DEEPCLOUDLABS

DEEP CLOUD LABS BİLİŞİM TEKNOLOJİLERİ
EĞİTİM VE DANIŞMANLIK TİC. LTD. ŞTİ.

www.deepcloudlabs.com

HOW TO SETUP HIVE 3.1.2 ON WINDOWS 10

1. Download the Hive binary from the following url:
   https://ftp.itu.edu.tr/Mirror/Apache/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz
2. Extract the **apache-hive-3.1.2-bin.tar.gz** to **C:\opt**
3. Download the Apache Derby binary from the following url:
   https://kozyatagi.mirror.guzel.net.tr/apache/db/derby/db-derby-10.14.2.0/db-derby-10.14.2.0-bin.zip
4. Extract the **db-derby-10.14.2.0-bin.zip** to **C:\opt**
5. Edit the following Hive configuration file with the following given content:
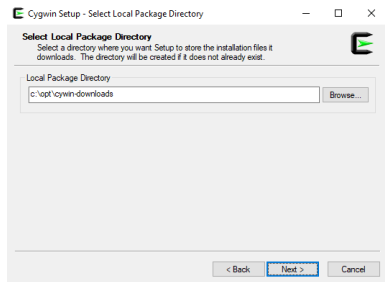
---

hive-site.xml

```xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
        <property>
                <name>javax.jdo.option.ConnectionURL</name>
                <value>jdbc:derby://localhost:1527/metastore_db;create=true</value>
                <description>JDBC connect string for a JDBC metastore</description>
        </property>
        <property>
                <name>javax.jdo.option.ConnectionDriverName</name>
                <value>org.apache.derby.jdbc.ClientDriver</value>
                <description>Driver class name for a JDBC metastore</description>
        </property>
        <property>
                <name>hive.server2.enable.doAs</name>
                <description>Enable user impersonation for HiveServer2</description>
                <value>true</value>
        </property>
        <property>
                <name>hive.server2.authentication</name>
                <value>NONE</value>
        </property>
        <property>
                <name>datanucleus.autoCreateTables</name>
                <value>True</value>
        </property>
</configuration>
```
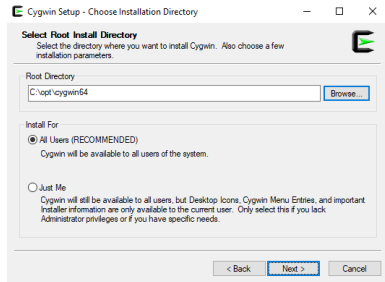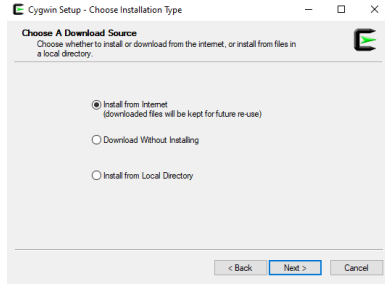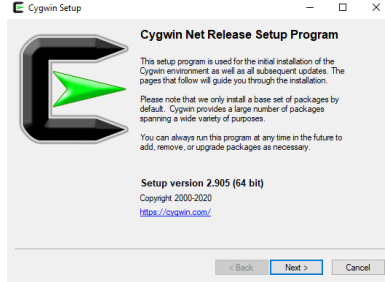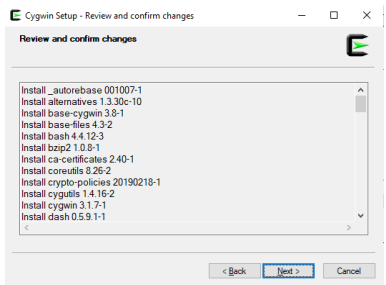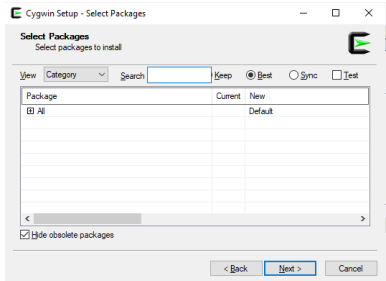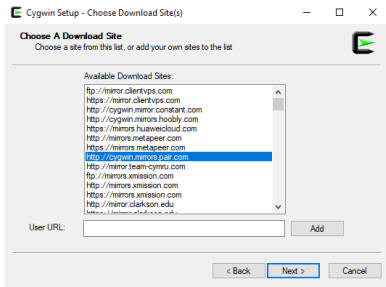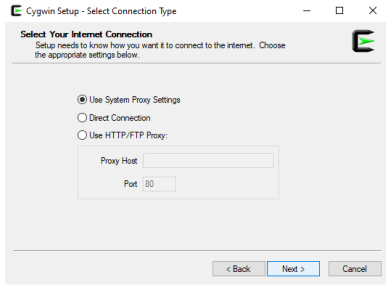

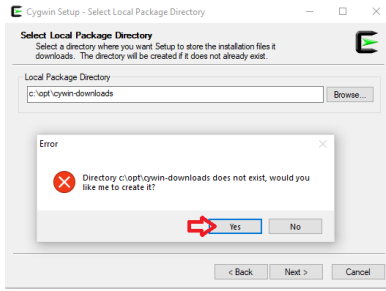6. Copy all the jar files from `c:\opt\db-derby-10.14.2.0\lib` to `d:\opt\apache-hive-3.1.2\lib`
7. Remove the file `c:\opt\apache-hive-3.1.2\lib\guava-19.0.jar`
8. Copy the file `c:\opt\hadoop-3.2.2\share\hadoop\common\lib\guava-27.0-jre.jar` to the folder `c:\opt\apache-hive-3.1.2\lib`
9. Run the following two commands on the windows command line
   ```
   mkdir c:\cygdrive
   mklink /J c:\cygdrive\c\ c:\
   ```

DEEPCLOUDLABS

DEEP CLOUD LABS BİLİŞİM TEKNOLOJİLERİ
EĞİTİM VE DANIŞMANLIK TİC. LTD. ŞTİ.

www.deepcloudlabs.com
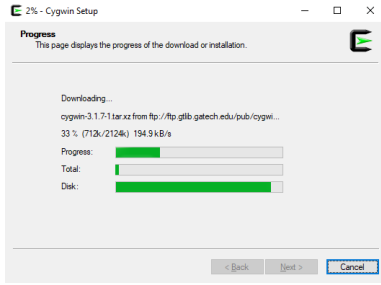
10. Start the Apache Derby server
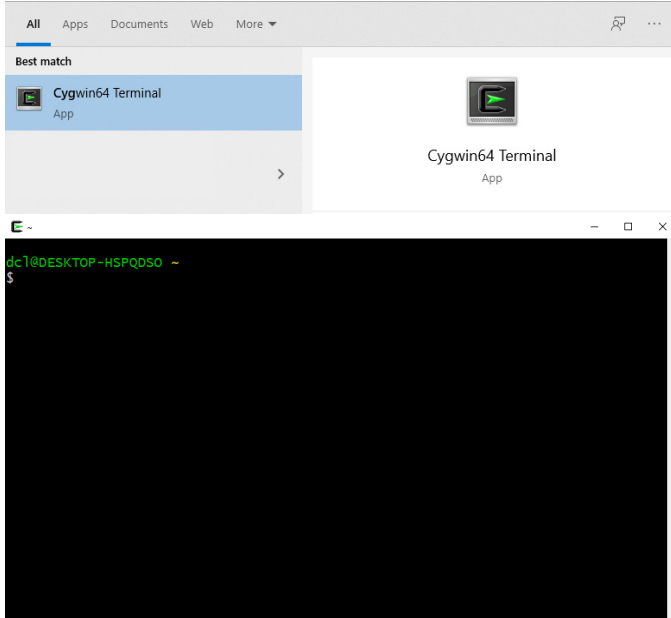```
cd c:\opt\db-derby-10.14.2.0
start bin\startNetworkServer -h 0.0.0.0
```

11. Install Cygwin (https://www.cygwin.com) using the following installer:
https://www.cygwin.com/setup-x86_64.exe

**DEEPCLOUDLABS**

DEEP CLOUD LABS BİLİŞİM TEKNOLOJİLERİ
EĞİTİM VE DANIŞMANLIK TİC. LTD. ŞTİ.

www.deepcloudlabs.com

**DEEPCLOUDLABS**

DEEP CLOUD LABS BİLİŞİM TEKNOLOJİLERİ
EĞİTİM VE DANIŞMANLIK TİC. LTD. ŞTİ.

www.deepcloudlabs.com

12. Run Cygwin Terminal



13. Set the following environment variables at the Cygwin Terminal:
    ```
    export HADOOP_HOME='/cygdrive/c/opt/hadoop-3.2.2'
    export JAVA_HOME='/cygdrive/c/opt/jdk1.8.0_241'
    export HIVE_HOME='/cygdrive/c/opt/apache-hive-3.1.2'
    export PATH=$PATH:$HADOOP_HOME/bin
    export PATH=$PATH:$HIVE_HOME/bin
    export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HIVE_HOME/lib/*.jar
    ```
14. Create the schemas in Apache Derby for Apache Hive at the Cygwin Terminal:
    ```
    schematool -initSchema -dbType derby
    ```
15. Run hive service at the Cygwin Terminal
    ```
    hive --service metastore
    ```
16. Run hive client at the Cygwin Terminal
    ```
    hive
    ```

**DEEPCLOUDLABS**

DEEP CLOUD LABS BİLİŞİM TEKNOLOJİLERİ
EĞİTİM VE DANIŞMANLIK TİC. LTD. ŞTİ.

www.deepcloudlabs.com

```
hive> show tables;
OK
Time taken: 1.164 seconds
hive> create table cities ( id int, name string, countrycode string,district string,population int)
    > row format delimited
    > fields terminated by ','
    > lines terminated by '\n'
    > stored as textfile;
OK
Time taken: 1.306 seconds
hive> show tables;
OK
cities
Time taken: 0.211 seconds, Fetched: 1 row(s)
hive> load data local inpath '/cygdrive/c/tmp/cities.csv' overwrite into table cities;
Loading data to table default.cities
OK
```

```
hive> select sum(population)
    > from cities
    > where countrycode='TUR';
Query ID = dcl_20210323155225_6a74964f-342c-4f8e-b4a2-0b70fa9a7943
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1616502461891_0001, Tracking URL = http://DESKTOP-HSPQDSO:8088/proxy/application_1616502461891_0001/
```