hadoop, spark, zookeeper 설치 가이드

|  |  |  |  |
| --- | --- | --- | --- |
| 버전 | 작성일 | 작성자 | 비고 |
| 1.0 | 2017-11-21 | 이승준 | 초안 작성 |
|  |  |  |  |
|  |  |  |  |

CentOS Linux release 7.1.1503 / 3.10.0-229.el7.x86\_64

spark-2.1.0-bin-hadoop2.7.tgz

hadoop-2.7.4.tar.gz

zookeeper-3.4.8.tar.gz

jdk-8u131-linux-x64.tar.gz

테스트 서버 정보

|  |
| --- |
| dphdata09 : 10.10.76.69 (NameNode, master)  dphdata10 : 10.10.76.70 (NameNode, master)  dphdata11 : 10.10.76.71 (DataNode, worker)  dphdata12 : 10.10.76.72 (DataNode, worker)  dphdata13 : 10.10.76.73 (DataNode, worker) |

서버별 컴포넌트 설치 리스트

|  |
| --- |
| dphdata09 : 10.10.76.69 (hadoop, spark, zookeeper, jdk)  dphdata10 : 10.10.76.70 (hadoop, spark, zookeeper, jdk)  dphdata11 : 10.10.76.71 (hadoop, spark, zookeeper, jdk)  dphdata12 : 10.10.76.72 (hadoop, spark, jdk)  dphdata13 : 10.10.76.73 (hadoop, spark, jdk) |

**가이드 문서에서는 dpadmin 계정을 사용하며, 모든 설치는 dpadmin 계정으로 실행한다.**

**동봉된 설정 값 및 파일들은 서버 환경(IP, Port, 서버 수 등)에 따라 바뀌기 때문에 절대 원본파일 그대로 사용하지 말아야 함.**

1. tar 파일 압축 해제

jdk를 제외한 나머지 tar를 압축 해제 (대상 tar는 설치 리스트 확인.)

1. jdk 1.8.0\_131 설치 (spark를 사용하는 모든 서버 대상)

tar –xvf jdk-8u131-linux-x64.tar.gz

cp jdk1.8.0\_131 /usr/local

ln –sf jdk1.8.0\_131 java

1. 키 생성 후 나머지 서버로 복사

cd /home/dpadmin/.ssh

ssh-keygen -t rsa

키 생성 후

ssh-copy-id -i id\_rsa.pub dpadmin@”host 명” 명령어로 자신을 제외한 서버에 key 전달

ex) dphdata09에서는 10~13으로 전달 ssh-copy-id -i id\_rsa.pub dpadmin@dphdata10

1. tar 압축 해제 및 링크 파일 생성 (1에서 옮겨둔 tar 기준)

tar –xvf spark-2.1.0-bin-hadoop2.7.tgz

ln –sf spark-2.1.0-bin-hadoop2.7 spark

tar –xvf hadoop-2.7.4.tar.gz

tar –xvf zookeeper-3.4.8.tar.gz

1. 컴포넌트별 설정

**Hadoop**

동봉된 core-site.xml, hdfs-site.xml, hadoop-env.sh, include\_server를

/home/dpadmin/hadoop/hadoop-2.7.4/etc/hadoop/ 안에 넣어준다.

(계정명, 버전에 따라 바뀔 수 도 있음.)

touch /home/dpadmin/hadoop/hadoop-2.7.4/etc/hadoop/exclude\_server

chown dpadmin. -R /data/

mkdir -p /data/zookeeper/tmp

vim /home/dpadmin/.bashrc

|  |
| --- |
| HADOOP\_HOME=/home/dpadmin/hadoop/hadoop-2.7.4/  HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop  HADOOP\_PREFIX=$HADOOP\_HOME  export HADOOP\_HOME HADOOP\_CONF\_DIR HADOOP\_PREFIX  export JAVA\_HOME=/usr/local/java  #SPARK  SPARK\_HOME=/home/dpadmin/spark  #ZOOKEEPER  export ZOOKEEPER\_HOME=/home/dpadmin/zookeeper  export ZOOKEEPER\_PREFIX=$ZOOKEEPER\_HOME  export ZOO\_LOG\_DIR=$ZOOKEEPER\_HOME/logs  export PATH=$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin:$SPARK\_HOME/bin:$SPARK\_HOME/sbin:$ZOOKEEPER\_HOME/bin:$JAVA\_HOME/bin:$PATH |

source /home/dpadmin/.bashrc

DataNode에서는 자신이 사용할 데이터 디렉토리를 만들어줘야함

현재 가이드에서는 data01~13까지 사용중

ex) mkdir –p /data01/hadoop/hdfs/data

**Zookeeper**

동봉된 zoo.cfg, java.env를 /home/dpadmin/zookeeper/conf에 복사

mkdir -p /data/zookeeper/tmp

zoo.cfg의 server.x값을 /data/zookeeper/tmp/myid에 넣어준다.

(현재 설정으로는 dphdata09 (1), dphdata10 (2), dphdata11 (3))

echo “1” >> /data/zookeeper/tmp/myid (서버별)

**Spark**

동봉된 slaves, spark-defaults.conf, spark-env.sh를 /home/dpadmin/spark/conf로 복사

1. 시작 및 initialize

주키퍼 start : /home/dpadmin/zookeeper/bin/zkServer.sh start

주키퍼 초기화 : hdfs zkfc -formatZK (**주키퍼 서버 실행 상태로 한곳에서만 실행** ex dphdata09)

네임 노드 Active (dphdata09)

저널노드 start : hadoop-daemon.sh start journalnode

네임노드 초기화 : hdfs namenode –format (**네임노드 포맷전에 반드시 저널노드 실행**)

저널노드 초기화 : hdfs namenode –initializeSharedEdits

네임노드 실행 : hadoop-daemon.sh start namenode

zkfc 실행 : hadoop-daemon.sh start zkfc

네임노드 Standby (dphdata10)

Standby 네임노드 준비 : hdfs namenode -bootstrapStandby

네임노드 실행 : hadoop-daemon.sh start namenode

zkfc 실행 : hadoop-daemon.sh start zkfc

데이터 노드 실행 (dphdata11, 12, 13)

hadoop-daemon.sh start datanode

**SPARK**

master node : $SPARK\_HOME/sbin/start-master.sh (dphdata09, 10)

worker node : $SPARK\_HOME/sbin/start-slave.sh spark://10.10.76.69:17077,10.10.76.70:17077

(master A/B의 ip를 적어준다. 현재 설정은 dphdata09 (Active), dphdata10 (Backup))

1. 웹 페이지 확인

hadoop : <http://10.10.76.69:10070/dfshealth.html#tab-overview>

spark : http://10.10.76.69:18080/