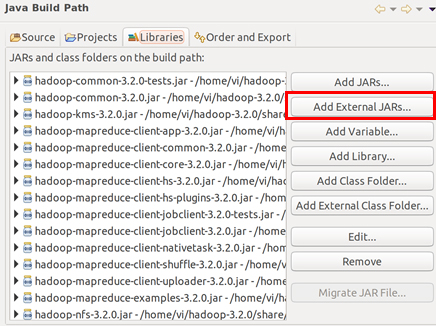
**Jar 파일 생성 부터 하둡 실행까지**

1. Elipse 프로젝트 생성 (프로젝트 명 : WordCount)

package는 생성 x.

2. WordCount, WordCountMapper, WordCountReducer 자바 파일을 복사 붙여 넣기.

3. 프로젝트 명에서 마우스 오른쪽 클릭(Build Path –> Configure build Path…)

a. Add External JARs...를 누른다.

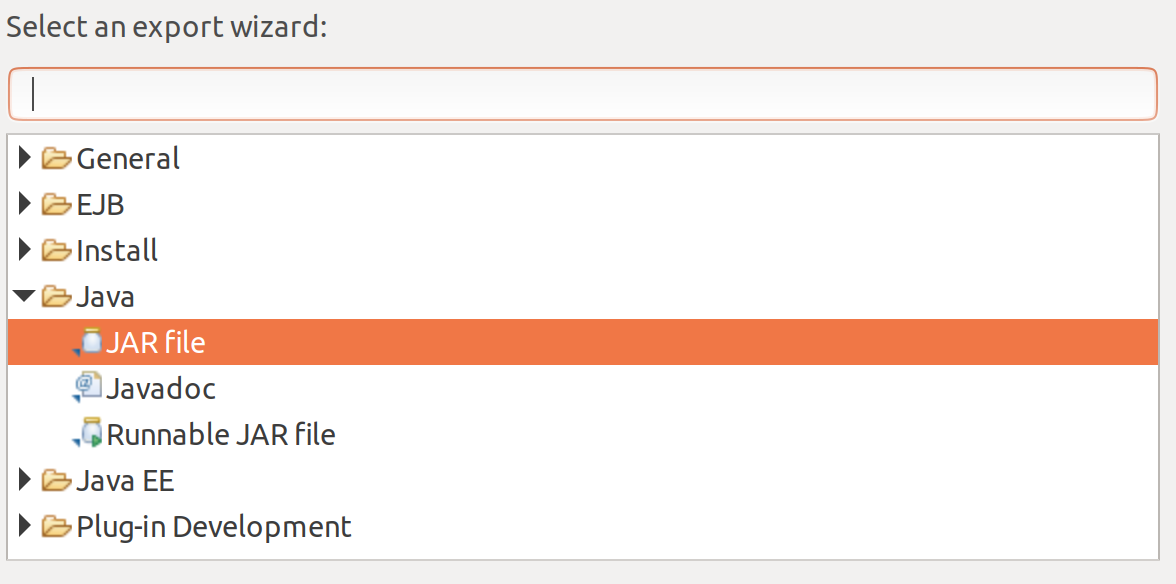
b. hadoop-3.2.0의 Share/hadoop/common 폴더에 있는 jar 파일 추가.

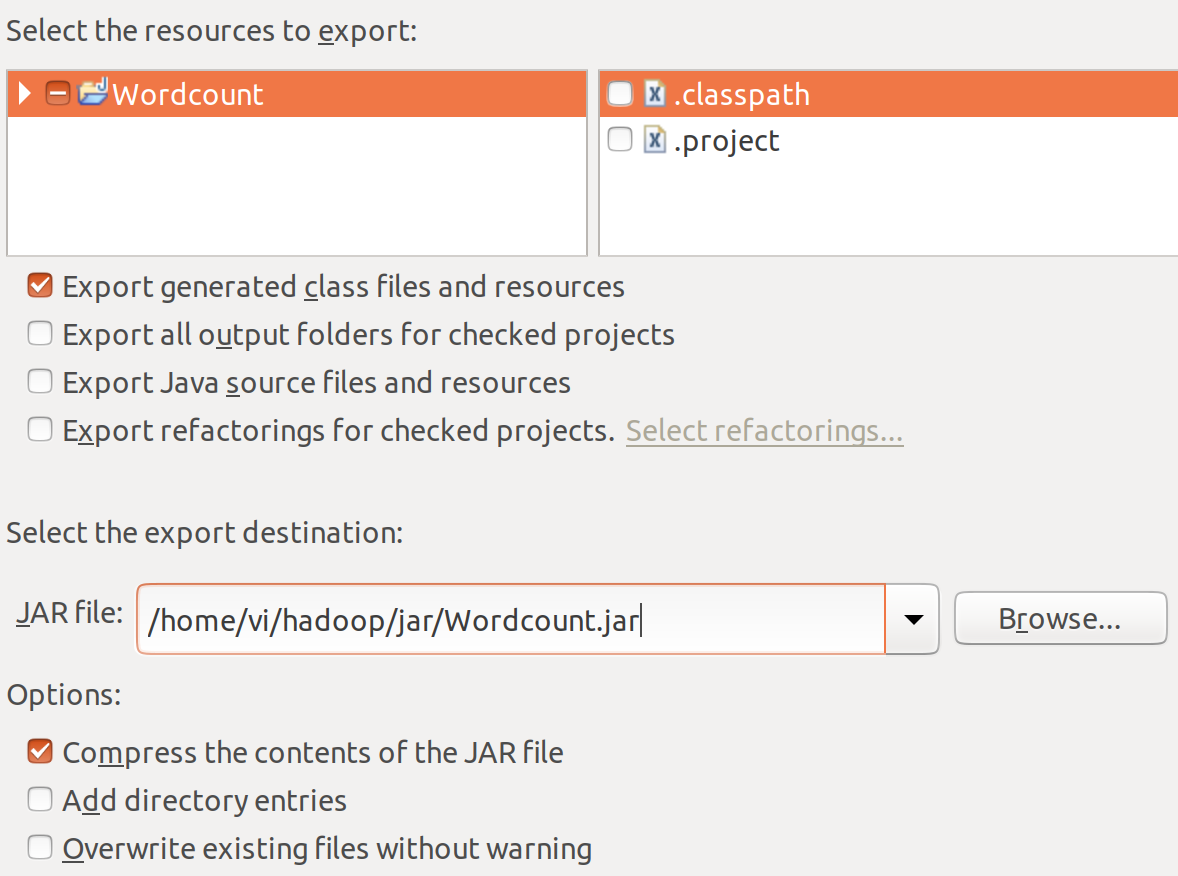
c. hadoop-3.2.0의 Share/hadoop/mapreduce 폴더에 있는 jar 파일 추가.

d. Apply and close 클릭.

4. jar 파일 만들기

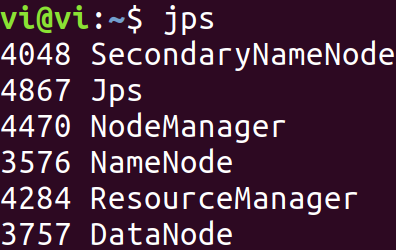
a. 프로젝트 명에서 마우스 오른쪽 클릭(Export… 클릭)

b. 위의 그림과 같이 Java-JAR file 선택 후, NEXT 버튼 클릭

c. .classpath, .project 체크를 풀고, JAR file에서 JAR이 생성될 경로와 파일 명 입력.

d. 확인 후, finish 버튼 클릭.

5. 하둡으로 실행해 보기.

하둡은 당연히 켜져 있는 상태.

|  |
| --- |
| **현재 HDFS에 저장되어 있는 내용. (MariaDB에서 테이블 생성 후, 여러 열을 가져 왔을 때)**  **HDFS 파일 경로 명 : /user/yunhyuck/test2/part-m-00000**  **-----------------------------------------**  100,P1,Computer,C413,A  100,P1,Computer,E412,A  200,P2,Electric,C123,B  300,P3,Computer,C312,A  300,P3,Computer,C324,C  300,P3,Computer,C413,A  400,P1,Computer,C312,A  400,P1,Computer,C324,A  400,P1,Computer,C413,B  400,P1,Computer,E412,C |

테스트 데이터가 있다고 가정하고, 실행해 본다. 현재 Map에서 공백을 기준으로 단어를 카운트 한다. 그렇기 때문에 맵 설정에서 replace 명령을 통해 , 를 공백으로 바꿔 주고 실행을 하였다.

a. yarn 실행

$ yarn jar /home/vi/hadoop/jar/Wordcount.jar WordCount /user/yunhyuck/test2/part-m-00000 output

/home/vi/hadoop/jar/Wordcount.jar → JAR 파일이 저장된 경로.

WordCount → 드라이버 클래스(WordCount.java에서 잡 객체를 생성할 때, WordCount로 지정. )

/user/yunhyuck/test2/part-m-00000 → 입력 파일 경로

output2 → 출력 파일 경로. 상세 경로를 적어 주지 않으면, /user/vi/output 경로에 출력 파일 생성.

|  |
| --- |
| 2019-04-07 19:15:12,642 INFO mapreduce.Job: map 0% reduce 0%  2019-04-07 19:15:16,722 INFO mapreduce.Job: map 100% reduce 0%  2019-04-07 19:15:22,769 INFO mapreduce.Job: map 100% reduce 100%  2019-04-07 19:15:22,785 INFO mapreduce.Job: Job job\_1554631447670\_0001 completed successfully  2019-04-07 19:15:22,873 INFO mapreduce.Job: Counters: 54  File System Counters  FILE: Number of bytes read=536  FILE: Number of bytes written=444121  FILE: Number of read operations=0  FILE: Number of large read operations=0  FILE: Number of write operations=0  HDFS: Number of bytes read=349  HDFS: Number of bytes written=108  HDFS: Number of read operations=8  HDFS: Number of large read operations=0  HDFS: Number of write operations=2  HDFS: Number of bytes read erasure-coded=0  Job Counters  Launched map tasks=1  Launched reduce tasks=1  Data-local map tasks=1  Total time spent by all maps in occupied slots (ms)=1996  Total time spent by all reduces in occupied slots (ms)=2256  Total time spent by all map tasks (ms)=1996  Total time spent by all reduce tasks (ms)=2256  Total vcore-milliseconds taken by all map tasks=1996  Total vcore-milliseconds taken by all reduce tasks=2256  Total megabyte-milliseconds taken by all map tasks=2043904  Total megabyte-milliseconds taken by all reduce tasks=2310144  Map-Reduce Framework  Map input records=10  Map output records=50  Map output bytes=430  Map output materialized bytes=536  Input split bytes=119  Combine input records=0  Combine output records=0  Reduce input groups=17  Reduce shuffle bytes=536  Reduce input records=50  Reduce output records=17  Spilled Records=100  Shuffled Maps =1  Failed Shuffles=0  Merged Map outputs=1  GC time elapsed (ms)=87  CPU time spent (ms)=1060  Physical memory (bytes) snapshot=501370880  Virtual memory (bytes) snapshot=5316816896  Total committed heap usage (bytes)=457703424  Peak Map Physical memory (bytes)=281010176  Peak Map Virtual memory (bytes)=2656210944  Peak Reduce Physical memory (bytes)=220360704  Peak Reduce Virtual memory (bytes)=2660605952  Shuffle Errors  BAD\_ID=0  CONNECTION=0  IO\_ERROR=0  WRONG\_LENGTH=0  WRONG\_MAP=0  WRONG\_REDUCE=0  File Input Format Counters  Bytes Read=230  File Output Format Counters  Bytes Written=108 |

[http://localhost:9870](http://localhost:9870/)

접속 해서 출력 파일 확인.

