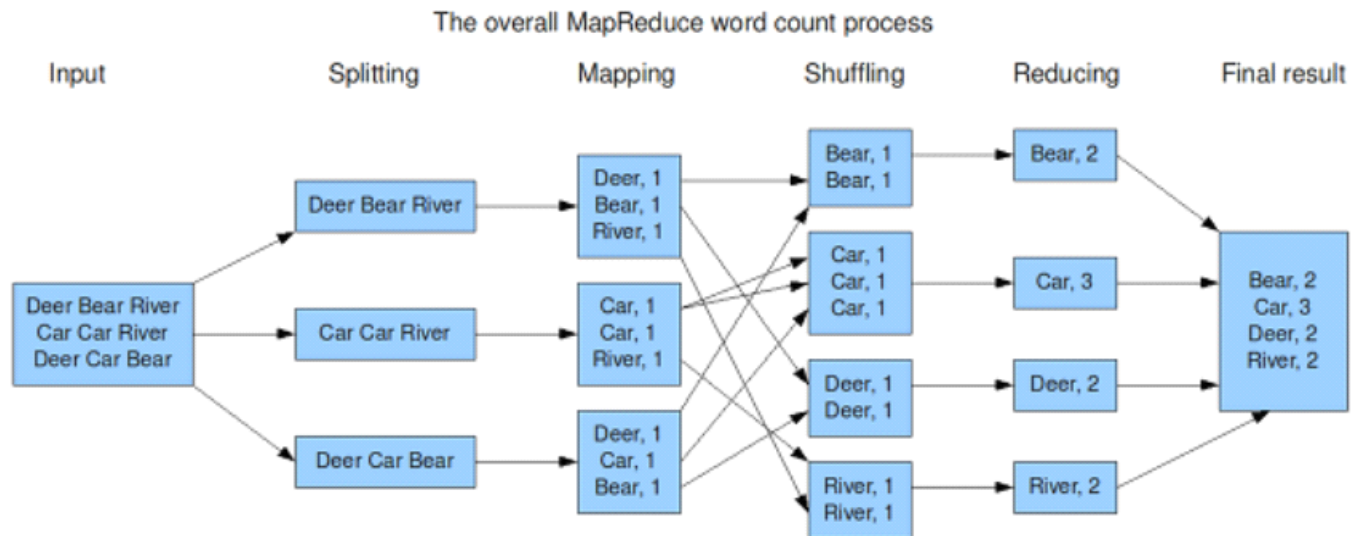


맵리듀스(Map/Reduce) 이해하기

2017년 10월 5일 목요일 오후 5:07



[참고] 맵리듀스 동작이해

2017년 11월 30일 목요일 오후 1:19

블로그 한지 3일째다.

실제 hadoop의 설치나 프로그래밍적 이슈는 언제 다룰 수 있을지는 모르겠다.

하지만 그런것들은 공식 웹사이트의 [getting started](#) 같은 것을 보면 정말 쉽게 수행할 수 있다.

hadoop은 수백대의 클러스터를 구성하여 linux 환경에서 돌리는 것이나 stand alone 환경으로 돌리는 것이나

hdfs와 mr의 큰 개념을 잡기에 전혀 차이가 없다. (매니징은 예외)

개념보다 당장 돌려보고 싶다면 추천! ㅎㅎ

아주 기본적인 Map Reduce의 개념

map reduce의 개념은 크게 어렵지 않다.

예전 우리프로젝트가 hadoop을 도입하기전에 매우 큰단위의 데이터를 처리하기 위해서 일괄적인 기준으로 데이터를 분배하여 분배된 데이터를 일관된 형식으로 처리하는 형식의 프레임워크를 구현한적이 있다.

예를들어서 1부터 1억까지의 임의의 순서로 정렬된 숫자 수열의 모든 숫자를 더하고 싶다고 가정하자 (일반화된 공식을 사용할 수 없고 이터레이팅으로만 계산할 수 있다고 가정하자). 그런데 싱글머신에서 저 이터레이팅 시간의 비용이 많이 든다고 추측되어

n대의 머신에 (1억/n)으로 나뉜 task 만큼 각 머신에 분배하여

1번 머신은 1부터 1000까지

2번 머신은 1001부터 2000 까지

....

n대의 머신이 각각 처리되게 한 후 마지막에 각 머신의 결과를 더했다고 가정하자.

그리고 위 사항을 일반화 시켜 x자리의 데이터의 덧셈에 대해 일관되게 작동하게 구현했다고 가정하자.

그렇다! 당신은 해당 계산에 대하여 map reduce를 적용한것이다 ㅎㅎ

(map 과정에 대한 설명이 다소 약한 예제같다)

이제부터 다시 교과서적인 설명을 보도록하자

MapReduce 는 클러스터링 된 컴퓨터 장비에서 페타 바이트 단위의 거대한 데이터를 분산 컴퓨팅 할 수 있는 목적으로 google에 의해 2004년 도입된 소프트웨어 프레임웍이다.

이프레임웍은 일반적으로 함수형 언어에서 사용되는 Map함수와 Reduce함수에서 힌트를 얻어서 만들어졌지만 그 목적이 원래의 그것과 완전히 동일하지 않다.

MR을 c++, c#, 얼랭, 자바, 오캄, 펄, 파이썬, PHP, 루비, F#, R등의 언어로 구현된 라이브러리들이 있다.

MR은 빅데이터(big data)를 많은 컴퓨터군(노드)를 이용하여 병렬처리 하기위한 프레임웍이다.

해당 작업을 파일 시스템(unstructured)이나 데이터베이스(structured)에 저장된 데이터에 대하여 수행 할 수 있다.

MR의 단계

Map 단계 - 마스터 노드는 입력 데이터를 받아 그것을 더 작은 단위로 분할하여 여러 워커 노드들에게 분산한다.

작업을 할당 받은 워커 노드들은 해당 데이터들을 더 작은 단위로 분할하여 다른 워커 노드에 배치하는 멀티레벨 트리 구조로 배치할 수도 있다.

결과적으로 워커 노드가 작은 단위로 분할된 데이터를 처리하고 나면 그 결과를 마스터 노드에게 알려준다.

Reduce 단계 - 마스터 노드는 작은 단위의 문제로 쪼개진 Map의 결과를 결합(combine)하고 수집(collect)하여 우리가 원하는 어떤 output을 만들어낸다.

MapReduce의 특징은 M과 R의 각 단계에서 병렬 처리가 가능 한 것이며, 각 Map의 처리 과정은 다른 Map의 처리과정과 완전 독립적이고 이론적으로 모든 병렬을 실행 할 수 있다. (실제로는 데이터와 CPU의 수에 따라 제한됨)

Reuce단계에서는 Map 단계에의 처리 결과를 키로 정렬하고 Reduce 처리단으로 보내게 되는데 이역시 병렬처리 가능하다.

MR의 일련 과정은 순차실행 알고리즘들에 비해 종종 비효율적으로 생각될 수 있는데 MR은 일반 서버가 취급할수 있는 데이터 양을 훨씬 초월한 데이터를 취급할 수 있다.

많은 서버를 가지고 있으면 MR은 페타 바이트 규모의 빅 데이터 정렬을 몇시간안에 처리할

수 있으며 프로세스 과정보다 병렬이기 때문에 어떤 서버나 스토리지의 일부에서 장애가 일어난다고 해도 해당 M과 R의 처리 노드만 다시 재할당하여 수행 될 수 있기 때문에 장애에 대해서도 지속적인 복구가 가능하다.

논리적 구성

Map Reduce 함수는 key와 value로 구조화된 데이터 페어에 대하여 수행 될 수 있다. (이하 k,v)

Map(k1, v1) -> list(k2, v2)

Map 함수는 입력 데이터에 대해 완전 병렬적으로 작동되며 해당 결과로 (k2, v2)에 대한 리스트가 생성된다.

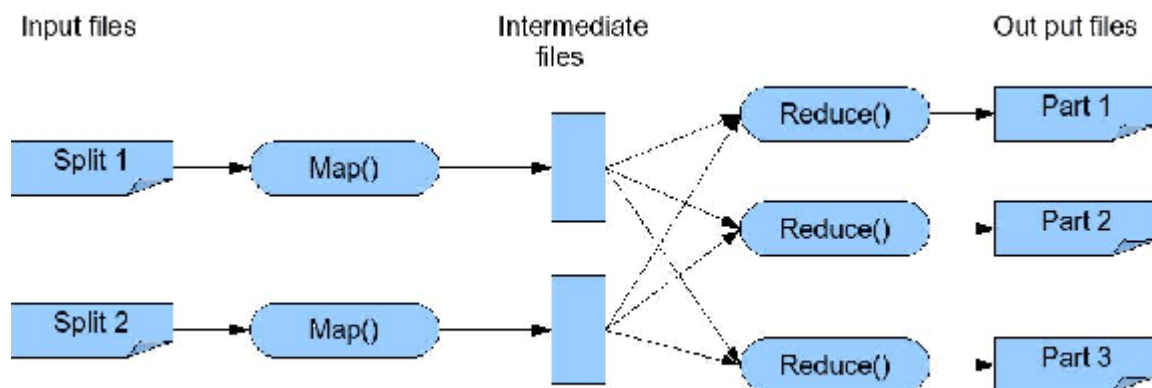
MR 프레임웍은 저런식으로 생성된 모든 리스트페어를 수집하여 같은 키에 대한 리스트들을 그룹핑한다.

하나의 그룹은 하나의 키로 가지고 그룹된 v들의 집합이다.

Reduce(k2, list(v2)) -> list(v3)

각 reduce 함수들은 1개의 결과나 결과가 없을수도 혹은 여러개의 결과를 뱉을 수도 있다.

결과적으로 MR 프레임웍은 k,v의 리스트를 다른 value의 리스트로 치환시키는 역할을 한다. 이것이 일반적인 함수형 언어의 맵, 리듀스와 다른점이다.



< Mapreduce 모델 - 한장도 그림없음 이상하니까>

예제

간단한 예제를 통하여 MR을 알아보자.

우리는 어떤 단어들이 나열된 문서 파일을 가지고 있다고 가정한다. 각 문서 파일끼리 서로

중복된 단어도 있을 수 있다고 가정한다.

우리는 모든문서에서 무슨 단어들이 총 몇번 출현했나를 알아보는 것에 관심이 있다고 가정하자.

데이터 형식은 아래와 같다고 가정한다.

doc 1

감자

고양이

호랑이

고양이

doc 2

사슴

호랑이

토끼

.....

등 등

Map 함수와 Reduce 함수는 아래와 같이 작성되었다고 가정한다.

```
void map (String name, String document) :
```

```
    // name : document name
```

```
    // document : document contents
```

```
    for each word w in document :
```

```
        EmitIntermediate (w "1");
```

```
void reduce (String word, Iterator partialCounts)
```

```
    // word : a word
```

```
    // partialCounts : a list of aggregated partial counts
```

```
    int sum = 0;
```

```
    for each pc in partialCounts :
```

```
        sum + = ParseInt (pc);
```

```
    Emit (word, AsString (sum));
```

각 응용프로그램들은 아래와 같은 함수들을 정의하여 작동되게 된다.

인풋 리더(Input reader)

맵 함수(Map function)
파티션 함수(partition function)
비교 함수(compare function)
리듀스 함수9Reduce function)
아웃풋 라이터(output writer)

위의 예제와 위의 함수들을 통하여 상황을 가정해본다.

인풋 리더(Input reader)

인풋리더는 doc들을 적절하게 지정된 사이즈로 데이터를 분할한다. (일반적으로 16mb ~ 128mb) 분할된 데이터는 프레임워크에 의하여 각 Map 함수단으로 넘어가게 된다. input 리더는 key, value 페어의 데이터를 읽을 수 있으며

지금 상황에서는 key가 파일명, value가 내용이라고 가정한다.

예제의 각 doc의 용량은 input reader가 우연찮게 각 파일당으로 분리했다고 가정하자.

그럼 인풋리더는 doc 1에 대하여 map 함수 할당을 요청하게 되고 병렬적으로 doc 2 에 대하여 map 함수 할당을 요청하게 된다. (임의로 doc1이 할당된 맵을 m1, doc2가 할당된 맵을 m2라고 하자)

맵 함수(Map function)

각 맵함수는 key / value 페어 데이터들을 받아 지정된 함수로 처리하고 결과를 리턴한다.

m1~2들은 받은 데이터가 종료될때까지 map(String name, String document) 함수를 call 하게 된다.

그럼 doc1 의 경우 함수 정의에 의하여 (감자, 1), (고양이,2), (호랑이,1) 등을 각각 리턴하게 된다.

doc2 는 (사슴,1), (호랑이,1), (토끼,1)

위와같은 과정에서 key가 단어로 value가 해당 문서에서 단어 출현 횟수로 변경됐다는 사실에 주목한다.

파티션 함수(Partition function)

맵함수의 결과가 리듀스 함수로 정리되어 넘어가게 해주는 역할을 하는 함수이며 [Sharding](#)을 목적으로 한다. 파티션 함수는 주어진 key를 지정된 리듀스 개수 만큼 분배하게 된다.

우리는 리듀스 머신 개수를 2개로 가정했다고 하자. 그럼 지정된 방법으로 파티션함수는

key의 값을 분석하여 % 2 (리듀스 머신 개수) 각 reduce 함수에게 처리될 결과들을 전달한다. (일반적으로 key값을 해시하여 모드연산을 취한다)

해당 과정에서 γ, ς 으로 시작하는 애들은 머신1에 $\text{ㅎ}, \text{ㅌ}$ 으로 시작하는 애들은 머신2에 분배되었다고 가정한다.

비교 함수(Comparison fuction)

해당과정에서는 지정된 방법에 따라 결과들을 정렬한다. 해당 과정을 세부분화하면, data의 reduce상에서의 순서, 그룹되는 단위 지정들을 할 수 있다. (일반적인 상황에선 신경을 덜 써도 무관할 것이라고 생각된다)

리듀스 함수(Reduce function)

맵함수의 결과는 각 키로 정렬되어 reduce 함수에서 수행되게 된다. 위에 말한 것처럼 해당 결과는 0개나 그이상의 결과로 출력될 수 있다.

γ 은 r1에 ς 은 r2에 ㅎ 은r3에 ㅌ 은 r4에서 처리 된다고 가정하자. 해당 과정은 머신 단위로 병렬이고 각 머신당 reduce 함수는 순차적으로 호출된다.

예제의 데이터로 따진다면

m1 에 (감자, (1)), (고양이, (2)), (사슴, (1))

m2 에 (토끼, (1)), (호랑이, (1, 1))

와 같이 배치 되고

reduce 함수의 결과로 (감자, 1), (고양이, 2), (사슴, 1), (토끼, 1), (호랑이, 2) 와 같은 데이터가 최종 생성되게 된다.

아웃풋 라이터(Output writer)

최종적으로 리듀스 함수의 결과를 아웃풋 라이터가 분산 파일 시스템에 저장하게 된다.

상당히 길게 쓰여진 것같은데 다시 요약하면

map 과정은 주어진 데이터를 key로 분리하여 reduce 단으로 넘기고

key로 분리된 데이터를 정렬, 그룹화하여 reduce 단에선 해당 그룹에 대한 일관적인 처리를 하게 된다.

참고

<http://en.wikipedia.org/wiki/MapReduce>

출처: <http://sonsworld.tistory.com/6?category=414618> [SonsWorld!]

출처: <<http://sonsworld.tistory.com/6?category=414618>>

리눅스 환경에서 처리하기

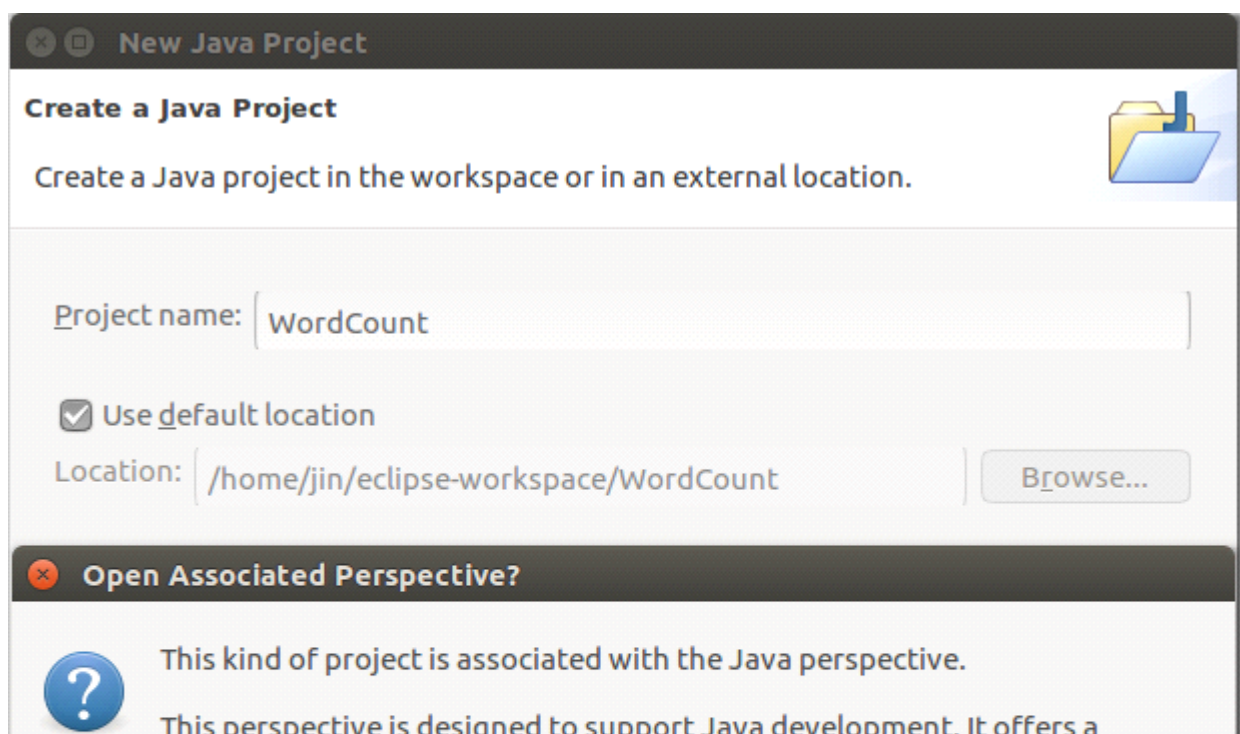
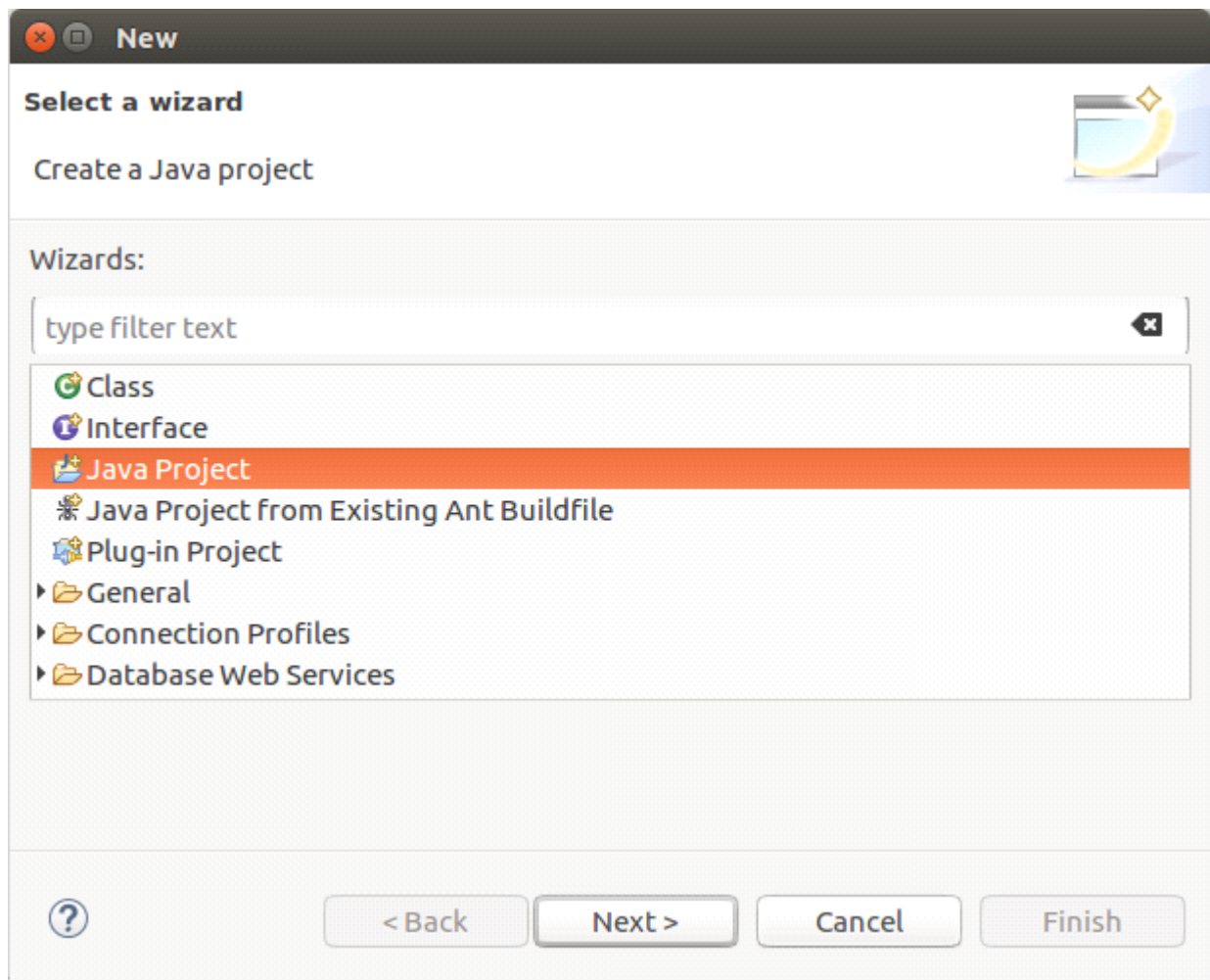
2017년 11월 24일 금요일 오전 10:27

환경설정

2017년 10월 5일 목요일 오후 3:28

프로젝트 만들기

2017년 10월 5일 목요일 오후 2:07





This kind of project is associated with the Java perspective.

This perspective is designed to support Java development. It offers a Package Explorer, a Type Hierarchy, and Java-specific navigation actions.

Do you want to open this perspective now?

☐ Remember my decision

No

Open Perspective

라이브러리 추가

2017년 10월 5일 목요일 오후 2:08

Properties for WordCount

type filter text

Resource Builders Coverage

Java Build Path

Java Code Style

Java Compiler

Java Editor

javadoc Location

Java Build Path

Source Projects Libraries Order and Export

JARs and class folders on the build path:

JRE System Library [JavaSE-1.8]

Add JARs...

Add External JARs...

Add Variable...

JAR Selection

최근 홈 바탕 화면 다운로드 문서 비디오 사진 음악 eclipse-installer

usr local hadoop share hadoop common

이름	크기	수정
jdifff		6월
lib		6월
sources		6월
templates		6월
hadoop-common-2.8.1.jar	3.7 MB	6월
hadoop-common-2.8.1-tests.jar	2.4 MB	6월
hadoop-nfs-2.8.1.jar	187.1 kB	6월

usr local hadoop share hadoop mapreduce

이름

jdifff

lib

lib-examples

sources

hadoop-mapreduce-client-app-2.8.1.jar

hadoop-mapreduce-client-common-2.8.1.jar

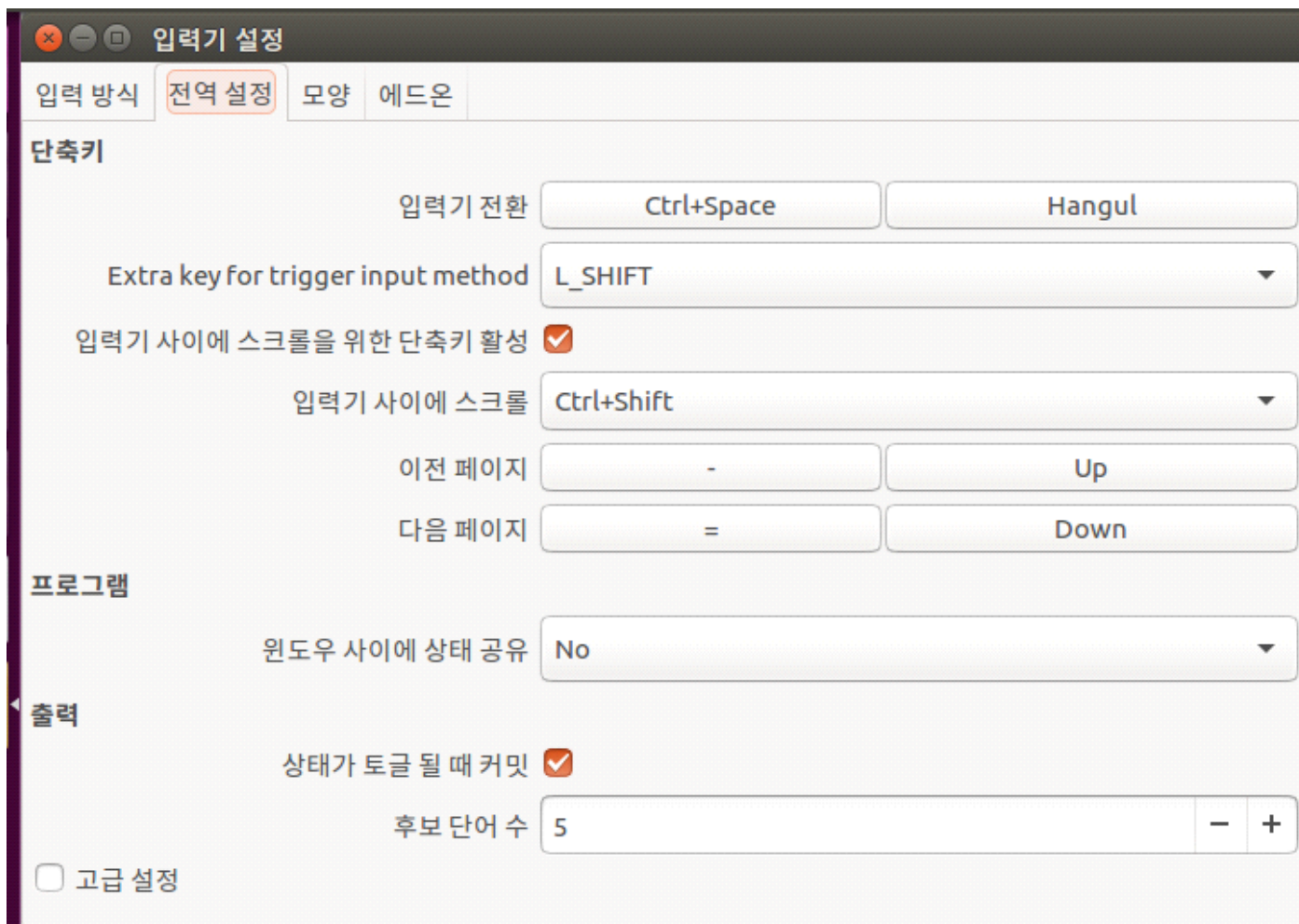
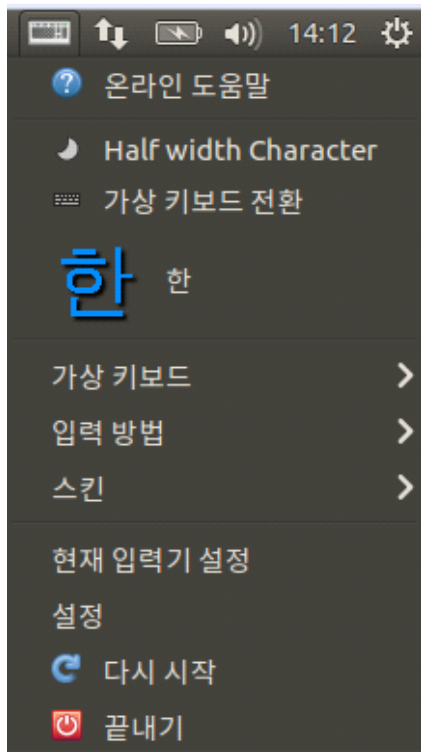
hadoop-mapreduce-client-core-2.8.1.jar

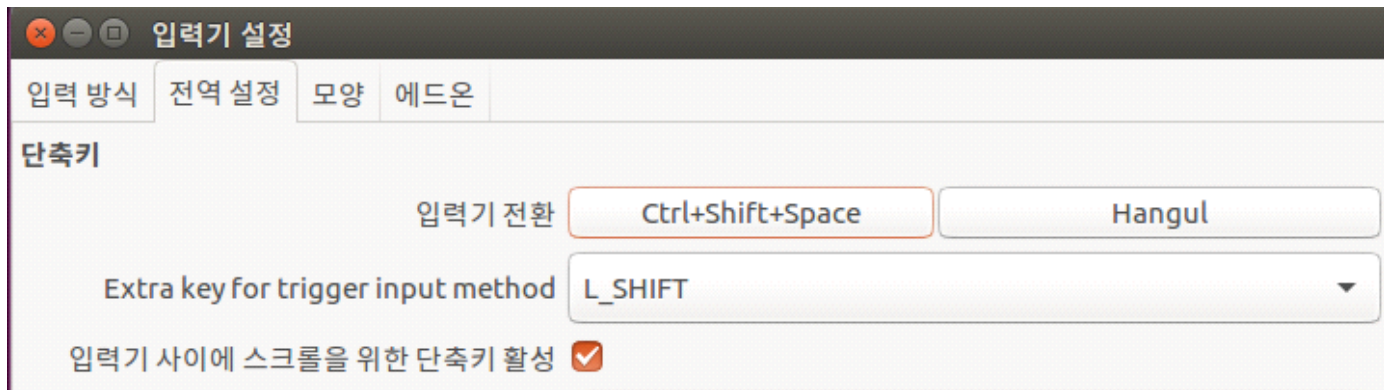
hadoop-mapreduce-client-hs-2.8.1.jar

hadoop-mapreduce-client-hs-plugins-2.8.1.jar

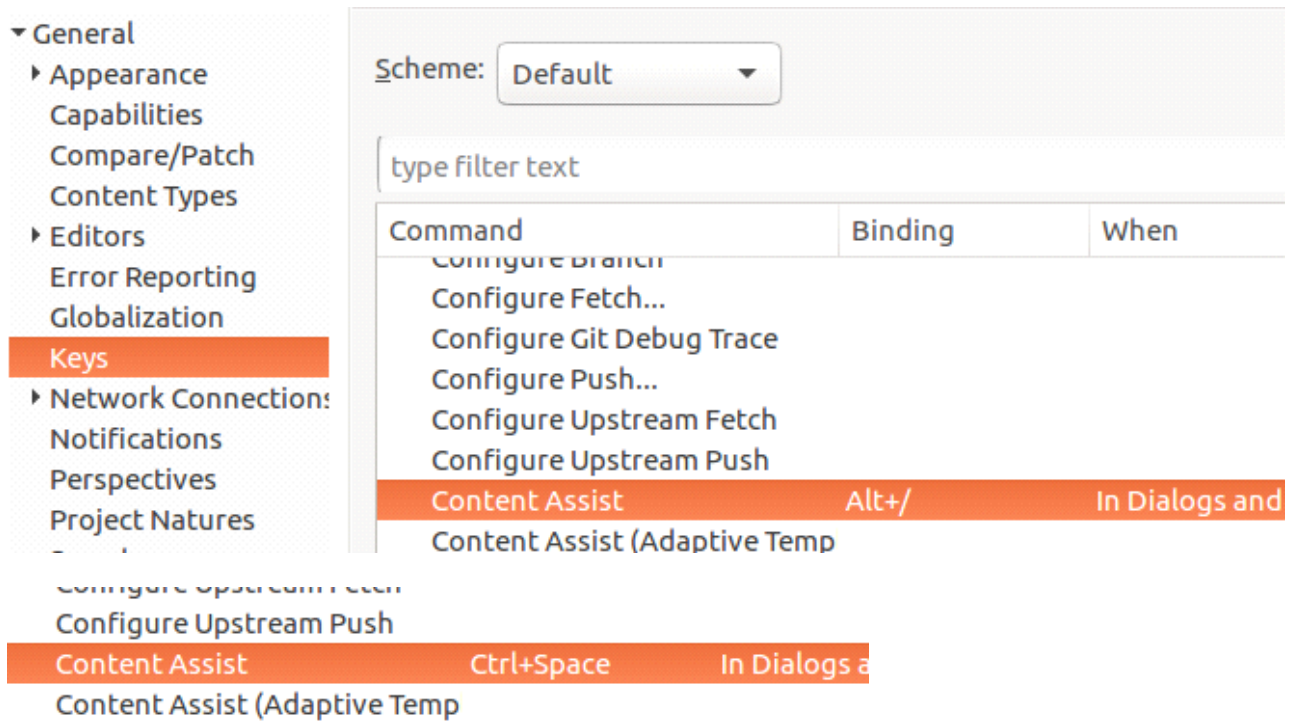
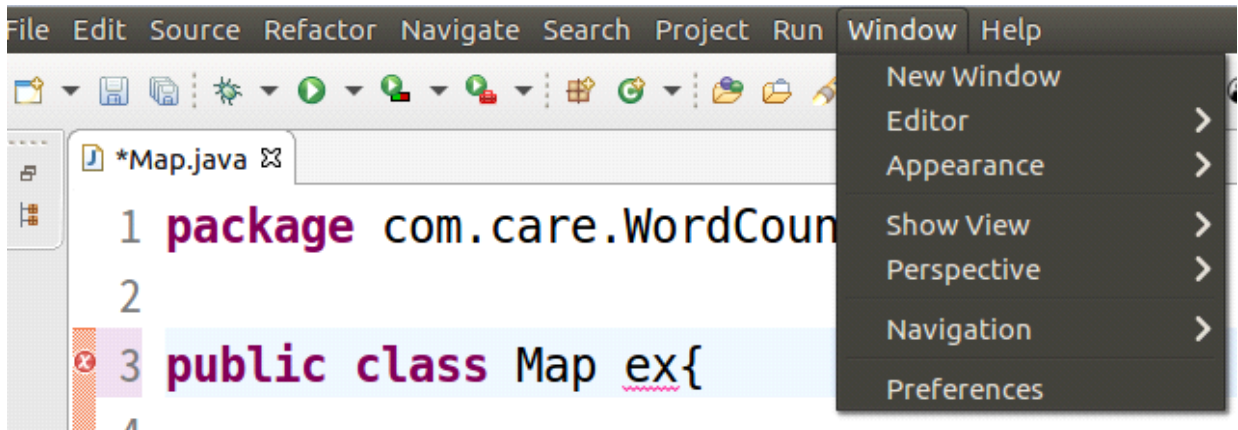
단축키 변경

2017년 10월 5일 목요일 오후 2:10





위처럼 변경



위처럼 변경

소스

2017년 10월 5일 목요일 오후 3:28

Map.java

2017년 10월 5일 목요일 오후 2:15

```
package com.care.WordCount;

import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

public class Map extends MapReduceBase
implements Mapper<LongWritable, Text, Text, IntWritable>{
    public void map(LongWritable key, Text value,
        OutputCollector<Text, IntWritable> output,
        Reporter reporter) throws IOException {
        StringTokenizer st = new StringTokenizer(value.toString().toLowerCase());

        while(st.hasMoreTokens()) {
            output.collect(new Text(st.nextToken()), new IntWritable(1));
        }
    }
}
```

```
package com.care.WordCount;

import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

public class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text,
IntWritable>{
    public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output,
Reporter reporter) throws IOException {
        StringTokenizer st = new StringTokenizer(value.toString().toLowerCase());
        while(st.hasMoreTokens()) {
            output.collect(new Text(st.nextToken()), new IntWritable(1));
        }
    }
}
```

Reduce.java

2017년 10월 5일 목요일 오후 2:48

```
package com.care.WordCount;

import java.io.IOException;
import java.util.Iterator;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;

public class Reduce extends MapReduceBase implements
    Reducer<Text, IntWritable, Text, IntWritable>{
    public void reduce(Text key, Iterator<IntWritable> values,
        OutputCollector<Text, IntWritable> output,
        Reporter reporter) throws IOException {
        int cnt=0;

        while(values.hasNext()) {
            cnt += values.next().get();
        }
        output.collect(key, new IntWritable(cnt));
    }
}
```

```
package com.care.WordCount;

import java.io.IOException;
import java.util.Iterator;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;
```

```
public class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text,
IntWritable>{
    public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text,IntWritable>
output, Reporter reporter) throws IOException {
        int count = 0;
        while(values.hasNext()) {
            count += values.next().get();
        }
        output.collect(key, new IntWritable(count));
    }
}
```

WordCount.java

2017년 10월 5일 목요일 오후 2:49

```
package com.care.WordCount;

import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.mapred.TextInputFormat;
import org.apache.hadoop.mapred.TextOutputFormat;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;
```

```

public class WordCount extends Configured implements Tool{

    @Override
    public int run(String[] args) throws Exception {
        JobConf conf = new JobConf(WordCount.class);
        conf.setJobName("wordcount");
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(Map.class);
        conf.setReducerClass(Reduce.class);
        conf.setInputFormat(TextInputFormat.class);
        conf.setOutputFormat(TextOutputFormat.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));

        JobClient.runJob(conf);
        return 0;
    }

    public static void main(String[] args) throws Exception {
        int exitCode = ToolRunner.run(new WordCount(), args);
        System.exit(exitCode);
    }
}

```

```
package com.care.WordCount;
```

```

import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.mapred.TextInputFormat;
import org.apache.hadoop.mapred.TextOutputFormat;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;

```

```
public class WordCount extends Configured implements Tool{
```

```

    @Override
    public int run(String[] args) throws Exception {
        JobConf conf = new JobConf(WordCount.class);
        conf.setJobName("wordcount");
    }
}

```

```

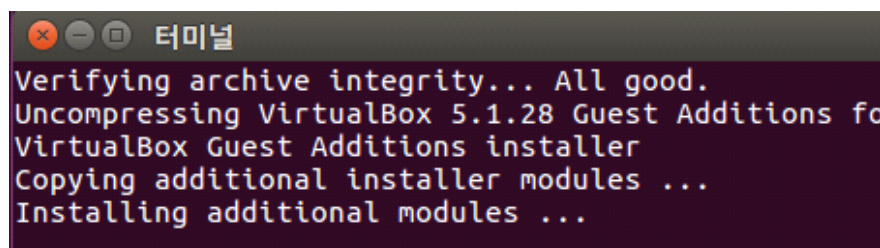
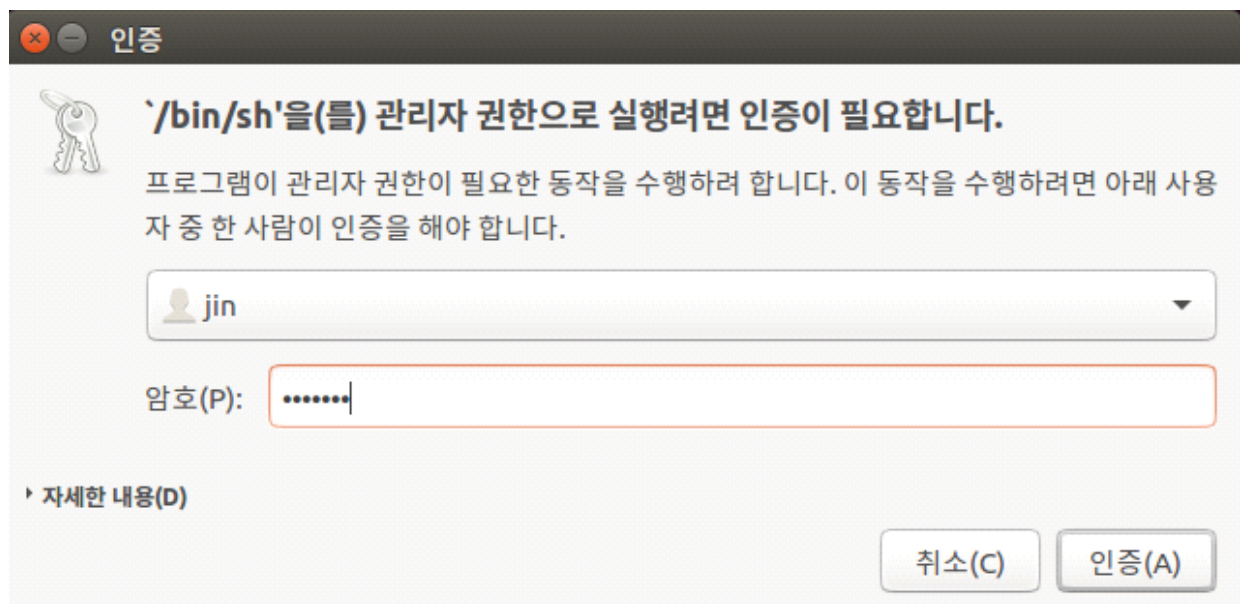
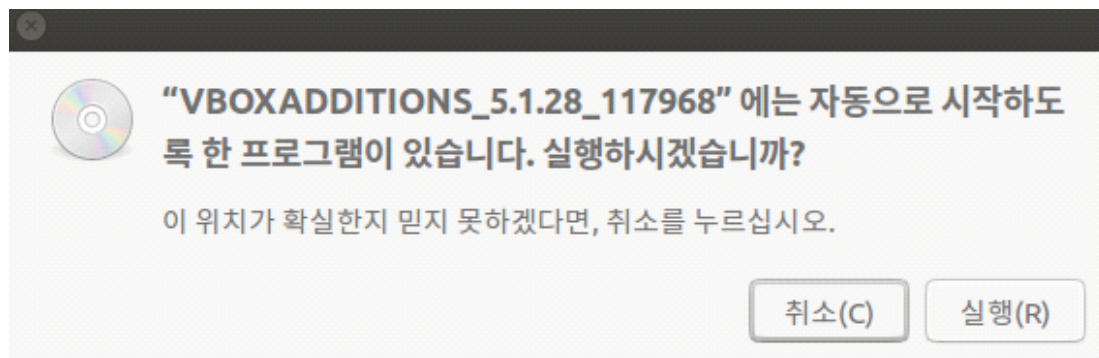
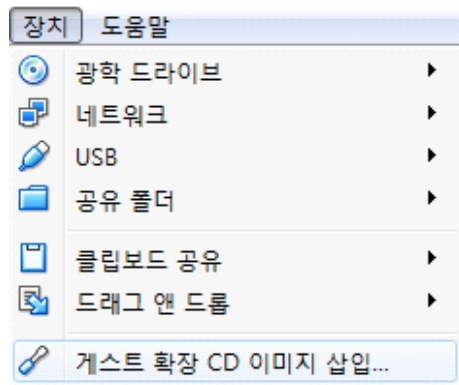
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(Map.class);
        conf.setReducerClass(Reduce.class);
        conf.setInputFormat(TextInputFormat.class);
        conf.setOutputFormat(TextOutputFormat.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));

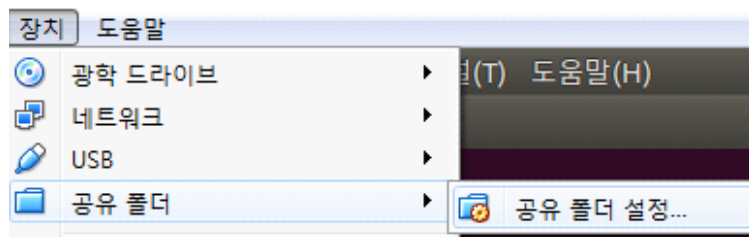
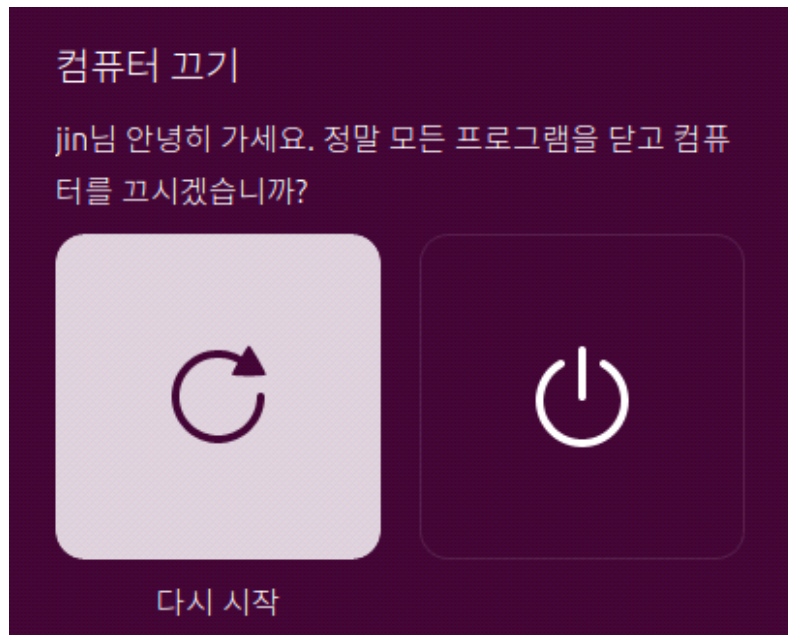
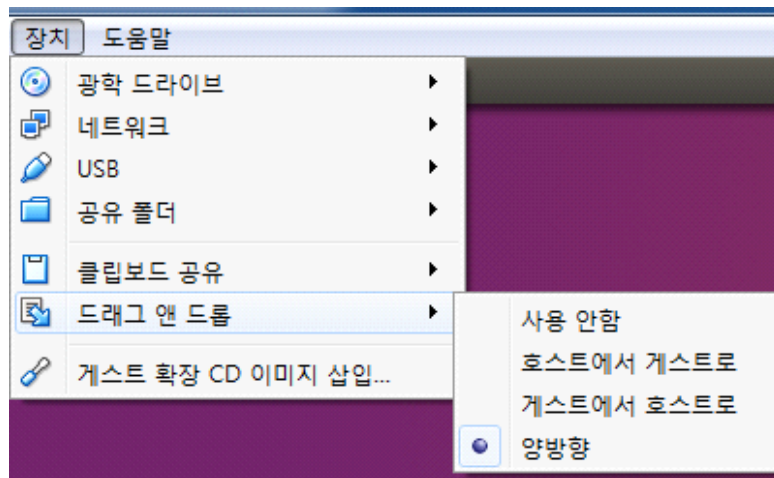
        JobClient.runJob(conf);
        return 0;
    }
    public static void main(String[] args) throws Exception {
        int exitCode = ToolRunner.run(new WordCount(), args);
        System.exit(exitCode);
    }
}

```

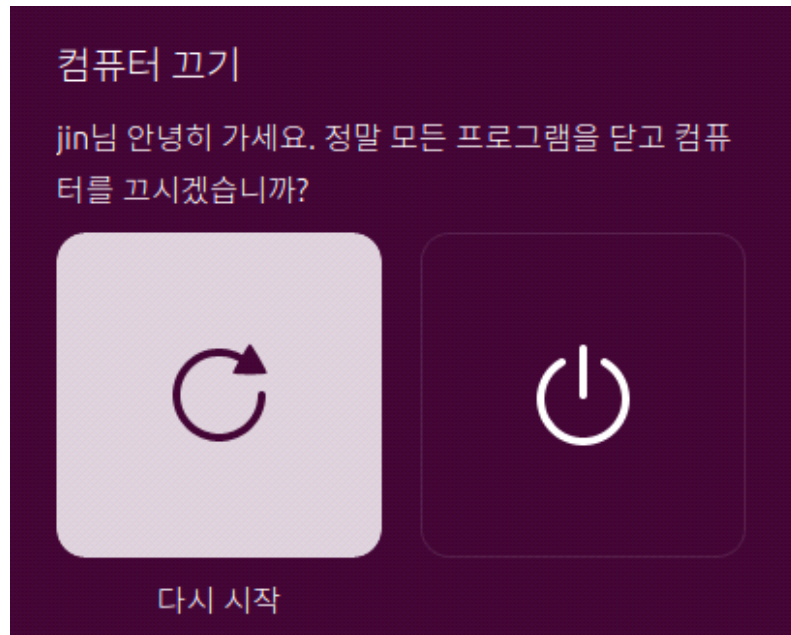
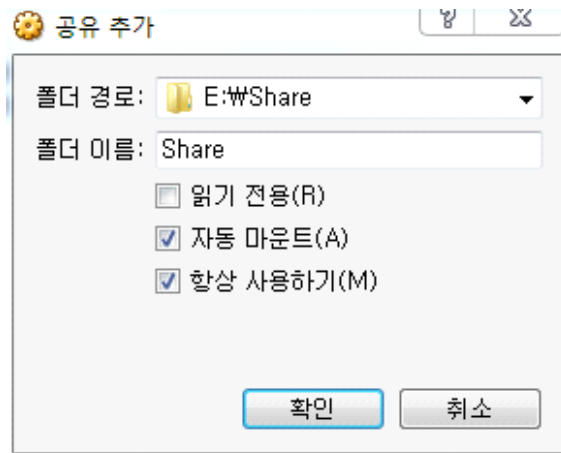
공유폴더 만들기

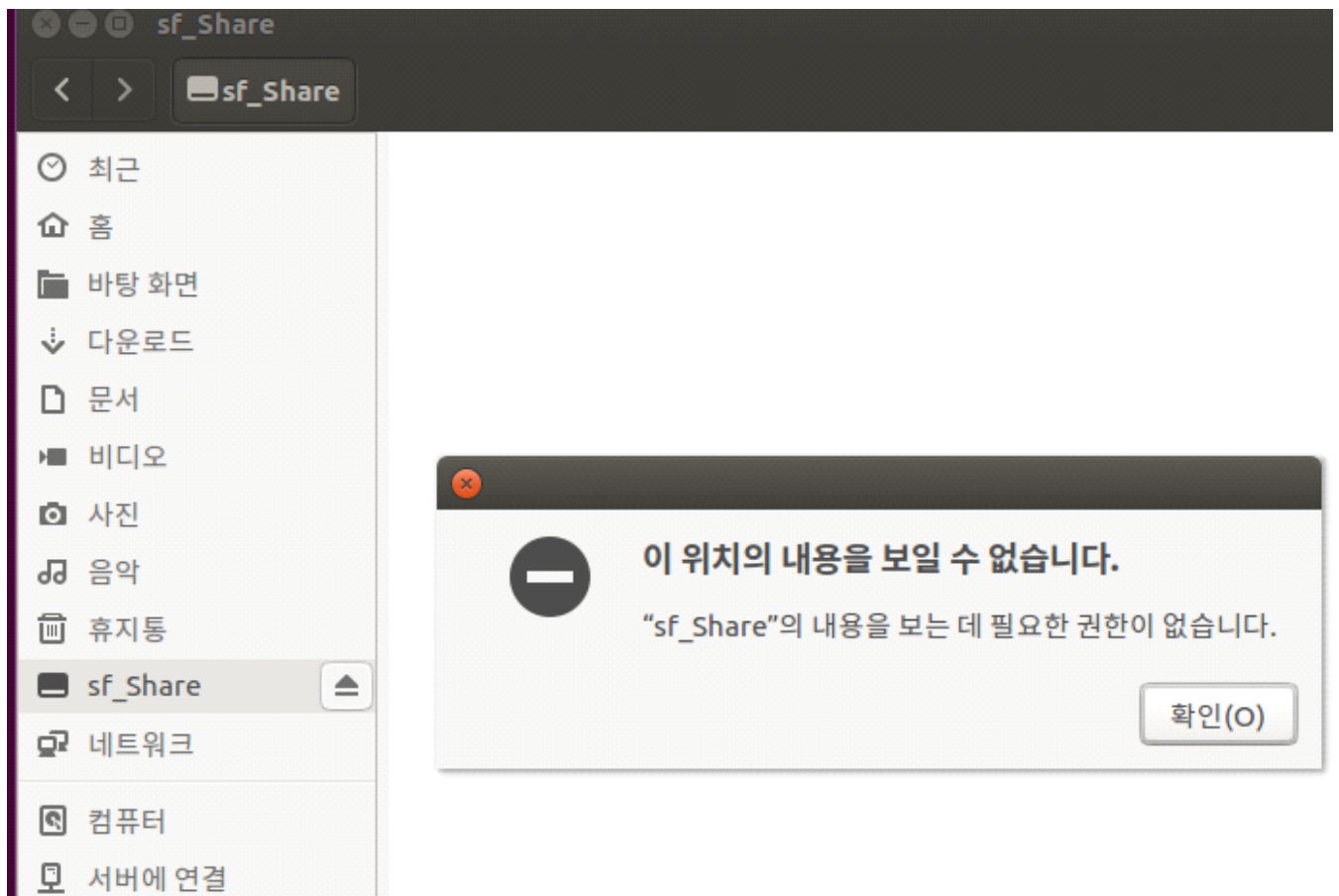
2017년 10월 5일 목요일 오후 3:33



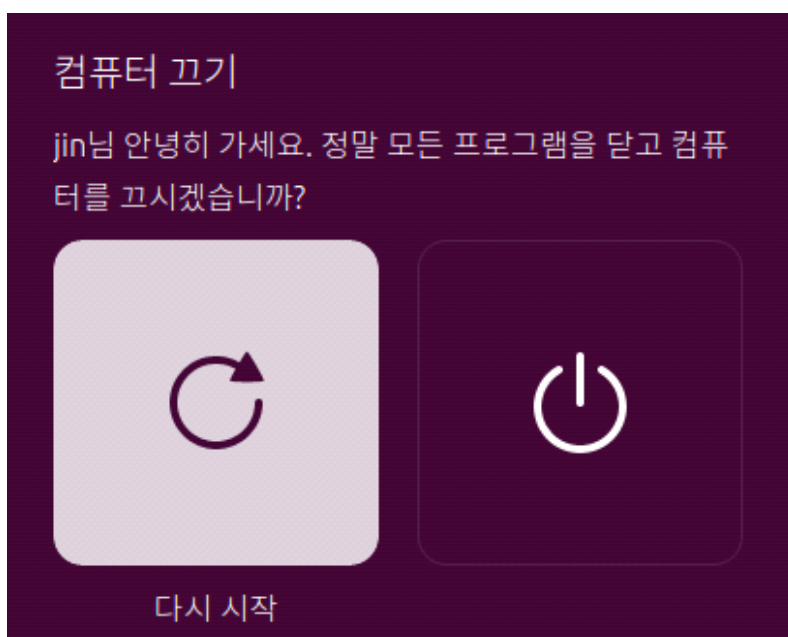


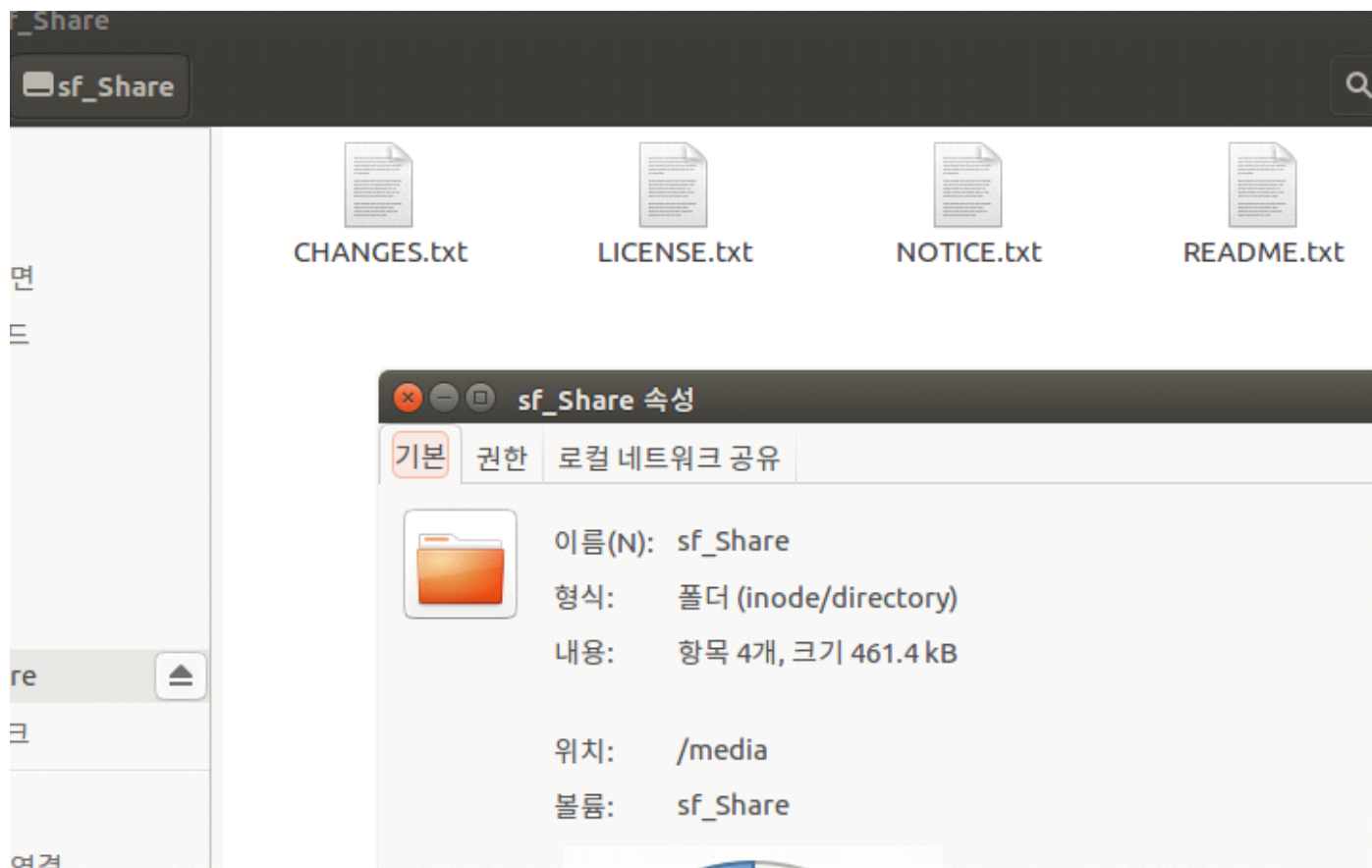
컴퓨터 > Data (E:) > Share





```
jin@jin-VirtualBox:~$ ll /media/
합계 12
drwxr-xr-x  4 root root  4096 10월  5 15:55 ./
drwxr-xr-x 25 root root  4096 10월  3 14:48 ../
drwxr-x---+  2 root root  4096 10월  5 15:53 jin/
drwxrwx---  1 root vboxsf    0 10월  5 15:56 sf_Share/
jin@jin-VirtualBox:~$ sudo gpasswd -a jin vboxsf
[sudo] password for jin:
사용자 jin을(를) vboxsf 그룹에 등록 중
```





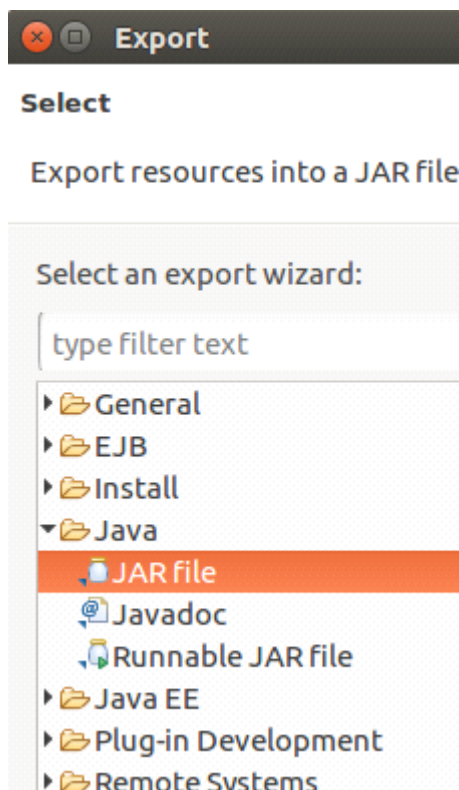
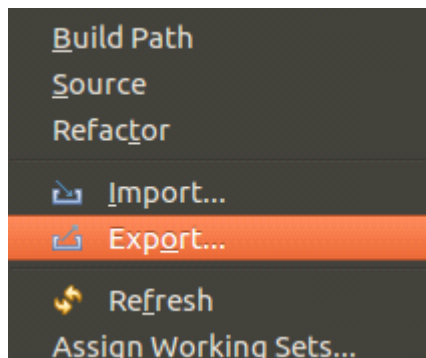
```
jin@jin-VirtualBox:~$ sudo cp /media/sf_Share/CHANGES.txt /usr/local/hadoop/CHANGES.txt
```

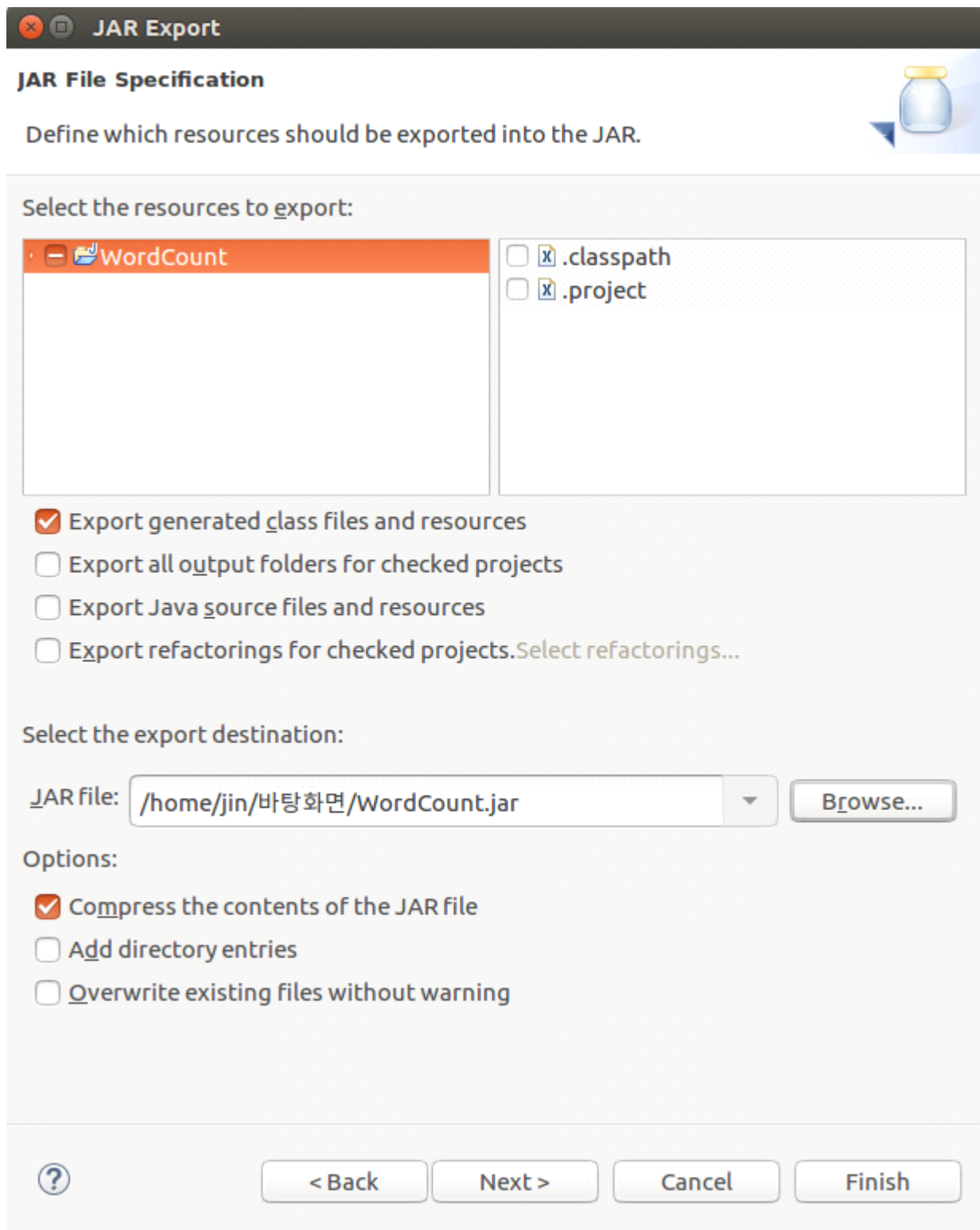
컴파일 및 실행

2017년 10월 5일 목요일 오후 3:41

Jar 생성

2017년 10월 5일 목요일 오후 2:50





```
jin@jin-VirtualBox:~$ cd 바탕화면/  
jin@jin-VirtualBox:~/바탕화면$ ls  
WordCount.jar  
jin@jin-VirtualBox:~/바탕화면$ sudo cp WordCount.jar /usr/local/hadoop  
[sudo] password for jin:
```

컴파일

2017년 10월 5일 목요일 오후 3:29



CHANGES

```
jin@jin-VirtualBox:~/바탕화면$ sudo cp /media/sf_share/*.txt /usr/local/hadoop
```

```
jin@jin-VirtualBox:~$ ssh jinuser@master
jinuser@master's password:
Welcome to Ubuntu 16.04.3 LTS (GNU/Linux 4.10.0-37-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

패키지 87개를 업데이트할 수 있습니다.
0 업데이트는 보안 업데이트입니다.

Last login: Tue Oct 17 21:12:05 2017 from 192.168.56.101
jinuser@jin-VirtualBox:~$ cd $HADOOP_HOME
jinuser@jin-VirtualBox:/usr/local/hadoop$ start-all.sh
```

```
jinuser@jin-VirtualBox:/usr/local/hadoop$ jps
4790 ResourceManager
4619 SecondaryNameNode
6764 Jps
4397 NameNode
```

```
jinuser@jin-VirtualBox:/usr/local/hadoop$ ssh slave
Welcome to Ubuntu 16.04.3 LTS (GNU/Linux 4.10.0-37-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

패키지 87개를 업데이트할 수 있습니다.
0 업데이트는 보안 업데이트입니다.

Last login: Tue Oct 17 21:13:05 2017 from 192.168.56.101
jinuser@jin-VirtualBox:~$ jps
2688 DataNode
4421 Jps
4143 NodeManager
jinuser@jin-VirtualBox:~$ exit
로그아웃
Connection to slave closed.
```

```
jinuser@jin-VirtualBox:/usr/local/hadoop$ hadoop fs -mkdir -p /wordcount/input
```

```
jinuser@jin-VirtualBox:/usr/local/hadoop$ hadoop fs -copyFromLocal /usr/local/hadoop/CHANGES.txt /wordcount/input/
17/10/05 16:18:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
copyFromLocal: /usr/local/hadoop/CHANGES.txt (허가 거부)
```



```
jinuser@jin-VirtualBox:/usr/local/hadoop$ ls -al
합계 604
drwxr-xr-x 10 jinuser hadoop 4096 10월 5 16:15 .
drwxr-xr-x 12 root root 4096 10월 3 14:59 ..
-rwxr-x--- 1 root root 446615 10월 5 16:15 CHANGES.txt
-rw-r--r-- 1 jinuser hadoop 99253 6월 2 15:24 LICENSE.txt
-rw-r--r-- 1 jinuser hadoop 15915 6월 2 15:24 NOTICE.txt
-rw-r--r-- 1 jinuser hadoop 1366 6월 2 15:24 README.txt
-rw-r--r-- 1 root root 3185 10월 5 15:30 WordCount.jar
drwxr-xr-x 2 jinuser hadoop 4096 6월 2 15:24 bin
```

위처럼 실행해 보았더니 허가 거부가 나왔다. 권한을 확인해 봤더니 CHANGES.txt와 WordCount.jar가 root 이다.

```
jinuser@jin-VirtualBox:/usr/local/hadoop$ sudo chown -R jinuser:hadoop /usr/local/hadoop
jinuser@jin-VirtualBox:/usr/local/hadoop$ ls -al
합계 604
drwxr-xr-x 10 jinuser hadoop 4096 10월 5 16:15 .
drwxr-xr-x 12 root root 4096 10월 3 14:59 ..
-rwxr-x--- 1 jinuser hadoop 446615 10월 5 16:15 CHANGES.txt
-rw-r--r-- 1 jinuser hadoop 99253 6월 2 15:24 LICENSE.txt
-rw-r--r-- 1 jinuser hadoop 15915 6월 2 15:24 NOTICE.txt
-rw-r--r-- 1 jinuser hadoop 1366 6월 2 15:24 README.txt
-rw-r--r-- 1 jinuser hadoop 3185 10월 5 15:30 WordCount.jar
drwxr-xr-x 2 jinuser hadoop 4096 6월 2 15:24 bin
```

```
jinuser@jin-VirtualBox:/usr/local/hadoop$ hadoop jar WordCount.jar com.care.WordCount.WordCount /wordcount/input /wordcount/output/
```

```
java.lang.Exception: java.io.IOException: Type mismatch in key from map: expected org.apache.hadoop.io.LongWritable, received org.apache.hadoop.io.Text
    at org.apache.hadoop.mapred.LocalJobRunner$Job.runTasks(LocalJobRunner.java:489)
    at org.apache.hadoop.mapred.LocalJobRunner$Job.run(LocalJobRunner.java:549)
Caused by: java.io.IOException: Type mismatch in key from map: expected org.apac
```

위처럼 실행했지만 에러 발생 아마도 타입 문제인듯....

확인결과 오타

```
conf.setJobName("wordcount");
conf.setOutputKeyClass(Text.class);
```

두 번째 줄의 내용을 빼먹었음.

```
jinuser@jin-VirtualBox:/usr/local/hadoop$ hadoop jar WordCount.jar com.care.WordCount.WordCount /wordcount/input /wordcount/output/
17/10/05 16:53:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/10/05 16:53:25 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
17/10/05 16:53:25 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
17/10/05 16:53:25 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://master:62350/wordcount/output already exists
    at org.apache.hadoop.mapred.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:131)
```


재 실행 결과 기존의 output 폴더가 존재하여 문제 발생

```
jinuser@jin-VirtualBox:/usr/local/hadoop$ hadoop fs -rmrdir /wordcount/output
17/10/05 16:56:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

위처럼 폴더 삭제 후 다시 분석하면 정상

```
sudo cp /media/sf_share/*.txt /usr/local/hadoop
```

```
ssh jinuser@master
```

```
cd $HADOOP_HOME
```

```
start-all.sh
```

```
jps
```

```
ssh slave
```

```
jps
```

```
exit
```

```
hadoop fs -mkdir -p /wordcount/input
```

```
sudo chown -R jinuser:hadoop /usr/local/hadoop
```

```
hadoop fs -copyFromLocal /usr/local/hadoop/*.txt /wordcount/input/
```

```
hadoop jar WordCount.jar com.care.wordcount.WordCnt /wordcount/input /wordcount/output
```

```
hadoop fs -ls /wordcount/output
```

```
hadoop fs -cat /wordcount/output/part-00000
```

[참고] 디렉토리 생성 문제 hadoop - Name node is in safe mode. Not able to leave - Stack Overflow

2018년 2월 20일 화요일 오후 12:42

hadoop dfsadmin -safemode leave

[Stack Overflow](#)

1. [Questions](#)
2. [Developer Jobs](#)
3. [Tags](#)
4. [Users](#)

[current community](#)

- [help chat](#)
- [Stack Overflow](#)
- [Meta Stack Overflow](#)

your communities

[Sign up](#) or [log in](#) to customize your list.

[more stack exchange communities](#)

[company blog](#)

- [Tour](#) Start here for a quick overview of the site
- [Help Center](#) Detailed answers to any questions you might have
- [Meta](#) Discuss the workings and policies of this site
- [About Us](#) Learn more about Stack Overflow the company
- [Business](#) Learn more about hiring developers or posting ads with us

[Log In](#) [Sign Up](#)

Join Stack Overflow to learn, share knowledge, and build your career.

[Email Sign Up](#) or sign in with

Google

Facebook

[Name node is in safe mode. Not able to leave](#)

[Ask Question](#)

```
93  root# bin/hadoop fs -mkdir t
24  mkdir: org.apache.hadoop.hdfs.server.namenode.SafeModeException: Cannot
    create directory /user/root/t. Name node is in safe mode.
    not able to create anything in hdfs
    I did
    root# bin/hadoop fs -safemode leave
    But showing
    safemode: Unknown command
    what is the problem?
```

[Solution](#)

[hadoop mapreduce](#)

[shareimprove this question](#) [edited Oct 24 '15 at 7:28](#)

asked Apr 4 '13 at 5:34



[Unmesha SreeVeni](#)

2,55763563

- [hadoop.apache.org/docs/stable/hdfs_user_guide.html#Safemode](#) might help. – [Brenden Brown](#) Apr 4 '13 at 5:39
- Hey Sree, please don't put [solved] in the title. This clutters up the questions and makes it tougher to parse them. Instead, if there's an answer that solves your problem, mark it as the accepted answer by clicking the checkbox next to the answer. This changes the color on the main page too so people know it's answered. Hope this helps! – [jmort253](#) Nov 16 '13 at 5:35

- ok jmort253.sorry i have included [solved] in every question. ok i vil remove. As i edited for all question my privilege is gone.now i have the permission to edit that only by tomorrow.is that fine if i edit tomorrow – [Unmesha SreeVeni Nov 16 '13 at 5:38](#)
- Hi Sree, I already fixed them for you. You're all set! :D – [jmort253 Nov 16 '13 at 5:57](#)
- ok.Thxs for correcting my mistake :) – [Unmesha SreeVeni Nov 16 '13 at 6:12](#)

| [show 1 more comment](#)

8 Answers

[active](#) [oldest](#) [votes](#)

165
accepted

In order to forcefully let the namenode leave safemode, following command should be executed:

```
bin/hadoop dfsadmin -safemode leave
```

You are getting Unknown command error for your command as -safemode isn't a sub-command for hadoop fs, but it is of hadoop dfsadmin.

Also after the above command, I would suggest you to once run `hadoop fsck` so that any inconsistencies crept in the hdfs might be sorted out.

Update:

Use `hdfs` command instead of `hadoop` command for newer distributions. The `hadoop` command is being deprecated:

```
hdfs dfsadmin -safemode leave
```

`hadoop dfsadmin` has been deprecated and so is `hadoop fs` command, all hdfs related tasks are being moved to a separate command `hdfs`.

[shareimprove this answer](#) [edited Apr 4 '16 at 13:34](#) answered Apr 4 '13 at 6:04



[slm](#)

7,20985367



[Amar](#)

8,99133966

- 1
actually why this displays 'namenode is in safemode' – [Unmesha SreeVeni Apr 4 '13 at 8:15](#)
- 1
Basically namenode enters in safe mode in unusual situations, for example when disk is full, also in the start-up phase. Read more here;
[hadoop.apache.org/docs/stable/hdfs_user_guide.html#Safemode](#) – [Amar Apr 4 '13 at 11:30](#)
- 1
I'm using Hadoop 2.0.0-cdh4.1.1 When I ran the `hadoop dfsadmin` command, it gave me this: `DEPRECATED: Use of this script to execute hdfs command is deprecated. Instead use the hdfs command for it. Safe mode is OFF` ... but still ran. Seems the suggestion by @user3316017 below is the recommended syntax. – [CJBS Feb 27 '14 at 0:14](#)
- @TimWolla why rollback my edit?? It was useful information that some people might miss given this response is incorrectly marked as the answer. – [MikeKulls Mar 17 '14 at 23:56](#)
- Updated my answer as per new distributions, if anyone can help us out in establishing the exact version of apache hadoop since when these deprecations have come into picture, it would be great. – [Amar Mar 19 '14 at 19:03](#)

| [show 2 more comments](#)

22

try this, it will work

```
sudo -u hdfs hdfs dfsadmin -safemode leave
```

[shareimprove this answer](#) answered Jan 3 '15 at 10:53



[Wesam Na](#)

684712

- this command worked for me as it can require superuser privilege..thanks @wesam – [nik Dec 14 '17 at 11:35](#)
- this work if the Access denied for user root – [hisi Jan 23 at 16:02](#)

[add a comment](#) |

19

The Command did not work for me but the following did

```
hdfs dfsadmin -safemode leave
```

I used the `hdfs` command instead of the `hadoop` command.
Check out <http://ask.gopivotal.com/hc/en-us/articles/200933026-HDFS-goes-into-readonly-mode-and-errors-out-with-Name-node-is-in-safe-mode-> link too

[shareimprove this answer](#) edited Feb 23 '14 at 12:04

answered Feb 23 '14 at 11:39



[Volker](#)

4,023¹¹227



[Kishan B](#)

53768

- Doc link on Hadoop Safe Mode: hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/ – [CJBS Feb 27 '14 at 0:27](#)

[add a comment](#) |

5

If you use Hadoop version 2.6.1 above, while the command works, it complains that its deprecated. I actually could not use the `hadoop dfsadmin -safemode leave` because I was running Hadoop in a Docker container and that command magically fails when run in the container, so what I did was this. I checked doc and found `dfs.safemode.threshold.pct` in documentation that says

Specifies the percentage of blocks that should satisfy the minimal replication requirement defined by `dfs.replication.min`. Values less than or equal to 0 mean not to wait for any particular percentage of blocks before exiting safemode. Values greater than 1 will make safe mode permanent.

so I changed the `hdfs-site.xml` into the following (In older Hadoop versions, apparently you need to do it in `hdfs-default.xml`:

```
<configuration>
  <property>
    <name>dfs.safemode.threshold.pct</name>
    <value>0</value>
  </property>
</configuration>
```

[shareimprove this answer](#) answered Nov 19 '15 at 9:46



[ambodi](#)

3,547¹²218

[add a comment](#) |

2

Namenode enters into safemode when there is shortage of memory. As a result the HDFS becomes readable only. That means one can not create any additional directory or file in the HDFS. To come out of the safemode, the following command is used:

```
hadoop dfsadmin -safemode leave
```

If you are using cloudera manager:

```
go to >>Actions>>Leave Safemode
```

But it doesn't always solve the problem. The complete solution lies in making some space in the memory. Use the following command to check your memory usage.

```
free -m
```

If you are using cloudera, you can also check if the HDFS is showing some signs of bad health. It probably must be showing some memory issue related to the namenode. Allot more memory by following the options available. I am not sure what commands to use for the same if you are not using cloudera manager but there must be a way. Hope it helps! :)

[shareimprove this answer](#) answered Sep 24 '14 at 6:15



[Amitesh Ranjan](#)

467178

[add a comment](#) |

1 Run the command below using the HDFS OS user to disable safe mode:
`sudo -u hdfs hadoop dfsadmin -safemode leave`

[shareimprove this answer](#) [edited Aug 23 '17 at 6:03](#) answered Aug 23 '17 at 5:43



[Stephen Rauch](#)

17.8k142440



[Mayank Gupta](#)

214

[add a comment](#) |

1 **Try this**
`sudo -u hdfs hdfs dfsadmin -safemode leave`
check status of safemode
`sudo -u hdfs hdfs dfsadmin -safemode get`
If it is still in safemode ,then one of the reason would be not enough space in your node, you can check your node disk usage using :
`df -h`
if root partition is full, delete files or add space in your root partition and retry first step.

[shareimprove this answer](#) [edited Nov 24 '17 at 11:02](#) answered Nov 21 '17 at 14:24



[Tarun Reddy](#)

212

[add a comment](#) |

0 use below command to turn off the safe mode
`$> hdfs dfsadmin -safemode leave`

[shareimprove this answer](#) answered Sep 14 '17 at 6:47



[Azam Khan](#)

13117

[add a comment](#) |

Your Answer

Sign up or [log in](#)

Sign up using Google

Sign up using Facebook

Sign up using Email and Password

Post as a guest

Name

Email

`<h3>Post as a guest</h3> <div class="form-item"> <table> <tr> <td class="vm"> <div> <label for="display-name">Name</label> <input id="display-name" name="display-name" type="text" size="30" maxlength="30" value="" tabindex="105"> </div> <div> <label for="m-address">Email</label> <input id="m-address" name="m-address" type="email" size="30" maxlength="100" value="" tabindex="106" placeholder="required, but never shown" /> </div> </td> </tr> </table> </div>`

[discard](#)

By posting your answer, you agree to the [privacy policy](#) and [terms of service](#).

Not the answer you're looking for? Browse other questions tagged [hadoop](#) [mapreduce](#) or [ask your own question](#).

asked 4 years, 10 months ago

viewed 133,332 times

active [2 months ago](#)

[Looking for a job?](#)

- [Software Engineer](#)
[Datadog](#)New York, NY
[Remote](#)
[gopython](#)
[dismiss this job](#)
- [소프트웨어 엔지니어 - 데이터 \(신입/경력\)](#)
[DEVSISTERS Corp.](#)Seoul, South Korea
[pythonscala](#)
[dismiss this job](#)
- [데이터 과학자](#)
[DEVSISTERS Corp.](#)Seoul, South Korea
[pythonr](#)
[dismiss this job](#)
- [왓차 - 웹 서버/프론트 엔지니어](#)
[Frograms](#)Seoul, South Korea
[reactjsruby-on-rails](#)
[dismiss this job](#)

Linked

[0](#)

[cant get past the error: "mkdir: Cannot create directory /user/hadoop. Name node is in safe mode."](#)

[4](#)

[Hadoop: Cannot delete a directory. Name node is in safe mode](#)

[2](#)

[Name node is in safe mode](#)

[2](#)

[Why is Google Dataproc HDFS Name Node in Safemode?](#)

[0](#)

[How to return to the local user in shell script?](#)

[0](#)

[Error in creating database in Hive](#)

[0](#)

[Unable to start Hive via Terminal on Cloudera CDH5.8 VM - Bad health,Name Node is in Safe Mode](#)

[1](#)

[Name node is in safe mode. Not able to leave. Hadoop 2.6.0](#)

[0](#)

[Hadoop load CSV file into HDFS - error Name node is in safe mode](#)

[0](#)

[Why is my test cluster running in safe mode?](#)

[see more linked questions...](#)

Related

[0](#)

[how to open hdfs://localhost:54310/user/hduser/](#)

[1](#)

[HDFS Safe mode issue](#)

[0](#)

[Namenode not coming out of safemode](#)

[4](#)

[Hadoop: Cannot delete a directory. Name node is in safe mode](#)

[1](#)

[rmr: Cannot delete /output. Name node is in safe mode](#)

[2](#)

[Name Node is in safe mode. Cannot be resolved](#)

[2](#)

[Name node is in safe mode](#)

[0](#)

[Hadoop load CSV file into HDFS - error Name node is in safe mode](#)

[0](#)

[Name node never leaves safe mode eventhough none of the files are corrupted and jps shows name node twice](#)

[0](#)

Cannot create directory in hdfs NameNode is in safe mode

Hot Network Questions

- [How to properly reject an intimate suggestion?](#)
- [Why are there both sharps and flats?](#)
- [the doubled "to" after "get used"](#)
- [Using the ENTER key as a delimiter in an itemized list](#)
- [Did the UK homicide rate go up after handgun ownership was banned?](#)
- [Why does the full moon appear?](#)
- [Would drawing darts from a "dart quiver" use up your free item interaction?](#)
- [Are there any languages that only allow CV syllables?](#)
- [How and why did T'Challa have the Black Panther power already?](#)
- [How to respond to "You should have skipped college" in an interview?](#)
- [Looping through table to get single intersection from N>2 geometries using PostGIS?](#)
- [Were there 18 school shootings in the US from Jan 1 to Feb 14, 2018?](#)
- [What is the world's oldest joke?](#)
- [Escaping tree well](#)
- [How would a society isolated by monster attacks get food?](#)
- [How to stop a colleague from talking to me about work when I'm on my lunch break?](#)
- [Reason for an exterior control panel on a space ship](#)
- [How would people in the 17th century react to modern day clothing](#)
- [How did this capacitor survive having a nail driven through it?](#)
- [Why does it make any physical sense for a body to have negative potential energy?](#)
- [How can I make a "meeting in VR" less dumb?](#)
- [Why is `char\[\]\[\] = {{...}, {...}}` not possible if explicitly given a multidimensional array?](#)
- [How can I find all files I do not have write access to in specific folder?](#)
- [Is a Manual of Golems' spell slot restriction covered by multiclassed spell slots?](#)

[question feed](#)

<div></div>

Stack Overflow

- [Questions](#)
- [Jobs](#)
- [Developer Jobs Directory](#)
- [Salary Calculator](#)
- [Help](#)
- [Mobile](#)

Stack Overflow

Business

- [Talent](#)
- [Ads](#)
- [Enterprise](#)

Company

- [About](#)
- [Press](#)
- [Work Here](#)
- [Legal](#)
- [Privacy Policy](#)
- [Contact Us](#)

Stack Exchange

Network

- [Technology](#)

- [Life / Arts](#)
- [Culture / Recreation](#)
- [Science](#)
- [Other](#)
- [Stack Overflow](#)
- [Server Fault](#)
- [Super User](#)
- [Web Applications](#)
- [Ask Ubuntu](#)
- [Webmasters](#)
- [Game Development](#)
- [TeX - LaTeX](#)
- [Software Engineering](#)
- [Unix & Linux](#)
- [Ask Different \(Apple\)](#)
- [WordPress Development](#)
- [Geographic Information Systems](#)
- [Electrical Engineering](#)
- [Android Enthusiasts](#)
- [Information Security](#)
- [Database Administrators](#)
- [Drupal Answers](#)
- [SharePoint](#)
- [User Experience](#)
- [Mathematica](#)
- [Salesforce](#)
- [ExpressionEngine® Answers](#)
- [Stack Overflow em Português](#)
- [Blender](#)
- [Network Engineering](#)
- [Cryptography](#)
- [Code Review](#)
- [Magento](#)
- [Software Recommendations](#)
- [Signal Processing](#)
- [Emacs](#)
- [Raspberry Pi](#)
- [Stack Overflow на русском](#)
- [Programming Puzzles & Code Golf](#)
- [Stack Overflow en español](#)
- [Ethereum](#)
- [Data Science](#)
- [Arduino](#)
- [Bitcoin](#)
- [**more \(28\)**](#)
- [Photography](#)
- [Science Fiction & Fantasy](#)
- [Graphic Design](#)
- [Movies & TV](#)
- [Music: Practice & Theory](#)
- [Worldbuilding](#)
- [Seasoned Advice \(cooking\)](#)
- [Home Improvement](#)
- [Personal Finance & Money](#)
- [Academia](#)
- [Law](#)
- [**more \(16\)**](#)
- [English Language & Usage](#)
- [Skeptics](#)
- [Mi Yodeya \(Judaism\)](#)
- [Travel](#)
- [Christianity](#)
- [English Language Learners](#)

- [Japanese Language](#)
- [Arcade \(gaming\)](#)
- [Bicycles](#)
- [Role-playing Games](#)
- [Anime & Manga](#)
- [Puzzling](#)
- [Motor Vehicle Maintenance & Repair](#)
- **[more \(32\)](#)**
- [MathOverflow](#)
- [Mathematics](#)
- [Cross Validated \(stats\)](#)
- [Theoretical Computer Science](#)
- [Physics](#)
- [Chemistry](#)
- [Biology](#)
- [Computer Science](#)
- [Philosophy](#)
- **[more \(10\)](#)**
- [Meta Stack Exchange](#)
- [Stack Apps](#)
- [API](#)
- [Data](#)
- [Area 51](#)
- [Blog](#)
- [Facebook](#)
- [Twitter](#)
- [LinkedIn](#)

site design / logo © 2018 Stack Exchange Inc; user contributions licensed under [cc by-sa 3.0](#) with [attribution required](#). rev 2018.2.19.29044

<div id="noscript-warning">Stack Overflow works best with JavaScript enabled </div>

<<https://stackoverflow.com/questions/15803266/name-node-is-in-safe-mode-not-able-to-leave>>에서 삽입

==== 윈도우 환경에서 처리하기

2017년 11월 24일 금요일 오전 10:27

환경설정

2017년 11월 24일 금요일 오전 10:29

프로젝트 만들기

2017년 11월 24일 금요일 오전 10:28

New	Java Project
Show In Alt+Shift+W	Project...
Copy Ctrl+C	Package
Copy Qualified Name	Class
Paste Ctrl+V	Interface
Delete Delete	Enum
Import...	Annotation
	Source Folder

Create a Java Project

Create a Java project in the workspace or in an external location.

Project name:

☒ Use default location

Location:

JRE

☒ Use an execution environment JRE:

☐ Use a project specific JRE:

☐ Use default JRE (currently 'jre1.8.0_144') [Configure JREs...](#)

Project layout

☐ Use project folder as root for sources and class files

☒ Create separate folders for sources and class files [Configure default...](#)

Working sets

☐ Add project to working sets

Working sets:

FileZilla 설치

2017년 11월 24일 금요일 오전 10:30

<https://filezilla-project.org/>



Download 클릭

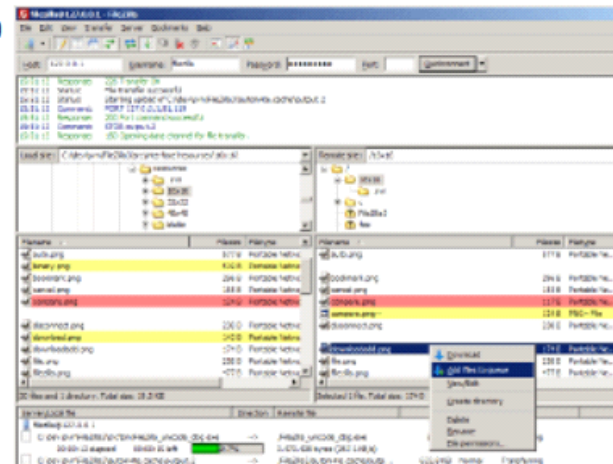
Download FileZilla Client for Windows (64bit)

The latest stable version of FileZilla Client is 3.29.0

Please select the file appropriate for your platform below.

 **Windows (64bit)** 

**Download
FileZilla Client**

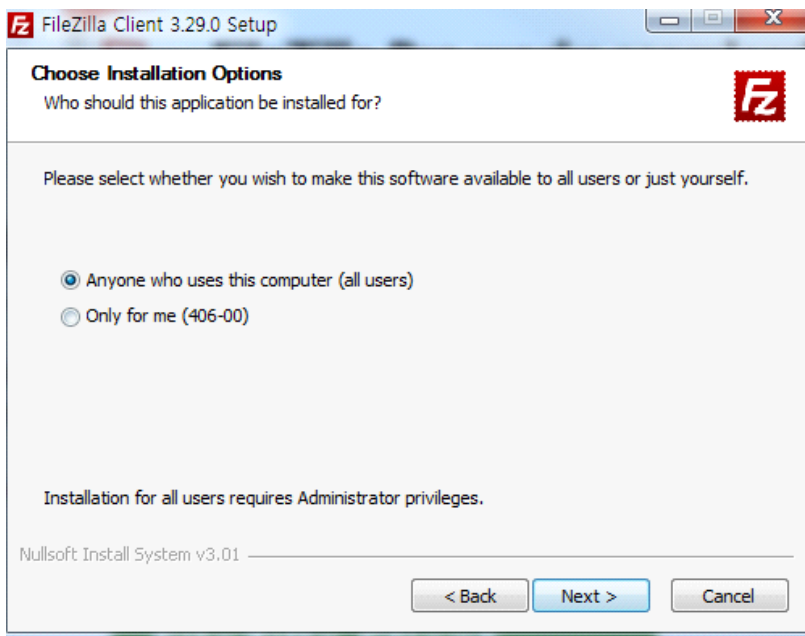


Please select your edition of FileZilla Client

	FileZilla Pro	FileZilla
Standard FTP	Yes	Yes
FTP over TLS	Yes	Yes
SFTP	Yes	Yes
Amazon S3	Yes	-
Priority customer support	Yes	-

Select

Download



기본 값으로 설치





Avast Free Antivirus
보호 기능을 설치에 추가하십시오. 무료입니다.

- ✓ 최신 보호 기능
- ✓ 가벼운 수준의 처리 작업
- ✓ PC 매거진 에디터스 초이스

PC EDITORS' CHOICE ★★★★★

☐ 본인은 Avast Antivirus (을)를 설치하고, 해당 최종 사용자 라이선스 동의 및 개인 정보 보호 정책으로의 동의서에 동의합니다.

Nullsoft Install System v3.01

< Back Next > Cancel

선택 해제하고 실행

Hadoop library 획득 및 추가

2017년 11월 24일 금요일 오전 10:33

호스트(H): 사용자명(U): 비밀번호(W): 포트(P): ▼

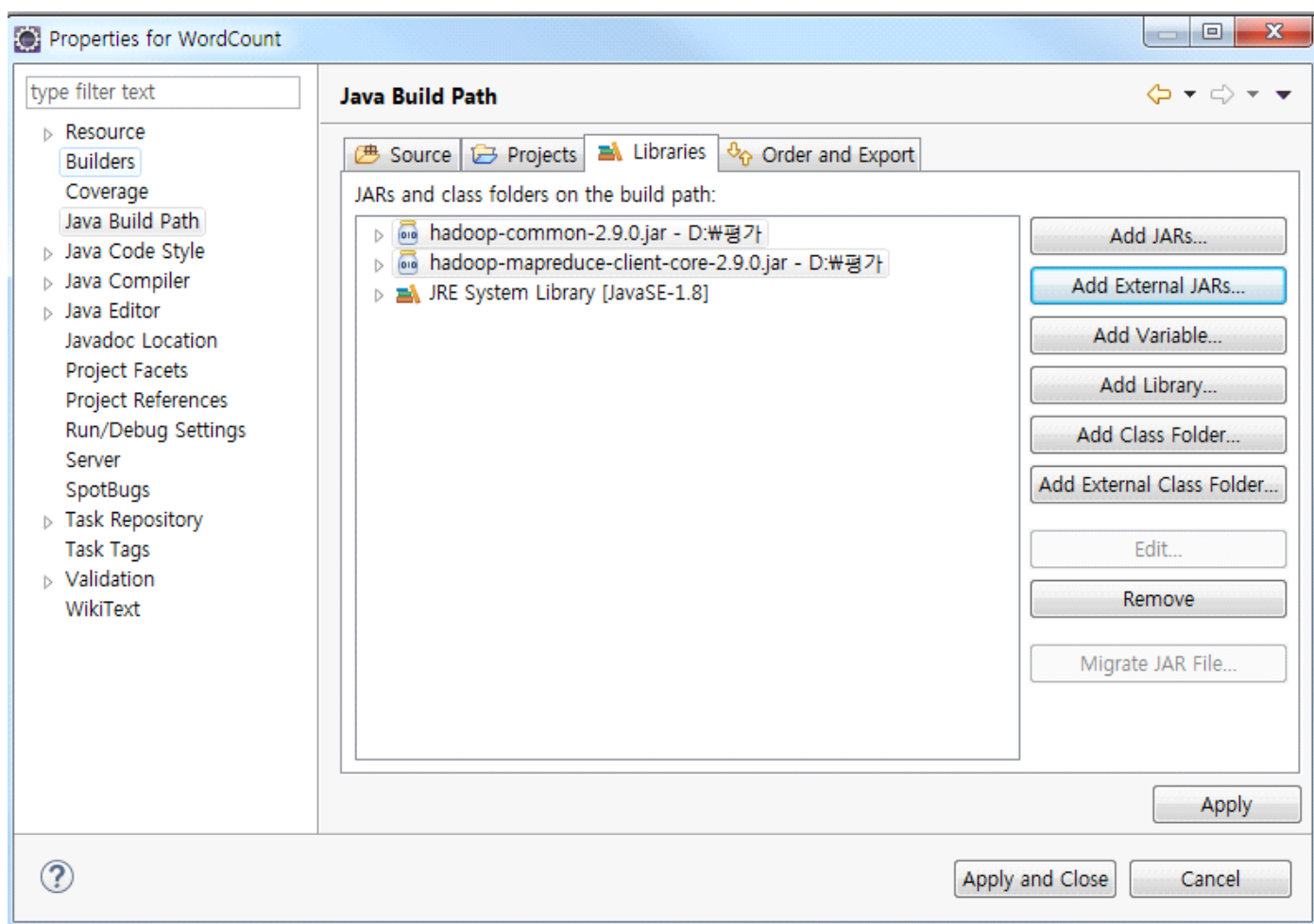
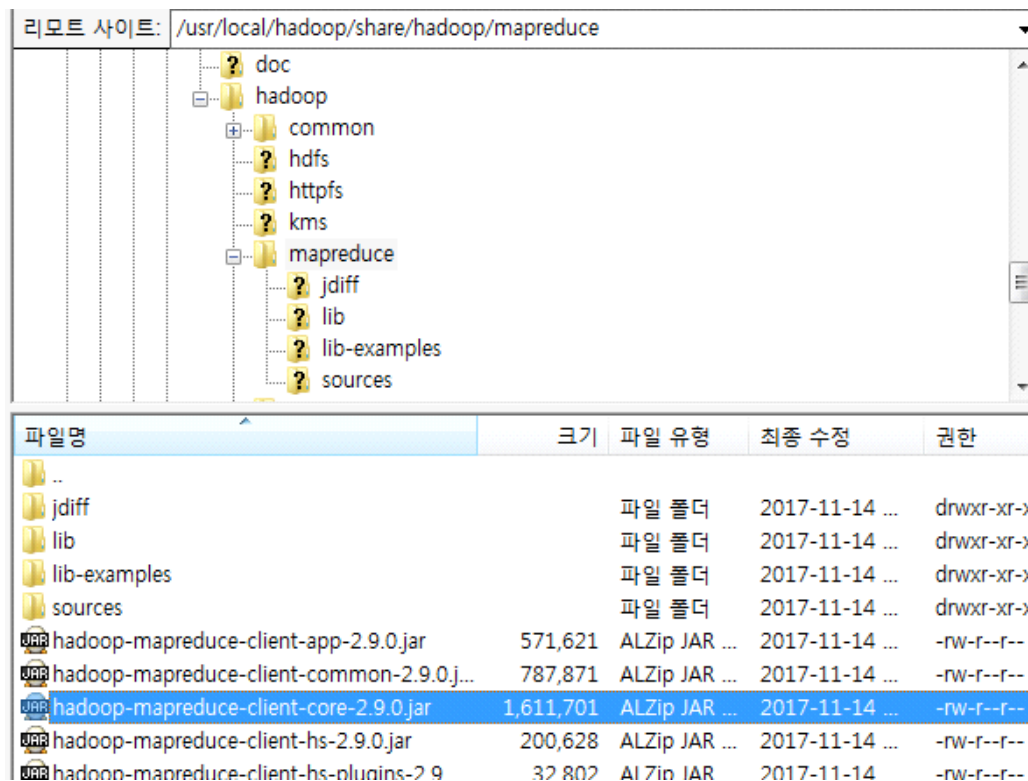
상태: Connected to 192.168.56.101
상태: 디렉터리 목록 조회...
상태: Listing directory /home/manager
상태: "/home/manager" 디렉터리 목록 조회 성공

/usr/local/hadoop/share/hadoop/common

리모트 사이트: /usr/local/hadoop/share/hadoop/common

파일명	크기	파일 유형	최종 수정	권한
..				
jdiff		파일 폴더	2017-11-14 ...	drwxr-xr-x
lib		파일 폴더	2017-11-14 ...	drwxr-xr-x
sources		파일 폴더	2017-11-14 ...	drwxr-xr-x
templates		파일 폴더	2017-11-14 ...	drwxr-xr-x
hadoop-common-2.9.0-tests.jar	2,618,633	ALZip JAR ...	2017-11-14 ...	-rw-r--r--
hadoop-common-2.9.0.jar	3,887,891	ALZip JAR ...	2017-11-14 ...	-rw-r--r--
hadoop-nfs-2.9.0.jar	188,717	ALZip JAR ...	2017-11-14 ...	-rw-r--r--

/usr/local/hadoop/share/hadoop/mapreduce



소스

2017년 10월 5일 목요일 오후 3:28

Map.java

2017년 10월 5일 목요일 오후 2:15

```
package com.care.WordCount;

import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

public class Map extends MapReduceBase
implements Mapper<LongWritable, Text, Text, IntWritable>{
    public void map(LongWritable key, Text value,
        OutputCollector<Text, IntWritable> output,
        Reporter reporter) throws IOException {
        StringTokenizer st = new StringTokenizer(value.toString().toLowerCase());

        while(st.hasMoreTokens()) {
            output.collect(new Text(st.nextToken()), new IntWritable(1));
        }
    }
}
```

```
package com.care.WordCount;

import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

public class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text,
IntWritable>{
    public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output,
Reporter reporter) throws IOException {
        StringTokenizer st = new StringTokenizer(value.toString().toLowerCase());
        while(st.hasMoreTokens()) {
            output.collect(new Text(st.nextToken()), new IntWritable(1));
        }
    }
}
```

Reduce.java

2017년 10월 5일 목요일 오후 2:48

```
package com.care.WordCount;

import java.io.IOException;
import java.util.Iterator;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;

public class Reduce extends MapReduceBase implements
    Reducer<Text, IntWritable, Text, IntWritable>{
    public void reduce(Text key, Iterator<IntWritable> values,
        OutputCollector<Text, IntWritable> output,
        Reporter reporter) throws IOException {
        int cnt=0;

        while(values.hasNext()) {
            cnt += values.next().get();
        }
        output.collect(key, new IntWritable(cnt));
    }
}
```

```
package com.care.WordCount;

import java.io.IOException;
import java.util.Iterator;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;
```

```
public class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text,
IntWritable>{
    public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text,IntWritable>
output, Reporter reporter) throws IOException {
        int count = 0;
        while(values.hasNext()) {
            count += values.next().get();
        }
        output.collect(key, new IntWritable(count));
    }
}
```

WordCount.java

2017년 10월 5일 목요일 오후 2:49

```
package com.care.WordCount;

import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.mapred.TextInputFormat;
import org.apache.hadoop.mapred.TextOutputFormat;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;
```

```

public class WordCount extends Configured implements Tool{

    @Override
    public int run(String[] args) throws Exception {
        JobConf conf = new JobConf(WordCount.class);
        conf.setJobName("wordcount");
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(Map.class);
        conf.setReducerClass(Reduce.class);
        conf.setInputFormat(TextInputFormat.class);
        conf.setOutputFormat(TextOutputFormat.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));

        JobClient.runJob(conf);
        return 0;
    }

    public static void main(String[] args) throws Exception {
        int exitCode = ToolRunner.run(new WordCount(), args);
        System.exit(exitCode);
    }
}

```

```
package com.care.WordCount;
```

```

import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.mapred.TextInputFormat;
import org.apache.hadoop.mapred.TextOutputFormat;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;

```

```
public class WordCount extends Configured implements Tool{
```

```

    @Override
    public int run(String[] args) throws Exception {
        JobConf conf = new JobConf(WordCount.class);
        conf.setJobName("wordcount");
    }
}

```

```

        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(Map.class);
        conf.setReducerClass(Reduce.class);
        conf.setInputFormat(TextInputFormat.class);
        conf.setOutputFormat(TextOutputFormat.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));

        JobClient.runJob(conf);
        return 0;
    }
    public static void main(String[] args) throws Exception {
        int exitCode = ToolRunner.run(new WordCount(), args);
        System.exit(exitCode);
    }
}

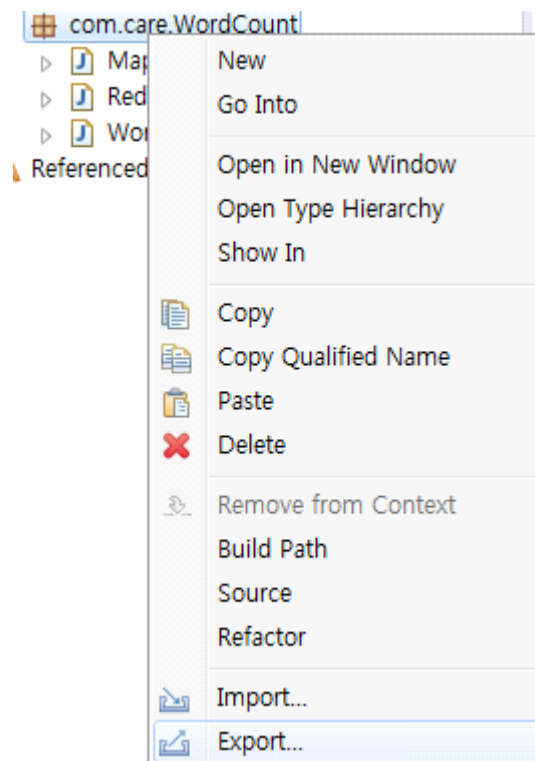
```

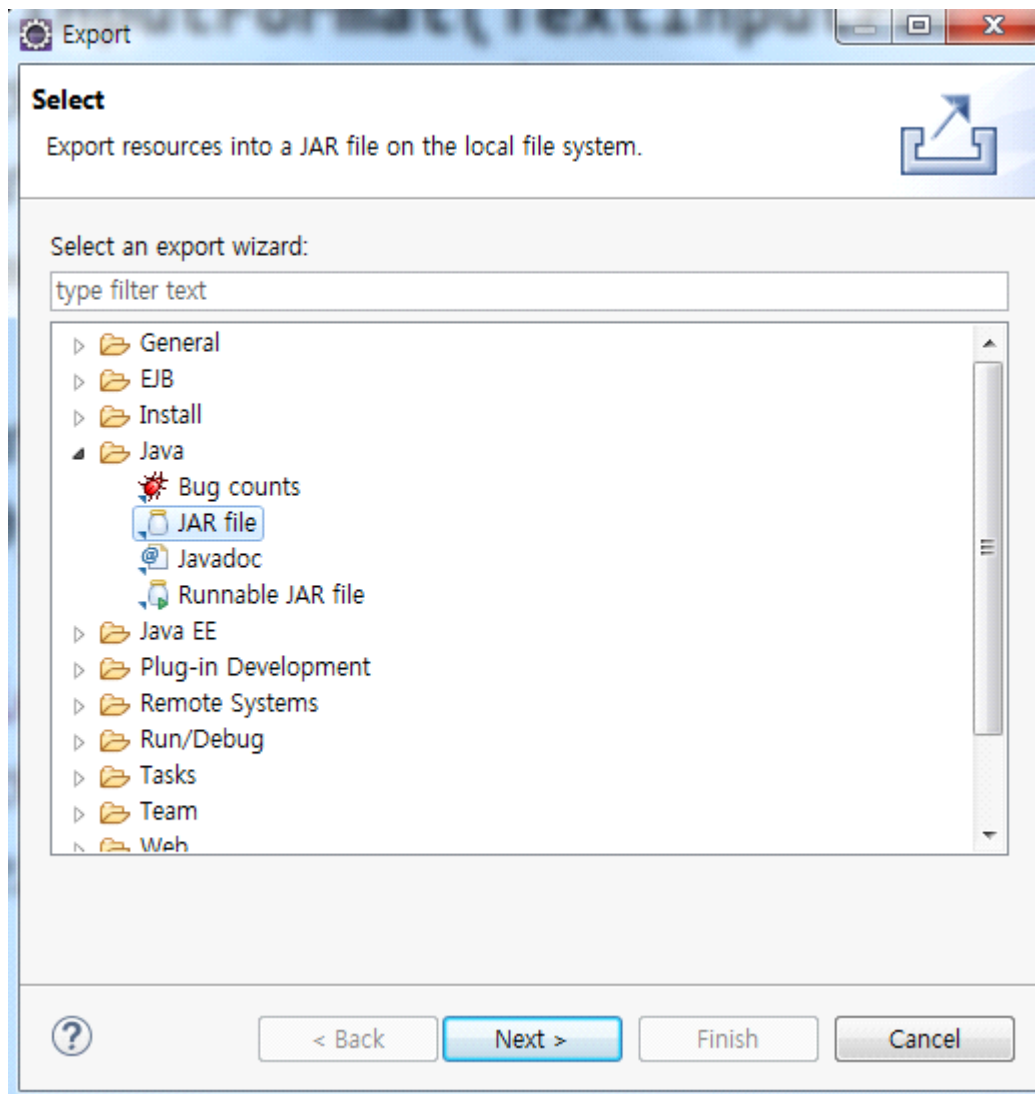

컴파일 및 실행

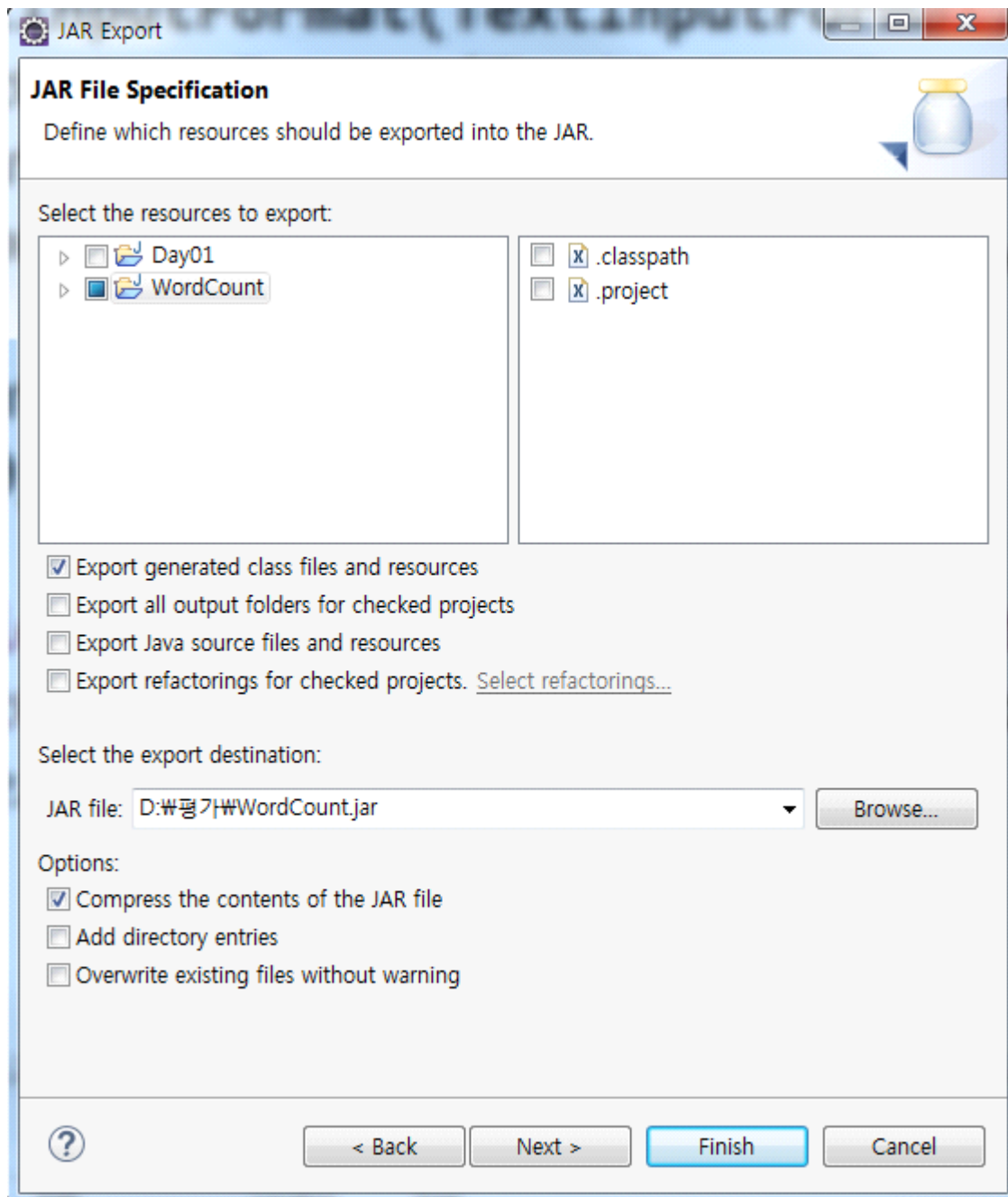
2017년 10월 5일 목요일 오후 3:41

Jar 생성

2017년 10월 5일 목요일 오후 2:50



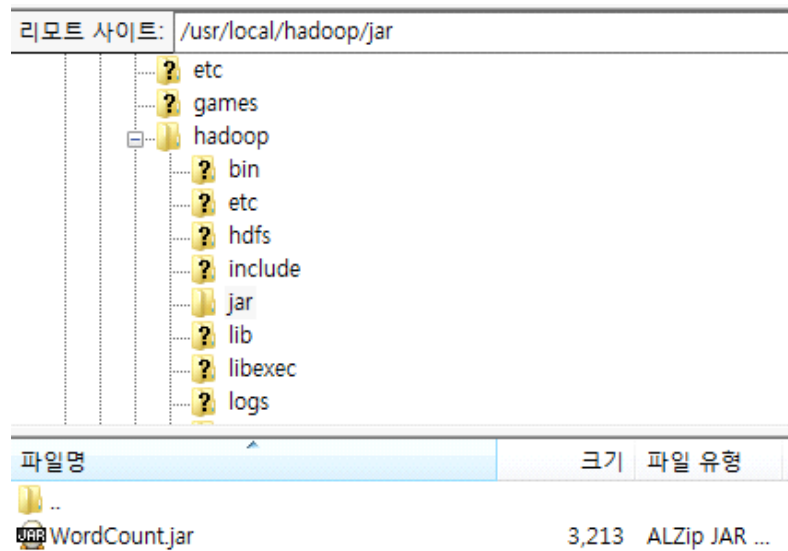




분석 폴더 생성 및 이동

2017년 11월 24일 금요일 오전 10:51

```
manager@master:~$ sudo mkdir -p /usr/local/hadoop/jar
manager@master:~$ sudo chown -R manager:hadoop /usr/local/hadoop/jar/
```



```
sudo mkdir -p /usr/local/hadoop/jar
sudo chown -R manager:hadoop /usr/local/hadoop/jar/
```

컴파일

2017년 10월 5일 목요일 오후 3:29



CHANGES

```
jিনuser@jিন-VirtualBox:/usr/local/hadoop$ hadoop fs -mkdir -p /wordcount/input
```

```
jিনuser@jিন-VirtualBox:/usr/local/hadoop$ hadoop fs -copyFromLocal /usr/local/hadoop/CHANGES.txt /wordcount/input/
17/10/05 16:18:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
copyFromLocal: /usr/local/hadoop/CHANGES.txt (허가 거부)
```

```
jিনuser@jিন-VirtualBox:/usr/local/hadoop$ ls -al
합계 604
drwxr-xr-x 10 jিনuser hadoop 4096 10월 5 16:15 .
drwxr-xr-x 12 root root 4096 10월 3 14:59 ..
-rwxr-x-- 1 root root 446615 10월 5 16:15 CHANGES.txt
-rw-r--r-- 1 jিনuser hadoop 99253 6월 2 15:24 LICENSE.txt
-rw-r--r-- 1 jინuser hadoop 15915 6월 2 15:24 NOTICE.txt
-rw-r--r-- 1 jინuser hadoop 1366 6월 2 15:24 README.txt
-rw-r--r-- 1 root root 3185 10월 5 15:30 WordCount.jar
drwxr-xr-x 2 jინuser hadoop 4096 6월 2 15:24 bin
drwxr-xr-x 2 jინuser hadoop 4096 6월 2 15:24 etc
```

위처럼 실행해 보았더니 허가 거부가 나왔다. 권한을 확인해 봤더니 CHANGES.txt와 WordCount.jar가 root 이다.

```
jინuser@jინ-VirtualBox:/usr/local/hadoop$ sudo chown -R jინuser:hadoop /usr/local/hadoop
jინuser@jინ-VirtualBox:/usr/local/hadoop$ ls -al
합계 604
drwxr-xr-x 10 jინuser hadoop 4096 10월 5 16:15 .
drwxr-xr-x 12 root root 4096 10월 3 14:59 ..
-rwxr-x-- 1 jინuser hadoop 446615 10월 5 16:15 CHANGES.txt
-rw-r--r-- 1 jინuser hadoop 99253 6월 2 15:24 LICENSE.txt
-rw-r--r-- 1 jინuser hadoop 15915 6월 2 15:24 NOTICE.txt
-rw-r--r-- 1 jინuser hadoop 1366 6월 2 15:24 README.txt
-rw-r--r-- 1 jინuser hadoop 3185 10월 5 15:30 WordCount.jar
drwxr-xr-x 2 jინuser hadoop 4096 6월 2 15:24 bin
drwxr-xr-x 2 jინuser hadoop 4096 6월 2 15:24 etc
```

```
jინuser@jინ-VirtualBox:/usr/local/hadoop$ hadoop jar WordCount.jar com.care.WordCount.WordCount /wordcount/input /wordcount/output/
```

```
17/10/05 16:25:15 WARN mapreduce.LocalJobRunner$Job: LocalJobRunner$Job_1103030302_0001
java.lang.Exception: java.io.IOException: Type mismatch in key from map: expected org.apache.hadoop.io.LongWritable, received org.apache.hadoop.io.Text
    at org.apache.hadoop.mapred.LocalJobRunner$Job.runTasks(LocalJobRunner.java:489)
    at org.apache.hadoop.mapred.LocalJobRunner$Job.run(LocalJobRunner.java:549)
Caused by: java.io.IOException: Type mismatch in key from map: expected org.apac
```

위처럼 실행했지만 에러 발생 아마도 타입 문제인듯....

확인결과 오타

```
conf.setJobName("wordcount");  
conf.setOutputKeyClass(Text.class);
```

두 번째 줄의 내용을 빼먹었음.

```
jlnuser@jin-VirtualBox:/usr/local/hadoop$ hadoop jar WordCount.jar com.care.Word  
Count.WordCount /wordcount/input /wordcount/output/  
17/10/05 16:53:23 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
17/10/05 16:53:25 INFO Configuration.deprecation: session.id is deprecated. Inst  
ead, use dfs.metrics.session-id  
17/10/05 16:53:25 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName  
=JobTracker, sessionId=  
17/10/05 16:53:25 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with proces  
sName=JobTracker, sessionId= - already initialized  
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException:  
Output directory hdfs://master:62350/wordcount/output already exists  
    at org.apache.hadoop.mapred.FileOutputFormat.checkOutputSpecs(FileOutput  
Format.java:121)
```

재 실행 결과 기존의 output 폴더가 존재하여 문제 발생

```
jlnuser@jin-VirtualBox:/usr/local/hadoop$ hadoop fs -rmrdir /wordcount/output  
17/10/05 16:56:11 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable
```

위처럼 폴더 삭제 후 다시 분석하면 정상

```
hadoop fs -mkdir -p /wordcount/input
```

```
hadoop fs -copyFromLocal /usr/local/hadoop/*.txt /wordcount/input/
```

```
hadoop jar WordCount.jar com.care.wordcount.WordCnt /wordcount/input /wordcount/output
```

```
hadoop fs -ls /wordcount/output
```

```
hadoop fs -cat /wordcount/output/part-00000
```