

## **Upload data to HDFS**

1. Introduction

- 2. Prerequistes
- 3. Set Environment Variables
- 4. <u>Setup SSH daemon</u>
- 5. <u>Download hadoop and place it</u> in the home directory.
- 6. <u>Unpack hadoop</u>
- 7. Configure Hadoop
- 8. Format the namenode
- 9. Setup hadoop plugin
- 10. Start the cluster
- 11. Setup hadoop location
- 12. Upload data
- 13. Create and run a test project.

🖸 ADD THIS 📲 🎡 🧤 ...)

We are now ready to run the first Map/Reduce project but data is still missing. This section explains how to upload data to the Hadoop Distributed File System (HDFS).

## **Upload Files To HDFS**

1. Open a new CYGWIN command window.



2. Execute the following commands in the new CYGWIN window as shown on the image above.

cd hadoop-0.19.1 bin/hadoop fs -mkdir In bin/hadoop fs -put \*.txt In

When the last of the above commands starts executing, you should see some activity in other Hadoop windows as shown in the image below.



User/In/NOTIC 09/03/10 17:3 Map updated: 5 cd hadoop-0.19.1 9/03/10 17:3 rs ip=/1 perme 9/03/10 17:3 User@BAHCLIENT ~/hadoop-0.19.1 09/03/10 17:3 User@BAHCLIENT ~/hadoop-0.19.1 09/03/10 17:3 Liser@BAHCLIENT ~/hadoop-0.19.1 Ser@BAHCLIENT ~/hadoop-0.19.1	
<pre> F -/hadoop-0  99/03/10 17:3 99/03/10 17:32:59 INFO common.Storage: Number of files under construction 99/03/10 17:32:59 INFO common.Storage: Edits file \tmp\hadoop-User\dfs dary\current\edits of size 1382 edits # 19 loaded in 0 seconds. 99/03/10 17:32:59 INFO common.Storage: Image file of size 1051 saved s. 99/03/10 17:32:59 INFO namenode.FSNamesystem: Number of transactions me for transactions(ms): 0 Number of syncs: 0 SyncTimes(ms): 0 99/03/10 17:33:00 INFO namenode.SecondaryNameNode: Posted URL 0.0.0 age=1&amp;port=50090&amp;machine=192.168.16.39&amp;token=-18:1626391377:0:1236 720436734 99/03/10 17:33:00 WARN namenode.SecondaryNameNode: Checkpoint do ze: 1051  F Start  A 6 Windows Comman </pre>	

The result of these commands is a newly created directory -- named  ${\bf In}$  -- in the HDFS which contains a set of text files that comes with the Hadoop distribution.

3. Close the Cygwin Window.

## Verify if the files were uploaded correctly

In this section we will check if the files were uploaded correctly.

- 1. Open the Eclipse environment.
- 2. Open **DFS locations** folder which is located in the Project Explorer tab of **Map/Reduce** perspective.
- 3. Open localhost folder in DFS locations folder.
- 4. Keep opening HDFS folders until you navigate to the newly created **In** directory, as shown in the image below.

Map/Reduce - LICENSE.txt (11	.1 Kb, r1) - Eclipse Platform
File Edit Source Refactor Navigate	Search Project Run Window Help
📫 • 🖫 👜 🗄 🏇 • 🔘 • 💁	•   🖀 📽 🎯 •   🅭 🖨 🖋 •   🖢 - 🖓 • 🗠
Project Explorer 🛿 📃 🗖	hdfs://localhost:9100/user/User/In/LICENSE.txt 🛛
E 🕏 🍹	
DFS Locations	Apache Lic
🖃 👧 localhost	Version 2.0, Jan
i - (2)	nttp://www.apache.o
■ 🔁 tmp (1)	TERMS AND CONDITIONS FOR USE, REPRODUCT
□ ··· 👉 user (1)	
□ □ □ □ □ □ □ □ □ □	1. Definitions.
📄 CHAN(	
📄 LICEN:	"License" shall mean the terms and
NOTIC	and discribucion as defined by Se
E READ	"Licensor" shall mean the copy
	the copyright owner that is gr
	<
	🔝 Problems 🖉 Tasks @ Javadoc 🦃 Map/Reduct
	Location
	🗉 🧖 localhost
	<
A start	mman
Start Start	

Verifying that the data was uploaded correctly

- 5. When you get to the **In** directory, double-click on the file **LICENCE.txt** to open it.
- 6. If you see something similar to the image above then the data was uploaded correctly and you can proceed to your first Hadoop project.

You can now move on to the next step.

Continue

🔼 ADD THIS 📲 🎡 🧤 ...

If you have questions comments suggestions regarding this tutorial you  $\underline{can post}$  them here or you can write me an email to tutorials **AT** v-lad.org.

2008 - Vlad Korolev