



# Mapreduce Jar Creation

...

## Running MR Job on HDFS File

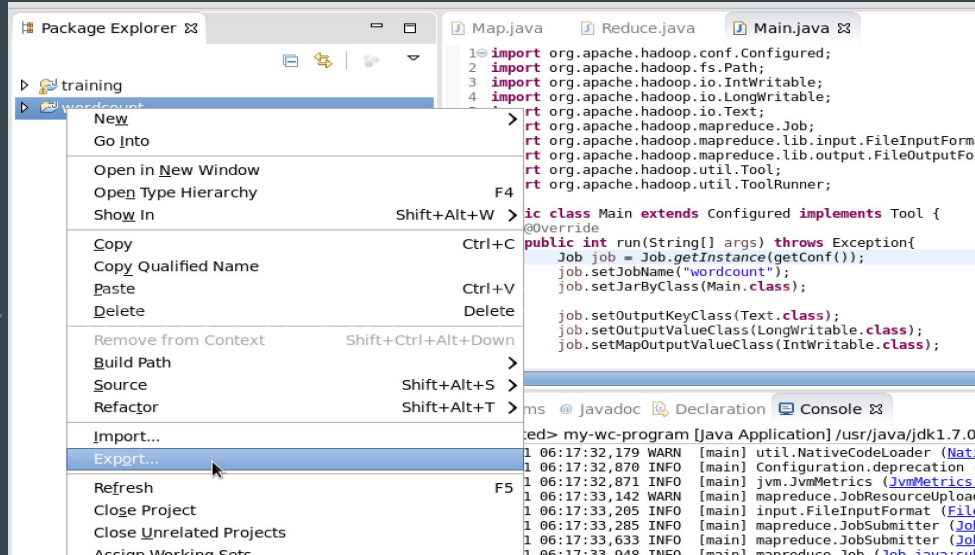


# Moving the input file from local to hdfs

```
cloudera@quickstart Desktop]$  
cloudera@quickstart Desktop]$ hadoop fs -mkdir /user/cloudera/inputfolder  
cloudera@quickstart Desktop]$ hadoop fs -put inputfolder/inputfile.txt inputfolder  
cloudera@quickstart Desktop]$ hadoop fs -ls /user/cloudera/inputfolder  
Found 1 items  
-rw-r--r-- 1 cloudera cloudera 94 2020-04-11 12:22 /user/cloudera/inputfo  
lder/inputfile.txt  
cloudera@quickstart Desktop]$ █
```

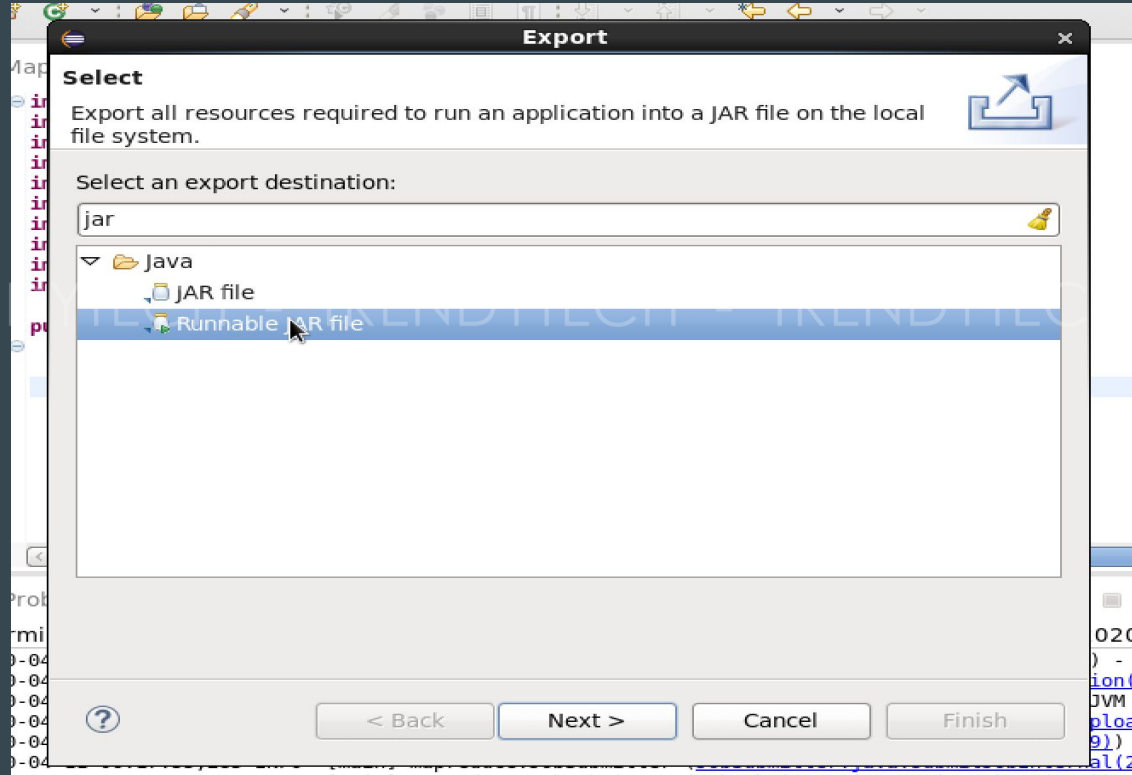
```
[cloudera@quickstart Desktop]$  
[cloudera@quickstart Desktop]$ hadoop fs -cat /user/cloudera/inputfolder/*  
we are all geared up to learn big data  
big data tech stack is fun  
yes it is an absolute fun  
  
[cloudera@quickstart Desktop]$ █
```

# Exporting our project as jar file



**Right click  
on project  
-> export ->  
runnable jar**

# Exporting our project as jar file



# Select the run configuration for your project



**Runnable JAR File Export**

**Runnable JAR File Specification**

⚠ Program arguments will not be part of the runnable JAR. Arguments can be passed on the command line when launching the JAR

Launch configuration:  
my-wc-program - wordcount

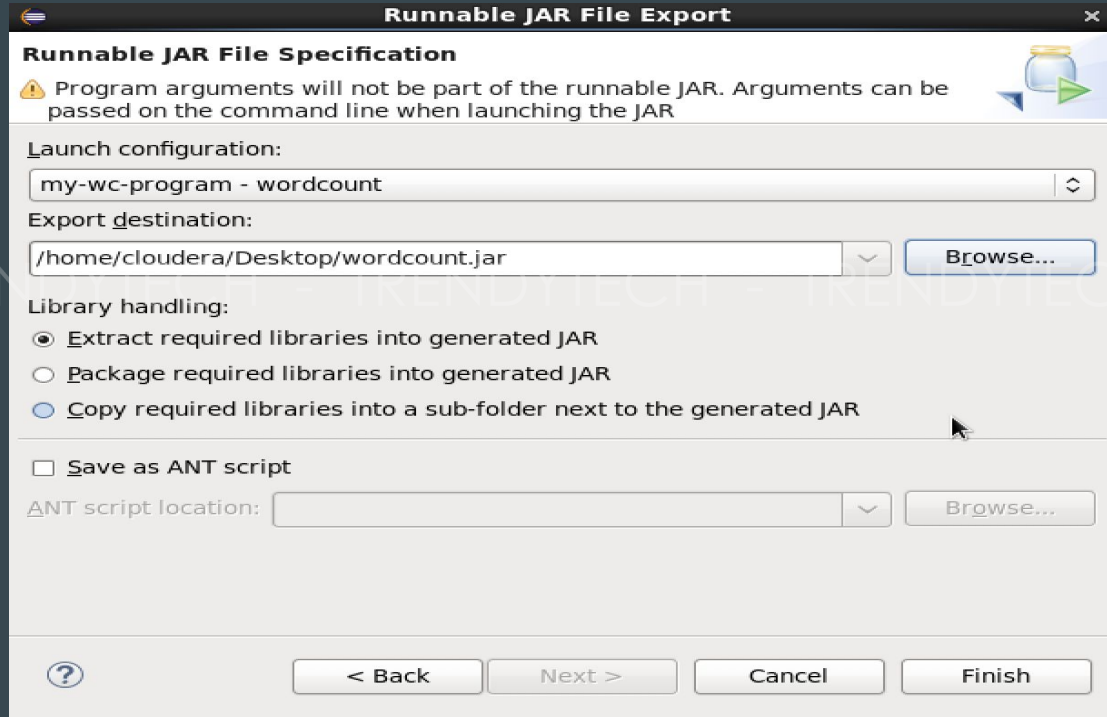
Export destination:  
 Browse...

Library handling:

- ☒ Extract required libraries into generated JAR
- ☐ Package required libraries into generated JAR
- ☐ Copy required libraries into a sub-folder next to the generated JAR

☐ Save as ANT script

# Give the name of the jar and also the location where jar should be created



**Runnable JAR File Export**

**Runnable JAR File Specification**

⚠ Program arguments will not be part of the runnable JAR. Arguments can be passed on the command line when launching the JAR

Launch configuration:  
my-wc-program - wordcount

Export destination:  
/home/cloudera/Desktop/wordcount.jar Browse...

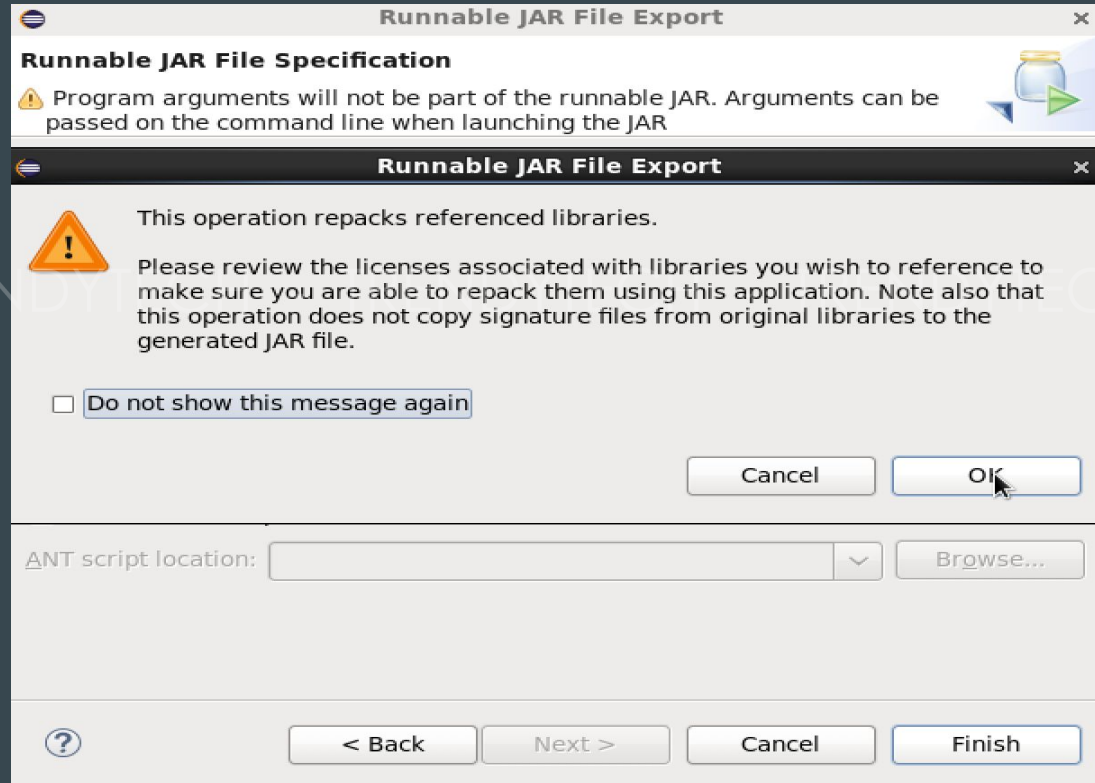
Library handling:  
☒ Extract required libraries into generated JAR  
☐ Package required libraries into generated JAR  
☐ Copy required libraries into a sub-folder next to the generated JAR

☐ Save as ANT script

ANT script location: Browse...

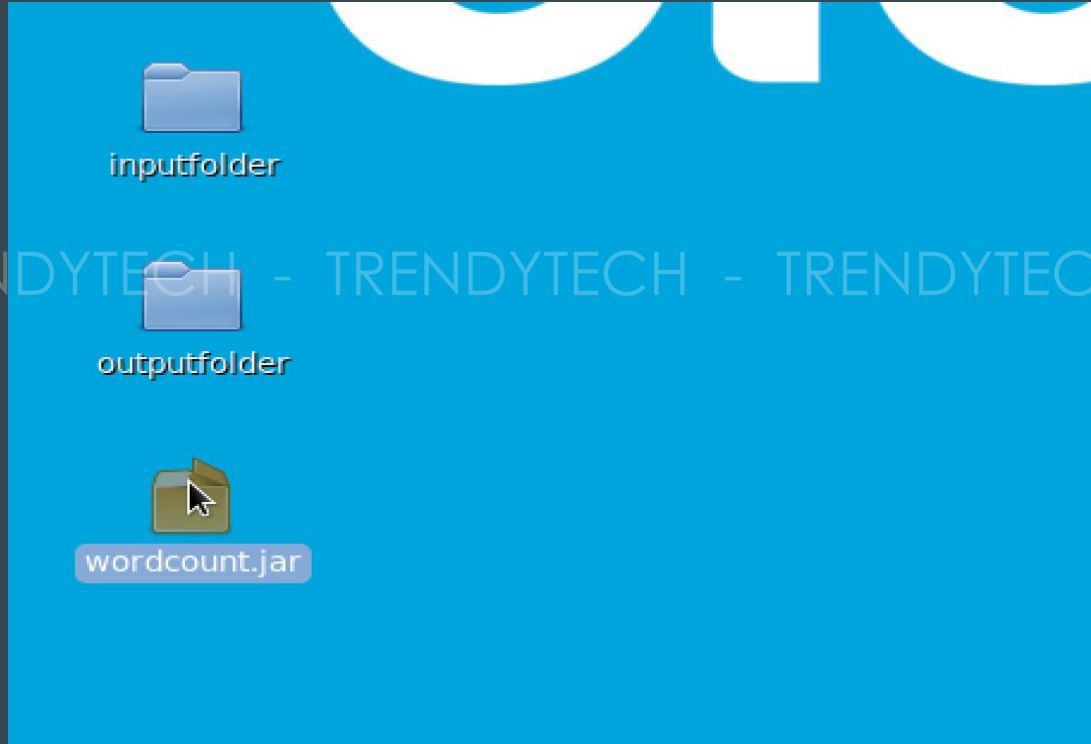
? < Back Next > Cancel Finish

# Ignore the warning while importing





# You can see the jar is now present on Desktop







# Run the jar now using below command

```
hadoop jar <jarname> <input path> <output path>
```

```
cloudera@quickstart ~]$  
cloudera@quickstart ~]$  
cloudera@quickstart ~]$ hadoop jar /home/cloudera/Desktop/wordcount.jar /user/cloudera/inputfolder/inputfile.txt /user/cloudera/outputfolder
```



## Job Tracking URL will be displayed when Job Runs (in the logs)

```
20/04/11 12:38:11 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1586388936914_0011
20/04/11 12:38:12 INFO impl.YarnClientImpl: Submitted application application_1586388936914_0011
20/04/11 12:38:12 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1586388936914_0011/
20/04/11 12:38:12 INFO mapreduce.Job: Running job: job_1586388936914_0011
20/04/11 12:38:24 INFO mapreduce.Job: Job job_1586388936914_0011 running in uber mode : false
20/04/11 12:38:24 INFO mapreduce.Job:  map 0% reduce 0%
```

**We can try opening the url in our cloudera browser.**

# The job tracking URL will show details related to the job

MapReduce Job job\_158... x

quickstart.cloudera:19888/jobhistory/job/job\_1586388936914\_0011

hadoop

MapReduce Job job\_1586388936914\_0011

Logged in as: dr.who

Application

Job

- Overview
- Counters
- Configuration
- Map tasks
- Reduce tasks

Tools

Job Overview

**Job Name:** wordcount  
**User Name:** cloudera  
**Queue:** root.cloudera  
**State:** SUCCEEDED  
**Uberized:** false  
**Submitted:** Sat Apr 11 12:38:11 PDT 2020  
**Started:** Sat Apr 11 12:38:22 PDT 2020  
**Finished:** Sat Apr 11 12:38:37 PDT 2020  
**Elapsed:** 14sec  
**Diagnostics:**  
**Average Map Time:** 5sec  
**Average Shuffle Time:** 3sec  
**Average Merge Time:** 0sec  
**Average Reduce Time:** 0sec

**ApplicationMaster**


Attempt Number	Start Time	Node	Logs
1	Sat Apr 11 12:38:17 PDT 2020	quickstart.cloudera:8042	<a href="#">logs</a>

Task Type	Total	Complete
<b>Map</b>	1	1
<b>Reduce</b>	1	1

Attempt Type	Failed	Killed	Successful
<b>Maps</b>	0	0	1
<b>Reduces</b>	0	0	1

# See list of all jobs on port 8088 (localhost:8088)

localhost:8088/cluster

 **All Applications**

Cluster

- About
- Nodes
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

**Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved
11	0	0	11	0	0 B	8 GB	0 B

**Cluster Nodes Metrics**

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

**User Metrics for dr.who**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores
<a href="#">application_1586388936914_0011</a>	cloudera	wordcount	MAPREDUCE	root.cloudera	Sat Apr 11 12:38:11 -0700 2020	Sat Apr 11 12:38:38 -0700 2020	FINISHED	SUCCEEDED	N/A	N/A
<a href="#">application_1586388936914_0010</a>	cloudera	orders.jar	MAPREDUCE	root.cloudera	Thu Apr 9 16:19:20 -0700 2020	Thu Apr 9 16:19:47 -0700 2020	FINISHED	SUCCEEDED	N/A	N/A
<a href="#">application_1586388936914_0009</a>	cloudera	orders.jar	MAPREDUCE	root.cloudera	Thu Apr 9 16:17:52 -0700 2020	Thu Apr 9 16:18:19 -0700 2020	FINISHED	SUCCEEDED	N/A	N/A
<a href="#">application_1586388936914_0008</a>	cloudera	orders.jar	MAPREDUCE	root.cloudera	Thu Apr 9 16:17:52 -0700 2020	Thu Apr 9 16:18:19 -0700 2020	FINISHED	SUCCEEDED	N/A	N/A



# See HDFS details through web UI on Port 50070 (localhost:50070)

localhost:50070/dfshealth.html#tab-overview

Search

☆ 📅 📧 ⬇ 🏠 💬

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities ▾

## Overview 'quickstart.cloudera:8020' (active)

<b>Started:</b>	Wed Apr 08 16:35:00 -0700 2020
<b>Version:</b>	2.6.0-cdh5.13.0, r42e8860b182e55321bd5f5605264da4adc8882be
<b>Compiled:</b>	Wed Oct 04 11:08:00 -0700 2017 by jenkins from Unknown
<b>Cluster ID:</b>	CID-a24185f9-a545-40fe-9553-84c3fdca489f
<b>Block Pool ID:</b>	BP-1067413441-127.0.0.1-1508775264580

## Summary

Security is off.

Safemode is off.

1,206 files and directories, 1,043 blocks = 2,249 total filesystem object(s).

Heap Memory used 181.64 MB of 313 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 39.21 MB of 39.63 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

<b>Configured Capacity:</b>	54.51 GB
<b>DFS Used:</b>	864.62 MB (1.55%)

# Browse HDFS filesystem

localhost:50070/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Browse the file system  
Logs

## Overview 'quickstart.cloudera:8020' (active)

<b>Started:</b>	Wed Apr 08 16:35:00 -0700 2020
<b>Version:</b>	2.6.0-cdh5.13.0, r42e8860b182e55321bd5f5605264da4adc8882be
<b>Compiled:</b>	Wed Oct 04 11:08:00 -0700 2017 by jenkins from Unknown
<b>Cluster ID:</b>	CID-a24185f9-a545-40fe-9553-84c3fdca489f
<b>Block Pool ID:</b>	BP-1067413441-127.0.0.1-1508775264580

Hadoop Overview Datanodes Snapshot Startup Progress Utilities ▾

## Browse Directory

/user/cloudera/outputfolder Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	0 B	Sat Apr 11 12:38:37 -0700 2020	1	128 MB	<a href="#">_SUCCESS</a>
-rw-r--r--	cloudera	cloudera	114 B	Sat Apr 11 12:38:37 -0700 2020	1	128 MB	<a href="#">part-r-00000</a>

Hadoop, 2017.



We have successfully Created & Executed the Jar

Happy Learning!!!



**5** Star Google Rated  
Big Data Course

**LEARN FROM THE EXPERT**



**9108179578**

**Call for more details**





# Follow US

**Trainer** Mr. Sumit Mittal

**Phone** 9108179578

**Email** trendytech.sumit@gmail.com

**Website** <https://trendytech.in/courses/big-data-online-training/>

**LinkedIn** <https://www.linkedin.com/in/bigdatabysumit/>

**Twitter** @BigdataBySumit

**Instagram** bigdatabysumit

**Facebook** <https://www.facebook.com/trendytech.in/>

**Youtube** TrendyTech