

# 빅데이터 저장시스템 개발

김은표

# 목차

1. VMWARE 설치
2. Centos 다운로드
3. VMWARE 설치
4. VMWARE 설정
5. Hadoop 설치
6. Sqoop 설치
7. Sqoop Import/Export 실습하기

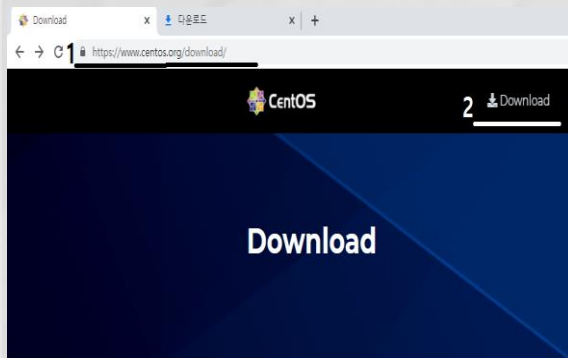
# 1. VMWARE 다운로드

1. <https://www.vmware.com/kr.html> 접속
2. 무료제품 다운로드 클릭
3. [Workstation Player](#) 클릭
4. Windows용 Workstation 16.0 Player 클릭



## 2. Centos 다운로드

1. <https://www.centos.org/> 접속
2. Download 클릭
3. 버전 7(2009) 클릭
4. X86\_64 버전 클릭
5. [http://mirror.kakao.com/centos/7.9.2009/isos/x86\\_64/](http://mirror.kakao.com/centos/7.9.2009/isos/x86_64/) 접속
6. [CentOS-7-x86\\_64-DVD-2009.iso](#) 다운로드



### CentOS Linux


8 (2004)	3 (2009)	610
ISO	Packages	Others
4 x86_64	RPMS	Cloud   Containers   Vagrant
ARM64 (aarch64)	RPMS	Cloud   Containers   Vagrant
IBM Power BE (ppc64)	RPMS	Cloud   Containers   Vagrant
IBM Power (ppc64le)	RPMS	Cloud   Containers   Vagrant
ARM32 (armhfp)	RPMS	Cloud   Containers   Vagrant
i386	RPMS	Cloud   Containers   Vagrant
Release Notes	Release Email	Documentation

The following mirrors in your region should have the ISO images available:

[http://mirror.kakao.com/centos/7.9.2009/isos/x86\\_64/](http://mirror.kakao.com/centos/7.9.2009/isos/x86_64/)  
[http://mirror.navercorp.com/centos/7.9.2009/isos/x86\\_64/](http://mirror.navercorp.com/centos/7.9.2009/isos/x86_64/)  
[http://mirror.anigil.com/CentOS/7.9.2009/isos/x86\\_64/](http://mirror.anigil.com/CentOS/7.9.2009/isos/x86_64/)  
[http://ftp.kaist.ac.kr/CentOS/7.9.2009/isos/x86\\_64/](http://ftp.kaist.ac.kr/CentOS/7.9.2009/isos/x86_64/)

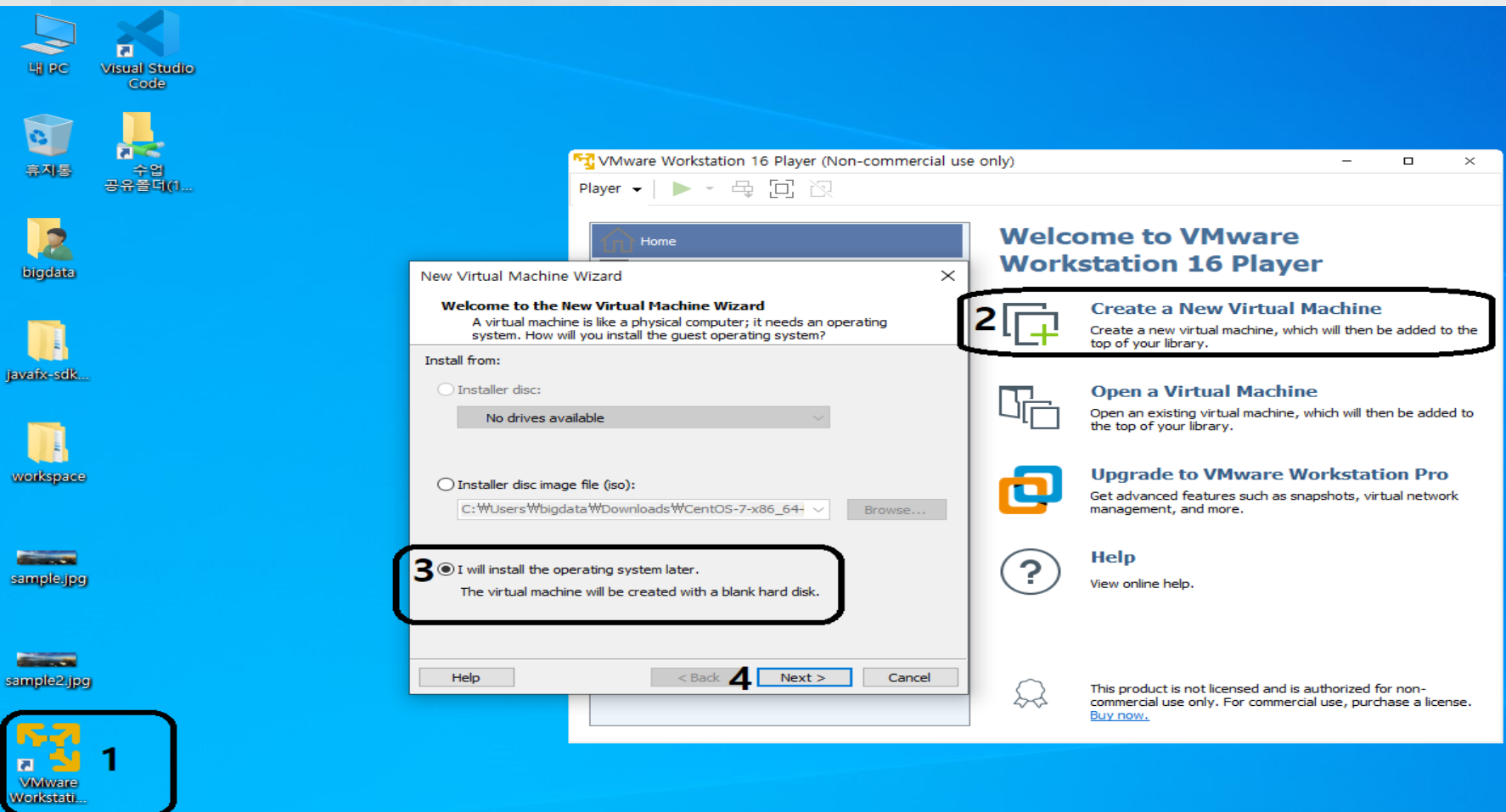
Other mirrors further away:

[http://mirrors.cat.net/centos/7.9.2009/isos/x86\\_64/](http://mirrors.cat.net/centos/7.9.2009/isos/x86_64/)  
[http://ftp.jaist.ac.jp/pub/Linux/CentOS/7.9.2009/isos/x86\\_64/](http://ftp.jaist.ac.jp/pub/Linux/CentOS/7.9.2009/isos/x86_64/)  
[http://ftp.riken.jp/Linux/centos/7.9.2009/isos/x86\\_64/](http://ftp.riken.jp/Linux/centos/7.9.2009/isos/x86_64/)  
[http://ftp.yz.yamagata-u.ac.jp/pub/linux/centos/7.9.2009/isos/x86\\_64/](http://ftp.yz.yamagata-u.ac.jp/pub/linux/centos/7.9.2009/isos/x86_64/)  
[http://ftp.sukuba.wide.ad.jp/Linux/centos/7.9.2009/isos/x86\\_64/](http://ftp.sukuba.wide.ad.jp/Linux/centos/7.9.2009/isos/x86_64/)  
[http://ftp-srv2.kddilabs.jp/Linux/packages/CentOS/7.9.2009/isos/x86\\_64/](http://ftp-srv2.kddilabs.jp/Linux/packages/CentOS/7.9.2009/isos/x86_64/)

 CentOS		
CentOS on the Web: <a href="#">CentOS.org</a>   <a href="#">Mailing Lists</a>   <a href="#">Mirror List</a>		
File Name ↓	File Size ↓	Date ↓
Parent directory/	-	-
0_README.txt	2495	2020-Nov-06 14:32
CentOS-7-x86_64-DVD-2009.iso	4712300544	2020-Nov-04 11:37
CentOS-7-x86_64-DVD-2009.torrent	180308	2020-Nov-06 14:44
CentOS-7-x86_64-Everything-2009.iso	10200547328	2020-Nov-02 15:18
CentOS-7-x86_64-Everything-2009.torrent	389690	2020-Nov-06 14:44
CentOS-7-x86_64-Minimal-2009.iso	1020264448	2020-Nov-03 14:55
CentOS-7-x86_64-Minimal-2009.torrent	39479	2020-Nov-06 14:44
CentOS-7-x86_64-NetInstall-2009.iso	602931200	2020-Oct-26 16:26
CentOS-7-x86_64-NetInstall-2009.torrent	23567	2020-Nov-06 14:44
sha256sum.txt	398	2020-Nov-04 11:38
sha256sum.txt.asc	1258	2020-Nov-06 14:37

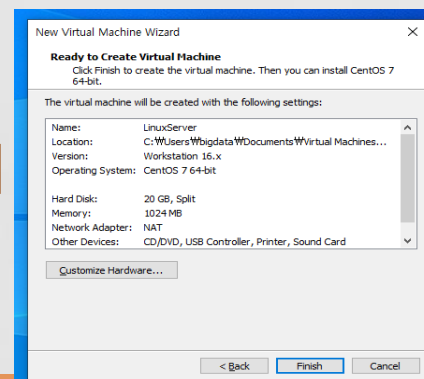
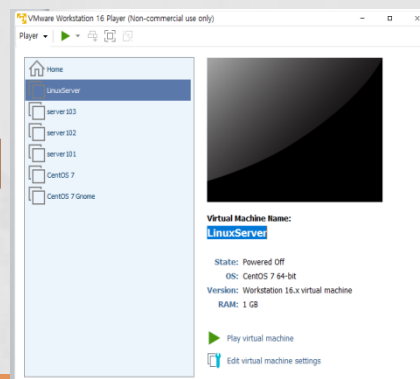
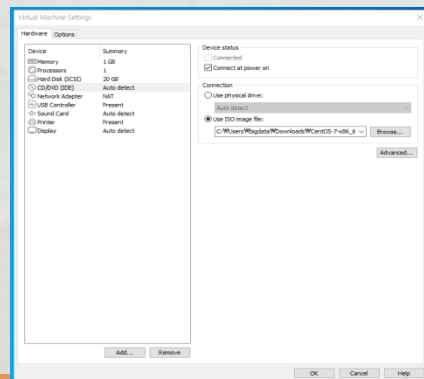
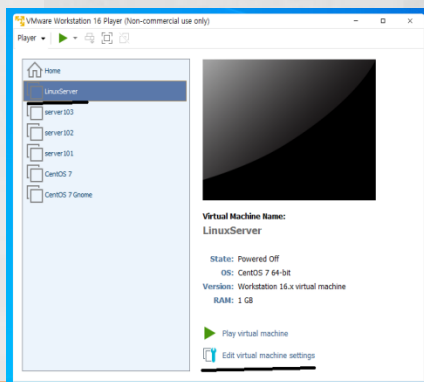
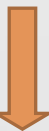
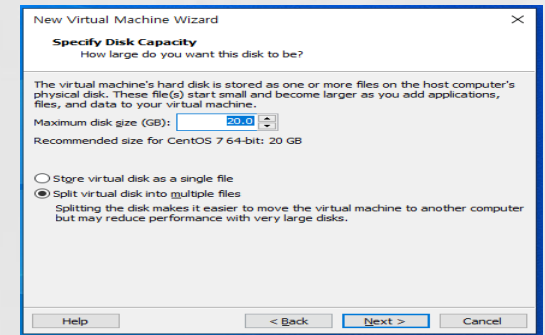
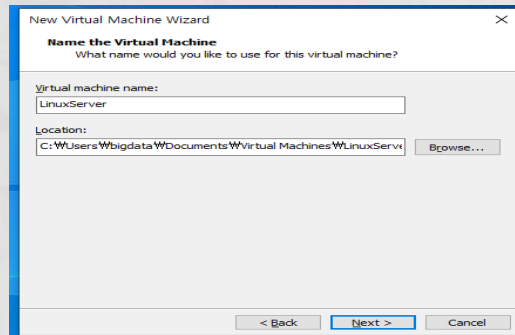
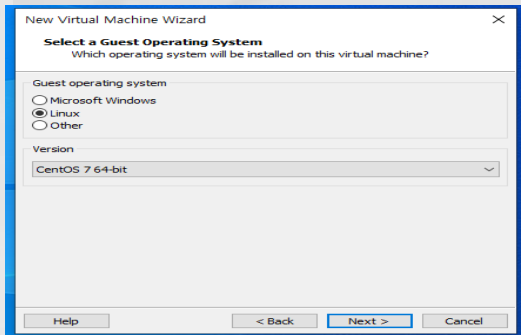
# 3. VMWARE 설치

1. Vmware 실행
2. Create a New Virtual Machine 클릭
3. I will install the oper 클릭 후 Next



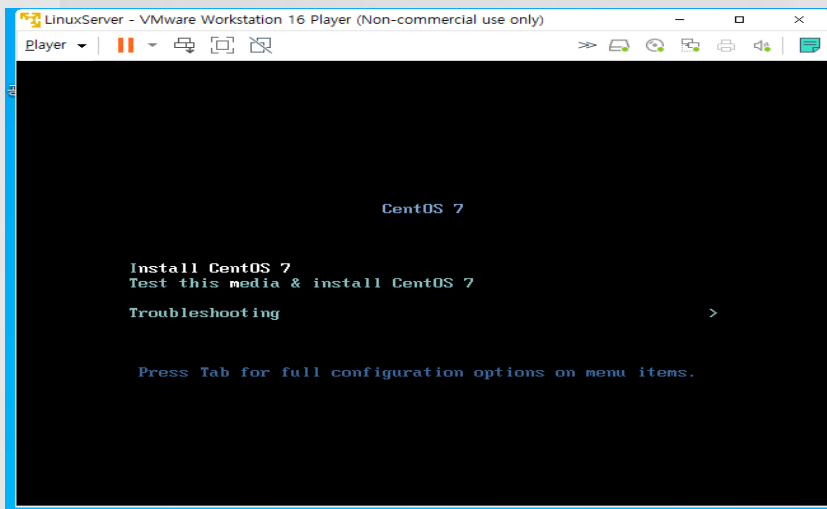
# 3. VMWARE 설치

1. Linux 선택
2. Virtual machine name 이름 설정
3. 용량 선택 후 끝내기
4. 실행 할 폴더 클릭
5. Edit virtual machine settings 클릭
6. CD/DVD (IDS) 선택 Use ISO 설정 하기  
(다운받은 Centos ISO 파일)



# 4. VMWARE 설정

## 1. Install



## 2. 언어선택



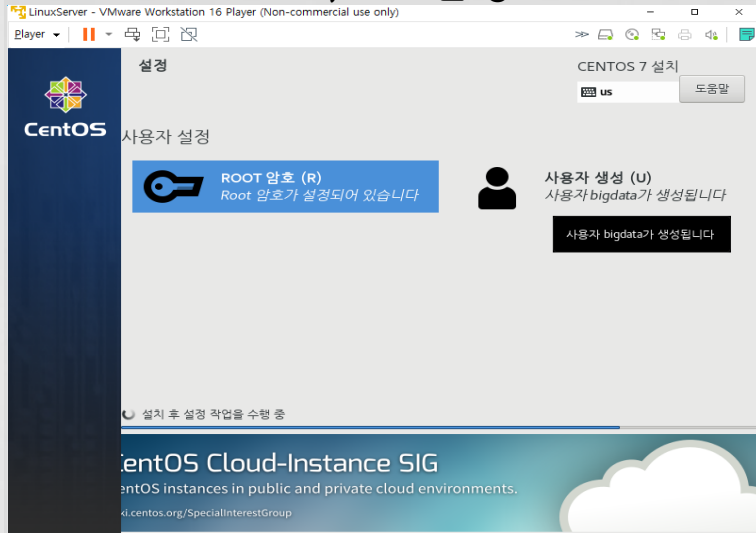
## 3. 설정하기

- ① 키보드 영어(미국)우선순위
- ② KDUMP 비활성화
- ③ 네트워크 및 호스트 우선으로 설정

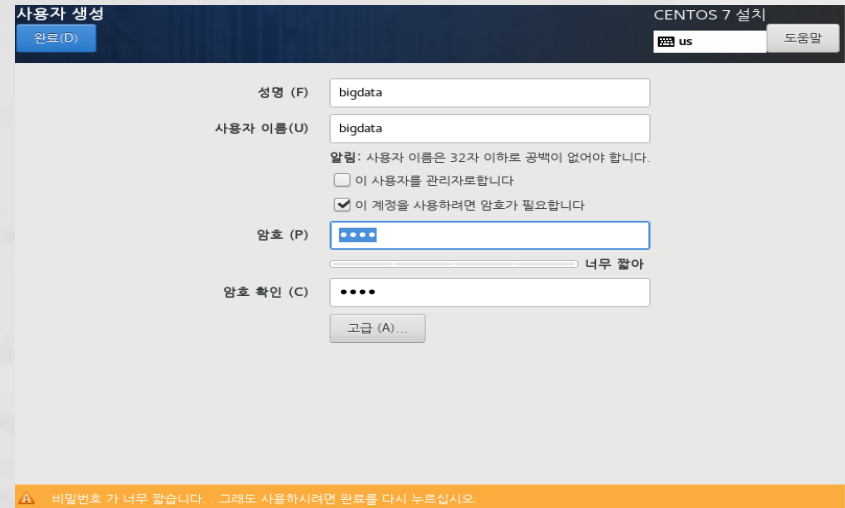
## 4. 설치시작

# 4. VMWARE 설정

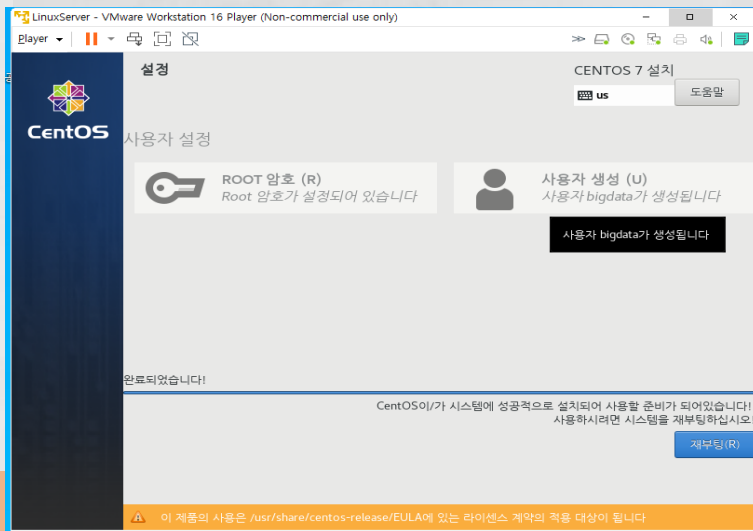
## 1. ROOT 암호 설정



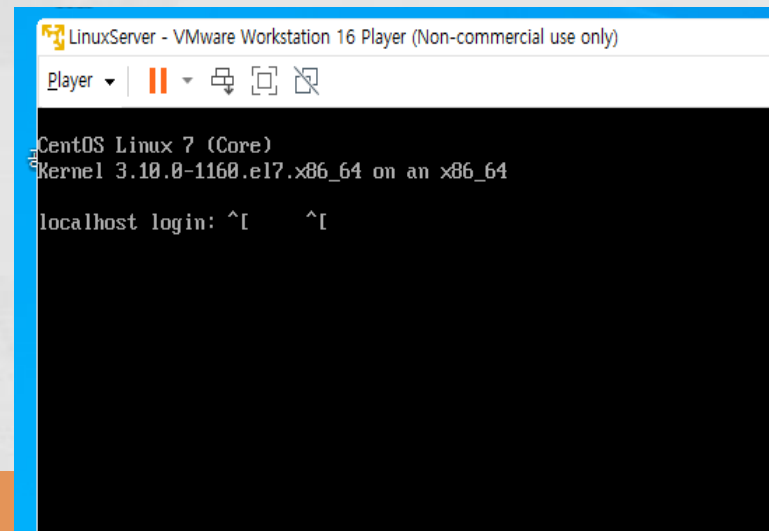
## 2. 사용자 생성



## 3. 설치 완료 후 재부팅

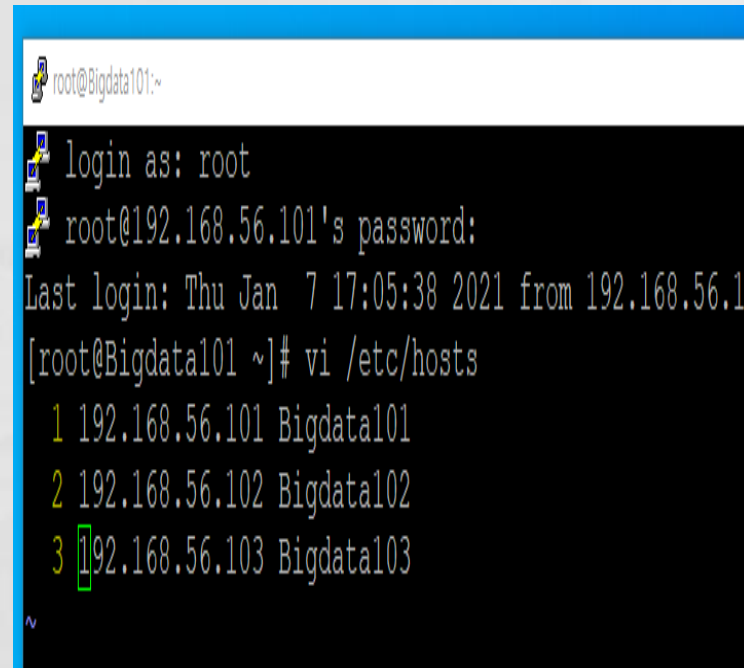


## 4. 로그인 하기






# 5.Hadoop 설치



하둡 시스템 가상 머신 최소 3대 생성,  
가상 머신 사양 하드디스크 최소 8GB, 메모리 최소 2GB 구성  
가상 머신 3대 CentOS7 최소 설치, 'Bigdata' 계정 생성  
필수 애플리케이션 설치 및 설정

# 5.Hadoop 설치

하둡 2.10.1 버전 다운로드/압축해제/이동/  
링크생성(Bigdata101, Bigdata102, Bigdata103 동일)

 Apache Hadoop

## Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

Version	Release date	Source download	Binary download	Release notes
2.10.1	2020 Sep 21	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a>	<a href="#">Announcement</a>


```
[root@Bigdata101 ~]# cd /home/bigdata/
[root@Bigdata101 bigdata]# ll
합계 4
drwxr-xr-x  8 root    root    200 1m  7 16:39 apache-flume-1.9.0-bin
lrwxrwxrwx  1 root    root      22 1m  7 15:17 flume -> apache-flume-1.9.0-bin
lrwxrwxrwx  1 root    root      14 1m  5 17:30 hadoop -> hadoop-2.10.1/
drwxr-xr-x 11 bigdata bigdata 204 1m  6 14:49 hadoop-2.10.1
drwxr-xr-x  8 root    root    114 1m  8 11:31 naver
lrwxrwxrwx  1 root    root      30 1m  6 17:24 sqoop -> sqoop-1.4.7.bin__hadoop-2.6.0/
drwxr-xr-x 10 bigdata bigdata 4096 1m  7 18:20 sqoop-1.4.7.bin__hadoop-2.6.0
drwxr-xr-x  6 root    root     97 1m  7 15:10 weather
```

# 5. Hadoop 설치

- Hadoop 상태 확인(bigdata101에서 #start-all.sh 실행)
- Hadoop 브라우저 확인
  - <http://192.168.100.101:50070> → Hadoop Web UI
  - <http://192.168.100.101:8088> → Hadoop Resource Manager

```
emanager-Bigdata101.out
Bigdata102: starting nodemanager, logging to /h
-nodemanager-Bigdata102.out
Bigdata103: starting nodemanager, logging to /h
-nodemanager-Bigdata103.out
Bigdata101: starting nodemanager, logging to /h
-nodemanager-Bigdata101.out
You have new mail in /var/spool/mail/root
[root@Bigdata101 bigdata]#
```

Configured Capacity:	18.56 GB
DFS Used:	30.09 MB (0.16%)
Non DFS Used:	11.63 GB
DFS Remaining:	6.9 GB (37.19%)
Block Pool Used:	30.09 MB (0.16%)
DataNodes usages% (Min/Median/Max/stdDev):	0.16% / 0.16% / 0.16% / 0.00%
Live Nodes	3 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	Fri Jan 08 11:49:54 +0900 2021
Last Checkpoint Time	Fri Jan 08 11:47:17 +0900 2021



[Cluster](#)  
[About](#)  
[Nodes](#)  
[Node Labels](#)  
[Applications](#)  
[NEW](#)  
[NEW SAVING](#)  
[SUBMITTED](#)  
[ACCEPTED](#)  
[RUNNING](#)  
[FINISHED](#)  
[FAILED](#)  
[KILLED](#)  
[Scheduler](#)

**Cluster Metrics**

Apps Submitted	Apps Pending
0	0

**Cluster Nodes Metrics**

Active Nodes	Decommissioning Nodes
3	0

**Scheduler Metrics**

Scheduler Type
Capacity Scheduler [ <a href="#">name=memory-mb de</a> ]

Show 20 entries

ID	User	Name	Application Type	Queue	Appli
▼	↕	↕	↕	↕	Prio








Showing 0 to 0 of 0 entries

# 6.Sqoop 설치

Sqoop-1.4.7.bin\_\_hadoop-2.6.0.tar.gz 다운로드 해야됨  
Sqoop 다운로드/압축해제/이동/링크 생성(Namenode 실행)

← → ↻ <https://downloads.apache.org/sqoop/1.4.7/>

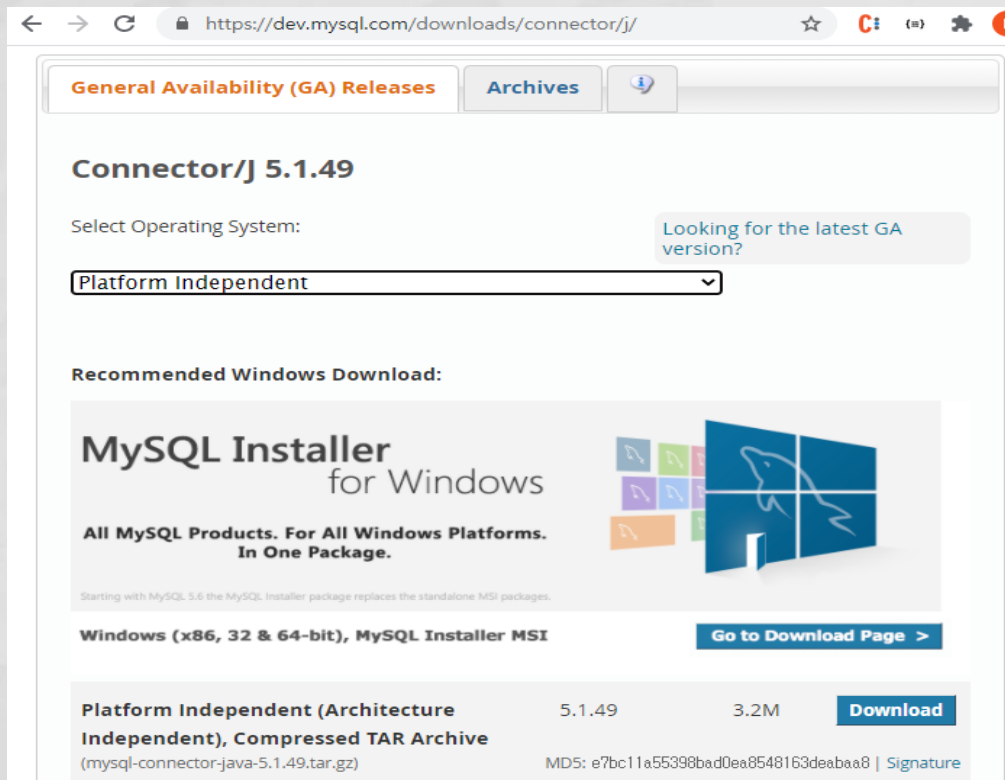
## Index of /sqoop/1.4.7

Name	Last modified	Size	Description
 <a href="#">Parent Directory</a>	-	-	-
 <a href="#">sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz</a>	2020-07-06 15:19	17M	
 <a href="#">sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz.asc</a>	2020-07-06 15:20	819	
 <a href="#">sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz.md5</a>	2020-07-06 15:19	71	
 <a href="#">sqoop-1.4.7.tar.gz</a>	2020-07-06 15:20	1.1M	
 <a href="#">sqoop-1.4.7.tar.gz.asc</a>	2020-07-06 15:19	819	
 <a href="#">sqoop-1.4.7.tar.gz.md5</a>	2020-07-06 15:20	53	

```
[root@Bigdata101 bigdata]# ll
합계 4
drwxr-xr-x  8 root    root    200 1m  7 16:39 apache-flume-1.9.0-bin
lrwxrwxrwx  1 root    root     22 1m  7 15:17 flume -> apache-flume-1
lrwxrwxrwx  1 root    root     14 1m  5 17:30 hadoop -> hadoop-2.10.1
drwxr-xr-x 11 bigdata bigdata 204 1m  6 14:49 hadoop-2.10.1
drwxr-xr-x  8 root    root    114 1m  8 11:31 naver
lrwxrwxrwx  1 root    root     30 1m  6 17:24 sqoop -> sqoop-1.4.7.bi
/
drwxr-xr-x 10 bigdata bigdata 4096 1m  7 18:20 sqoop-1.4.7.bin__hadoop
drwxr-xr-x  6 root    root     97 1m  7 15:10 weather
You have new mail in /var/spool/mail/root
```

# 6.Sqoop 설치

- RDBMS 드라이버 설치(Namenode 실행)  
<http://dev.mysql.com/downloads/connector/j/>에서  
리눅스용 mysql-connector-java-5.1.49.tar.gz 파일 다운로드
- FileZilla FTP로 mysql-connector-java-5.1.49.tar.gz 파일을 /root 경로에 업로드
- mysql-connector-java-5.1.49.tar.gz 압축해제/라이브러리 복사



# 7. Sqoop Import 실습하기

- Hadoop 실행 후, Sqoop import 실행
- HDFS 확인

http://192.168.xxx.101:50070 브라우저 확인

Utilities – Browse the file system - /sqoop/User1 확인 및 파일 내용 확인

```
root@Bigdata101:~  
at org.apache.sqoop.mapreduce.ImportJobBase.runImport (ImportJobBase.java:270)  
at org.apache.sqoop.manager.SqlManager.importTable (SqlManager.java:127)  
at org.apache.sqoop.manager.MySQLManager.importTable (MySQLManager.java:127)  
at org.apache.sqoop.tool.ImportTool.importTable (ImportTool.java:628)  
at org.apache.sqoop.tool.ImportTool.run (ImportTool.java:628)  
at org.apache.sqoop.Sqoop.run (Sqoop.java:147)  
at org.apache.hadoop.util.ToolRunner.run (ToolRunner.java:76)  
at org.apache.sqoop.Sqoop.runSqoop (Sqoop.java:183)  
at org.apache.sqoop.Sqoop.runTool (Sqoop.java:234)  
at org.apache.sqoop.Sqoop.runTool (Sqoop.java:243)  
at org.apache.sqoop.Sqoop.main (Sqoop.java:252)  
  
You have new mail in /var/spool/mail/root  
[root@Bigdata101 bigdata]# systemctl status httpd  
Unit httpd.service could not be found.  
You have new mail in /var/spool/mail/root  
[root@Bigdata101 bigdata]# cd  
[root@Bigdata101 ~]# systemctl status httpd  
● httpd.service - The Apache HTTP Server  
   Loaded: loaded (/usr/lib/systemd/system/httpd.service; disabled; vendor preset: disabled)  
   Active: inactive (dead)  
     Docs: man:httpd(8)  
           man:apachectl(8)
```

주요 요약 | http://192.168.56.101:50070/explorer.html#/

## Hadoop

Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

### Browse Directory

/ Go!

Show 25 entries Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Jan 06 16:01	0	0 B	HTML
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Jan 07 16:15	0	0 B	flume
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Jan 07 17:51	0	0 B	sqoop
<input type="checkbox"/>	drwx-----	root	supergroup	0 B	Jan 07 17:51	0	0 B	tmp

# 7.Sqoop Export 실습하기

- Hadoop 실행 후, Sqoop Export 실행
- HeidiSQL 접속 – sqoop 데이터베이스 – User2 테이블 데이터 확인

```
[root@Bigdata101 ~]# sqoop export -- connect jdbc:mysql://192.168.56.101:3306/sqoop --table User2 --export-dir /sqoop/User1 --username sqoop --password 1234 -m
```

Bigdata101@sqoop@User2 - HeidiSQL 11.1.0.6116

파일 편집 검색 쿼리 도구 이동 도움말

데이터베이스 필터 테이블 필터

EMPLOYEE.sql\* mysql.sql

호스트: 192.168.56.101 데이터

sqoop.User2: 6 행 (총) (대략적)

uid	name	hp	age
A101	???	010-1234-1111	23
A102	???	010-1234-2222	21
A103	???	010-1234-3333	35
A101	???	010-1234-1111	23
A102	???	010-1234-2222	21
A103	???	010-1234-3333	35

Bigdata101

- information\_schema
- mydb
- mysql
- performance\_schema
- sqoop 32.0 KIB
  - User1 16.0 KIB
  - User2 16.0 KIB
  - weather