Group B:- Big Data Analysis

Write a code in JAVA for a simple WordCount application that counts the number of occurrences of each word in a given input set using the Hadoop Map-Reduce framework on a local-standalone set-up.

Hadoop Installation Video Ref :-
https://www.youtube.com/watch?v=O1GcLdXuYc8

Documentation :-
https://github.com/bhushanasati25/SPPU-TE-Practical/tree/main/6th%20Sem/Data%20Science%20%26%20Big%20Data/Hadoop%20Program

WordCount Program :-
https://www.youtube.com/watch?v=QfdxiXFTcak&list=PLI6DC0kKx3wt-M-Ij25U1YCkeip5RvQuD&index=2

✕   –   ⊡   Inbox -   ✕   The-Sp   ✕   TE_Cor   ✕   ⊘ Brows   ✕   All App   ✕   TE_Cor   ✕   PR 2/   ✕   Practic   ✕   Untitle   ✕   ⊘ New Ta   ✕   +

← → C ⌂   ⓘ localhost:50070/explorer.html#/user/hduser/bhushanoutput

★ Bookmarks   Facebook   Flipkart   Gmail   Google   IRCTC   SPPU   Onlinesbi   Software Daw...   You Tube   iNeuron.ai   Create Androi...   »

Hadoop   Overview   Datanodes

Browse Directory

File information - part-r-00000   ✕

Download   Head the file (first 32K)   Tail the file (last 32K)

Block information --   [Block 0 ▾]

Block ID: 1073741832

Block Pool ID: BP-996143458-127.0.1.1-1647935279153

Generation Stamp: 1008

Size: 121

Availability:

- iHack-PC

/user/hduser/bhushanoutput   Go!   📁 ☁ ▤

Show [25] entries   Search: [   ]

☐	Permission	Owner			ock Size	Name
☐	-rw-r--r--	hduser			8 MB	_SUCCESS	🗑
☐	-rw-r--r--	hduser			8 MB	part-r-00000	🗑

File contents

```
Association's  1
College        1
District       1
Education      1
I              1
India          2
Pune           2
USA 2
```

Showing 1 to 2 of 2 entries   Previous   1   Next

Hadoop, 2017.

✕   –   ⊡   Inbox -   ✕   The-Sp   ✕   TE_Cor   ✕   ⊘ Brows   ✕   All App   ✕   TE_Cor   ✕   PR 2/   ✕   Practic   ✕   Untitle   ✕   ⊘ New Ta   ✕   +

← → C ⌂   ⓘ localhost:50070/explorer.html#/user/hduser/bhushanoutput

★ Bookmarks   Facebook   Flipkart   Gmail   Google   IRCTC   SPPU   Onlinesbi   Software Daw...   You Tube   iNeuron.ai   Create Androi...   »

Hadoop   Overview   Datanodes   Datanode Volume Failures   Snapshot   Startup Progress   Utilities ▾
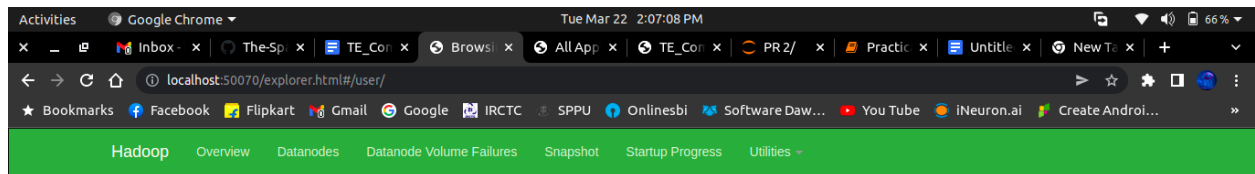
Browse Directory

/user/hduser/bhushanoutput   Go!   📁 ☁ ▤

Show [25] entries   Search: [   ]

☐	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	-rw-r--r--	hduser	supergroup	0 B	Mar 22 13:29	1	128 MB	_SUCCESS	🗑
☐	-rw-r--r--	hduser	supergroup	121 B	Mar 22 13:29	1	128 MB	part-r-00000	🗑

Showing 1 to 2 of 2 entries   Previous   1   Next

Hadoop, 2017.

Browse Directory

/user/   Go!

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
	drwxr-xr-x	hduser	supergroup	0 B	Mar 22 13:25	0	0 B	bhushan	🗑
	drwxr-xr-x	hduser	supergroup	0 B	Mar 22 13:28	0	0 B	hduser	🗑
	drwxr-xr-x	hduser	supergroup	0 B	Mar 22 13:26	0	0 B	input	🗑

Showing 1 to 3 of 3 entries   Previous 1 Next

Hadoop, 2017.

Hadoop   **All Applications**

Logged in as: dr.who

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
1	0	0	1	0	0 B	8 GB	0 B	0	8	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries   Search:

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
application_1647935230954_0001	hduser	word count	MAPREDUCE	default	0	Tue Mar 22 13:28:42 +0550 2022	Tue Mar 22 13:29:03 +0550 2022	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	0.0	0.0		History	0

Showing 1 to 1 of 1 entries   First Previous Next Last