**##결국 모든게 저장 경로 설정, 확인…. 암때나 저장하지 말자…**

파일 몇 개 다운받아서 뭐 합쳐야 함 (bin)파일에 winutils의 bin 합치더라

자바홈, 하둡홈, 자바빈, 자바 s빈 : 경로 설정

winutils :linux환경에서 돌아가는 것들 window에서 돌아가도록 도와줌

set : 환경변수 설정 되어있는거 다 보여줌

하둡 ; 데이터 분산 처리

하둡-> [https://hadoop.apache.org/releases.html 에서 다운 3.12](https://hadoop.apache.org/releases.html%20에서%20다운%203.12) binary 다운

SSH안하기 위해서 가분산모드 : local에서 주인이랑, slave랑 같이 한다. 안나누고

Hadoop-env.sh : linux용

Hadoop-env.sh : window용

<< C:\bigdata\hadoop-3.1.2\etc\hadoop \ Hadoop-env.sh>>

<cmd 에서 간략한 path보는 방법>

cd \

dir /x :c\의 디렉터리 보여줌 거기서 간략한 path와 full path볼 수 이ㅏㅆ다. 하둡에서 경로 설정시 빈 문자열 있음 잘 안되는 경우 많음 그래서 “ “이렇게 적어주거나 간략한 path사용(언더바적힌 부분 빈 문자로 인식)

<바꿔야 하는 부분>

set JAVA\_HOME=C:\PROGRA~1\Java\jdk1.8.0\_211

set HADOOP\_IDENT\_STRING=admin (cmd 에서 set치면 나옴)

<입력>

set HADOOP\_PREFIX=C:\bigdata\hadoop-3.1.2

set HADOOP\_CONF\_DIR=%HADOOP\_PREFIX%\etc\hadoop

set YARN\_CONF\_DIR=%HADOOP\_CONF\_DIR%

set PATH=%PATH%;%HADOOP\_PREFIX%\bin

<< C:\bigdata\hadoop-3.1.2\etc\hadoop \ core-site.xml>>

<configuration>

    <property>

        <name>fs.default.name</name>

        <value>hdfs://0.0.0.0:9000</value>

    </property>

    <property>

        <name>hadoop.tmp.dir</name>

        <value>/c:/bigdata/hadoop-3.1.2/tmp</value>

    </property>

</configuration>

++c:\windows\System32\drivers\etc\host->컴퓨터 전체의 경로

# localhost name resolution is handled within DNS itself.

# 127.0.0.1 localhost <-이부분이 잘 설정되어있어야 한다. (보안 때문에 실행안되는 경우 해결 할 수 있다.\_

# ::1 localhost

192.168.99.23 datanode1 ->이렇게 되어있음 다른 slave컴퓨터 등록된다.

<< C:\bigdata\hadoop-3.1.2\etc\hadoop \ hdfs-site.xml>>

<configuration>

    <property>

        <name>dfs.replication</name> <-자가복제 몇 개 할꺼냐

        <value>1</value>

    </property>

    <property>

        <name>dfs.permissions</name>

        <value>false</value>

    </property>

    <property>

        <name>dfs.namenode.name.dir</name> <-namenode라는 폴더 내가 만듬

        <value>/C:/hadoop-2.7.1/data/namenode</value>

    </property>

    <property>

        <name>dfs.datanode.data.dir</name><-slave를 어디에 저장할 것인지

        <value>/C:/hadoop-2.7.1/data/datanode</value><- datanode라는 폴더 내가만듬

    </property>

</configuration>

<< C:\bigdata\hadoop-3.1.2\etc\hadoop \ mapred-site.xml>>

<configuration>

    <property>

        <name>mapreduce.framework.name</name>

        <value>yarn</value>

    </property>

    <property>

        <name>mapred.job.tracker</name>

        <value>0.0.0.0:9001</value>

    </property>

</configuration>

<<yarn-site.xml>>

<configuration>

<!-- Site specific YARN configuration properties -->

    <property>

        <name>yarn.nodemanager.aux-services</name>

        <value>mapreduce\_shuffle</value>

    </property>

    <property>

        <name>yarn.nodemanager.aux-services.mapreduce\_shuffle.class</name>

        <value>org.apache.hadoop.mapred.ShuffleHandler</value>

    </property>

    <property>

        <name>yarn.log-aggregation-enable</name>

        <value>true</value>

    </property>

    <property>

        <name>yarn.nodemanager.pmem-check-enabled</name>

        <value>false</value>

    </property>

    <property>

        <name>yarn.nodemanager.vmem-check-enabled</name>

        <value>false</value>

    </property>

</configuration>

(node라는 패키지 관리자있으면 에러날 수 있음, python conda같은게 node의 yarn같은거임 이거 깔려 있음 위에 부분 에러 날 수 있음)

<<workers>>->연결할 컴퓨터 주르륵 적어주면 된다.

하둡 2.~->slaves

하둡 3.~->works

<<cmd>>

ping localhost치면 제대로 되었나 확인

<<cmd>>관리자 권한으로 실행

hdfs namenode -format

-> 2020-02-10 13:57:14,422 INFO common.Storage: Storage directory C:\bigdata\hadoop-3.1.2\namenode has been successfully formatted. 이런거 나와야 함

start-dfs.cmd : 창세개 뜸

start-yarn.cmd : 창 두개 뜸

jps : NameNode, ResourceManager, DataNode, NodeManagerm Jps 다섯개 나오면 잘 된 것!