-2018 13:12	SECRET File	1 KB
httpfs-site.xml	17-05-2018 13:12	XML Document	1 KB
kms-acls.xml	17-05-2018 13:12	XML Document	4 KB
kms-env.sh	17-05-2018 13:12	Shell Script	2 KB
kms-log4j.properties	17-05-2018 13:12	PROPERTIES File	2 KB
kms-site.xml	17-05-2018 13:12	XML Document	6 KB
log4j.properties	17-05-2018 13:12	PROPERTIES File	11 KB
mapred-env.cmd	17-05-2018 13:12	Windows Comma...	1 KB
mapred-env.sh	17-05-2018 13:12	Shell Script	2 KB
mapred-queues.xml.template	17-05-2018 13:12	Dev-C++ Templat...	5 KB
mapred-site.xml	17-05-2018 13:12	XML Document	1 KB
mapred-site.xml.template	20-05-2018 21:33	Dev-C++ Templat...	2 KB
slaves	17-05-2018 13:12	File	1 KB
ssl-client.xml.example	17-05-2018 13:12	EXAMPLE File	3 KB
ssl-server.xml.example	17-05-2018 13:12	EXAMPLE File	3 KB
yarn-env.cmd	17-05-2018 13:12	Windows Comma...	3 KB
yarn-env.sh	17-05-2018 13:12	Shell Script	5 KB
yarn-site.xml	20-05-2018 21:28	XML Document	2 KB

## 2.5 Create two folders by the name Namenode and Datanode



## 2.6 Go to cmd and cd to Hadoop/sbin and type start-all.cmd



4 extra command prompts will pop out

```
Apache Hadoop Distribution - hadoop - namenode
19/10/05 23:02:30 INFO namenode.LeaseManager: Number of blocks under construction: 0
19/10/05 23:02:30 INFO hdfs.StateChange: STATE= Safe mode ON
The reported blocks 0 needs additional 10 blocks to reach the threshold 0.9990 of total blocks 10.
The number of live datanodes 0 has reached the minimum number 0. Safe mode will be turned off automatically once the thresholds have been reached.
19/10/05 23:02:30 INFO blockmanagement.DatanodeDescriptor: Number of failed storage changes from 0 to 0
19/10/05 23:02:30 INFO ipc.Server: IPC Server Responder: starting
19/10/05 23:02:30 INFO ipc.Server: IPC Server listener on 19000: starting
19/10/05 23:02:30 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:19000
19/10/05 23:02:30 INFO namenode.FSNamesystem: Starting services required for active state
19/10/05 23:02:30 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds
19/10/05 23:02:30 INFO hdfs.StateChange: BLOCK= registerDatanode: from DatanodeRegistration(127.0.0.1:50010, datanodeUuid=46b04633-0f12-454d-83b1-51a648a37ca0, infoPort=50075, infoSecurePort=0, ipcPort=50020, storageInfo=156-c1d5-rtn-4d246ba-724a-4a3a-b325-8b369031f1a1, ssid=1f85393525-c10), storage: 46b04633-0f12-454d-83b1-51a648a37ca0

Apache Hadoop Distribution - yarn - nodemanager
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Oct 05, 2019 11:02:28 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.9.05/02/2011 11:17 AM'
Oct 05, 2019 11:02:28 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
Oct 05, 2019 11:02:28 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider

Apache Hadoop Distribution - yarn - resourcemanager
19/10/05 23:02:26 INFO ipc.CallQueueManager: Using callQueue class java.util.concurrent.LinkedBlockingQueue
19/10/05 23:02:26 INFO ipc.Server: Starting Socket Reader #1 for port 8031
19/10/05 23:02:26 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the server
19/10/05 23:02:26 INFO ipc.Server: IPC Server Responder: starting
19/10/05 23:02:26 INFO ipc.Server: IPC Server listener on 8031: starting
19/10/05 23:02:26 INFO ipc.CallQueueManager: Using callQueue class java.util.concurrent.LinkedBlockingQueue
19/10/05 23:02:26 INFO ipc.Server: Starting Socket Reader #1 for port 8030
19/10/05 23:02:26 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ApplicationMasterProtocolPB to the server

Apache Hadoop Distribution - hadoop - datanode
STARTUP_MSG: java = 1.8.0_221
19/10/05 23:02:25 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2.properties
19/10/05 23:02:25 INFO impl.MetricsSystemImpl: Scheduled snapshot period at 10 second(s).
19/10/05 23:02:25 INFO impl.MetricsSystemImpl: DataNode metrics system started
19/10/05 23:02:25 INFO datanode.BlockScanner: Initialized block scanner with targetBytesPerSec 1048576
19/10/05 23:02:25 INFO datanode.DataNode: Configured hostname is User-PC
19/10/05 23:02:25 INFO datanode.DataNode: Starting DataNode with maxLockedMemory = 0
19/10/05 23:02:25 INFO datanode.DataNode: Opened streaming server at /0.0.0.0:50010
19/10/05 23:02:25 INFO datanode.DataNode: Balancing bandwidth is 1048576 bytes/s
19/10/05 23:02:25 INFO datanode.DataNode: Number threads for balancing is 5
19/10/05 23:02:26 INFO morthay.log: Logging to org.slf4j.impl.Log4jLoggerAdapter(org.morthay.log) via org.morthay.log.Slf4jLog
19/10/05 23:02:26 INFO server.AuthenticationFilter: Unable to initialize FileSignerSecretProvider, falling back to use random secrets.
19/10/05 23:02:26 INFO http.HttpRequestLog: Http request log for http requests.datanode is not defined
19/10/05 23:02:26 INFO http.HttpServer2: Added global filter 'safety' (class=org.apache.hadoop.http.HttpServer2$QuotingInputFilter)
19/10/05 23:02:26 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context datanode
19/10/05 23:02:26 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context static
19/10/05 23:02:26 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context logs
19/10/05 23:02:26 INFO morthay.log: Jetty bound to port 50135
19/10/05 23:02:26 INFO morthay.log: jetty-S.1.26
19/10/05 23:02:26 INFO morthay.log: Started HttpServer2$SelectChannelConnectorWithSafeStartup@localhost:50135
19/10/05 23:02:27 INFO web.DatanodeHttpServer: Listening HTTP traffic on /0.0.0.0:50075
19/10/05 23:02:27 INFO datanode.DataNode: dnUserName = User
```

2.7 Verify the hive is working via with browser:

- Namenode (hdfs) - http://localhost:50070
- Datanode - http://localhost:50075
- All Applications (cluster) - http://localhost:8088










## All Applications

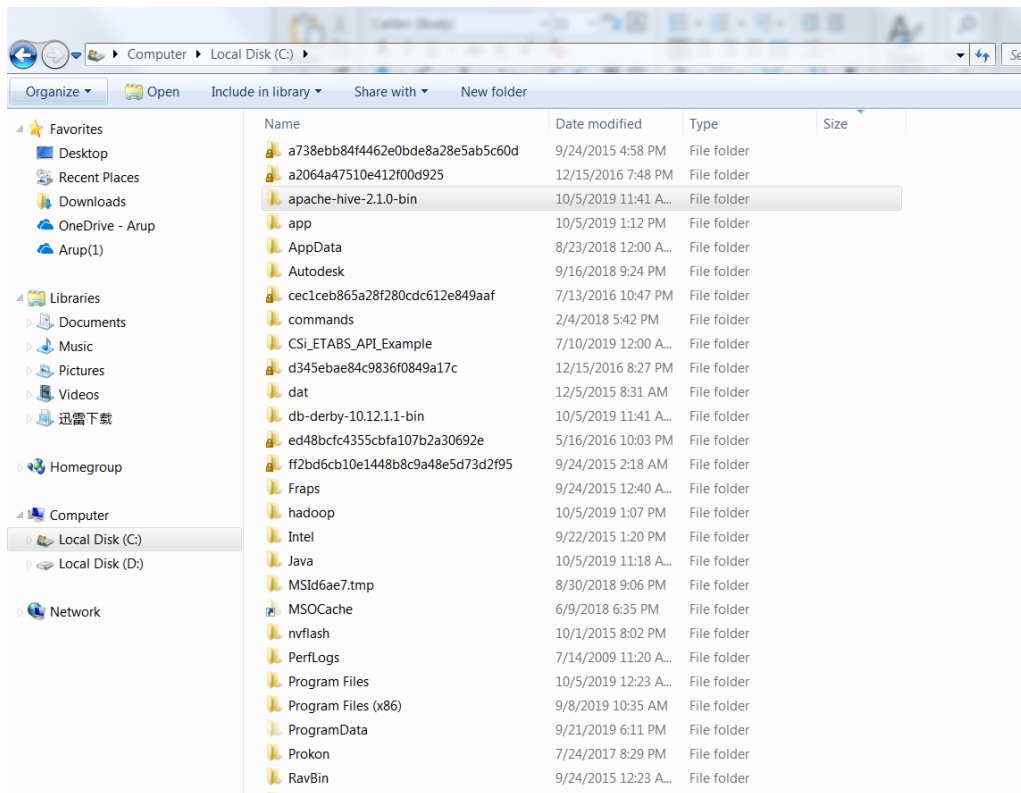
## Step 4: Installing Hive

4.1 Download Hive file at <https://archive.apache.org/dist/hive/hive-2.1.0/>

# Index of /dist/hive/hive-2.1.0

Name	Last modified	Size	Description
 <a href="#">Parent Directory</a>		-	
 <a href="#">apache-hive-2.1.0-bin.tar.gz</a>	2016-06-21 01:26	143M	←
 <a href="#">apache-hive-2.1.0-bin.tar.gz.asc</a>	2016-06-21 01:26	819	
 <a href="#">apache-hive-2.1.0-bin.tar.gz.md5</a>	2016-06-21 01:26	70	
 <a href="#">apache-hive-2.1.0-src.tar.gz</a>	2016-06-21 01:26	18M	
 <a href="#">apache-hive-2.1.0-src.tar.gz.asc</a>	2016-06-21 01:26	819	
 <a href="#">apache-hive-2.1.0-src.tar.gz.md5</a>	2016-06-21 01:26	70	

4.2 Unzip the file and put it at your drive



## Step 5: Configuring Hive

Go to Hive->conf and configure hive-site.xml file for derby

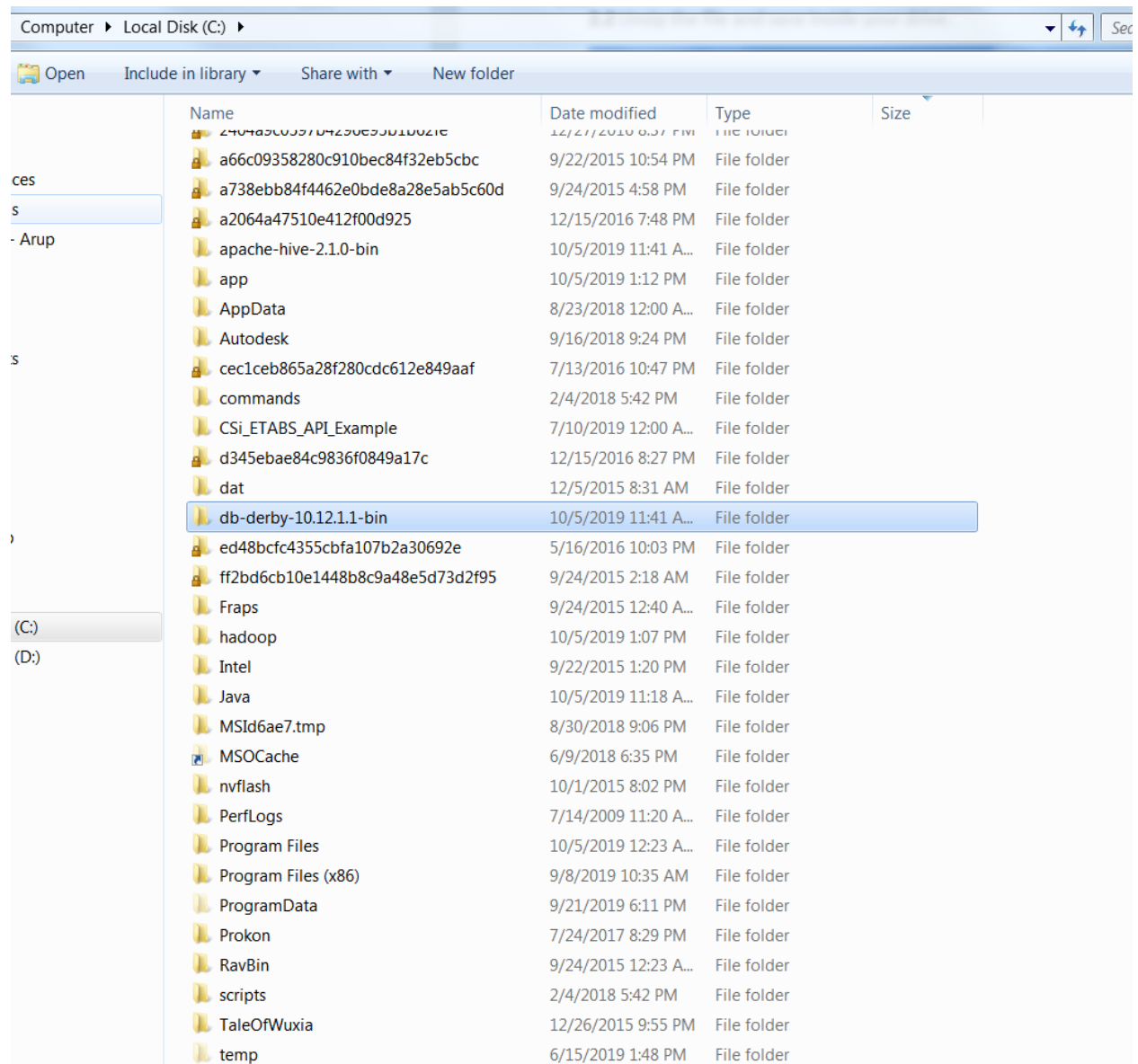
```
1 <configuration>
2   <property>
3     <name>javax.jdo.option.ConnectionURL</name>
4     <value>jdbc:derby://localhost:1527/metastore_db;create=true</value>
5     <description>JDBC connect string for a JDBC metastore</description>
6   </property>
7   <property>
8     <name>javax.jdo.option.ConnectionDriverName</name>
9     <value>org.apache.derby.jdbc.ClientDriver</value>
10    <description>Driver class name for a JDBC metastore</description>
11  </property>
12  <property>
13    <name>hive.server2.enable.impersonation</name>
14    <description>Enable user impersonation for HiveServer2</description>
15    <value>true</value>
16  </property>
17  <property>
18    <name>hive.server2.authentication</name>
19    <value>NONE</value>
20    <description>Client authentication types. NONE: no authentication check LDAP: LDAP/AD based authentication KERBEROS: Kerberos/GSSAPI authentication CUSTOM: Custom authentication provider (Use with property
hive.server2.custom.authentication.class) </description>
21  </property>
22  <property>
23    <name>datanucleus.autoCreateTables</name>
24    <value>True</value>
25  </property>
26 </configuration>
```

## Step 6: Downloading and Installing Apache Derby

Go to the link <https://archive.apache.org/dist/db/derby/db-derby-10.12.1.1/> download the .tar file

Name	Last modified	Size	Description
 Parent Directory		-	
 db-derby-10.12.1.1-bin.tar.gz	2015-10-10 14:38	18M	←
 db-derby-10.12.1.1-bin.tar.gz.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-bin.tar.gz.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-bin.zip	2015-10-10 14:38	20M	
 db-derby-10.12.1.1-bin.zip.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-bin.zip.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-lib-debug.tar.gz	2015-10-10 14:38	14M	
 db-derby-10.12.1.1-lib-debug.tar.gz.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-lib-debug.tar.gz.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-lib-debug.zip	2015-10-10 14:38	14M	
 db-derby-10.12.1.1-lib-debug.zip.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-lib-debug.zip.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-lib.tar.gz	2015-10-10 14:38	12M	
 db-derby-10.12.1.1-lib.tar.gz.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-lib.tar.gz.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-lib.zip	2015-10-10 14:38	12M	
 db-derby-10.12.1.1-lib.zip.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-lib.zip.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-src.tar.gz	2015-10-10 14:38	27M	
 db-derby-10.12.1.1-src.tar.gz.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-src.tar.gz.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-src.zip	2015-10-10 14:38	33M	
 db-derby-10.12.1.1-src.zip.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-src.zip.md5	2015-10-10 14:38	33	

## 6.1. unzip the file and put It at your drive



**6.2** go into C:\db-derby-10.12.1.1-bin\lib and copy all the file and paste it into C:\apache-hive-2.1.0-bin\lib



**6.3** Go into command prompt and type `startNetworkServer -h 0.0.0.0` to start derby server.

```
Administrator: C:\Windows\System32\cmd.exe - startNetworkServer -h 0.0.0.0
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Windows\system32>cd C:\db-derby-10.12.1.1-bin

C:\db-derby-10.12.1.1-bin>cd bin

C:\db-derby-10.12.1.1-bin>startNetworkServer -h 0.0.0.0
Sat Oct 05 23:28:57 SGT 2019 : Security manager installed using the Basic server security policy.
Sat Oct 05 23:28:57 SGT 2019 : Apache Derby Network Server - 10.12.1.1 - (1704137) started and ready to accept connections on port 1527
```

## Step 7: Configuring Metastore of Hive

**7.1** Go into command prompt and `cd` to `C:\apache-hive-2.1.0-bin\bin` and type `jps -m` to check `NetworkServerControl`.

```
Administrator: C:\Windows\System32\cmd.exe
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Windows\system32>cd C:\apache-hive-2.1.0-bin

C:\apache-hive-2.1.0-bin>cd bin

C:\apache-hive-2.1.0-bin\bin>jos -m
'jos' is not recognized as an internal or external command,
operable program or batch file.

C:\apache-hive-2.1.0-bin\bin>jps -m
8784 DataNode
9072 NameNode
9412 NodeManager
12440 NetworkServerControl start -h 0.0.0.0
3112 ResourceManager
9480 Jps -m
```

**7.2** type `hive` to execute hive server.

```
C:\apache-hive-2.1.0-bin\bin>hive
ERROR StatusLogger No log4j2 configuration file found. Using default configuration: logging only errors to the console.
Connecting to jdbc:hive2://
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/hadoop/hadoop-2.7.3/share/hadoop/common/lib/slf4j-log4j12-1.7.10.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]
Connected to: Apache Hive (version 2.1.0)
Driver: Hive JDBC (version 2.1.0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 2.1.0 by Apache Hive
hive>
```

## Step 8: Verifying Hive Installation

Try to create database and show table in Hive

```
hive> SHOW DATABASES;
OK
default
test
training
3 rows selected (0.441 seconds)
hive> show tables;
OK
testhive
1 row selected (0.109 seconds)
hive>
```

Try insert value into testhive and show values and check the submit application at cluster

```
hive> INSERT INTO testhive (col1,col2) VALUES('xxxadada','adawdqweq');
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = User_20191005233753_0cb4ba8e-0e44-4ee2-9c85-5a0e39e89005
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1570287746286_0001, Tracking URL = http://User-PC:8088/proxy/application_1570287746286_0001/
Kill Command = C:\hadoop\hadoop-2.7.3\bin\hadoop.cmd job -kill job_1570287746286_0001
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2019-10-05 23:38:02.837 Stage-1 map = 0%, reduce = 0%
2019-10-05 23:38:08.010 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.574 sec
MapReduce Total cumulative CPU time: 1 seconds 574 msec
Ended Job = job_1570287746286_0001
Stage-4 is selected by condition resolver
Stage-3 is filtered out by condition resolver
Stage-5 is filtered out by condition resolver
Moving data to directory hdfs://localhost:19000/user/hive/warehouse/testhive/.hive-staging_hive_2019-10-05_23-37-53_753_7186381147795138655-1/~ext-10000
Loading data to table default.testhive
MapReduce Jobs Launched:
Stage=Stage-1: Map: 1 Cumulative CPU: 1.574 sec HDFS Read: 4308 HDFS Write: 91 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 574 msec
OK
No rows affected (15.691 seconds)
hive> select * from testhive;
OK
adadada adad
xxxadada adawdqweq
2 rows selected (0.098 seconds)
hive>
```



### All Applications

Logged in as: dr.who

Cluster		Cluster Metrics															
About Nodes Node Labels Applications	Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	V-Cores Used	V-Cores Total	V-Cores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	
	1	0	0	1	0	0 B	9.77 GB	0 B	0	8	0	1	0	0	0	0	
Scheduler Metrics																	
NEW SCHEDULER				Scheduling Resource Type				Minimum Allocation				Maximum Allocation					
Capacity Scheduler				[MEMORY]				<memory:500, vCores:1>				<memory:8192, vCores:8>					
Show 20 entries																	
ID	User	Name	Application Type	Queue	Start Time	Finish Time	State	Final Status	Progress	Tracking UI	Backlisted Nodes						
application_1570287746286_0001	User	INSERT INTO testhive .xadada.'adawdqweq' (Stage-1)	MAPREDUCE	default	Sat Oct 5 23:37:56 +0800 2019	Sat Oct 5 23:38:07 +0800 2019	FINISHED	SUCCEEDED		History	N/A						
Showing 1 to 1 of 1 entries																	
First Previous 1 Next Last																	
Tools																	