

		docker-hadoop-spark-workbench	Running (6/6)			
		spark-master e0bf5eb35a28	bde2020/spark-master:2.1.0-f	Running	7077:7077 8080:8080	2 minutes ago
		namenode b48fbf1949ea	bde2020/hadoop-namenode:1	Running	50070:50070	2 minutes ago
		spark-worker-1 e6370092f438	bde2020/spark-worker:2.1.0-f	Running	8081:8081	1 minute ago
		datanode-1 1e5e7524f7a8	bde2020/hadoop-datanode:1	Running	50075:50075	1 minute ago
		spark-notebook 34c70afd8346	bde2020/spark-notebook:2.1	Running	9001:9001	2 minutes ago
		hue-1 2bf852832c5d	bde2020/hdfs-filebrowser:3.1	Running	8088:8088	2 minutes ago

Showing 15 items

```
PS C:\Users\marai\Documents\Data_Engineer\Docker_Hadoop_Spark_Image\docker-hadoop-spark-workbench-master> docker-compose up -d
[+] Running 6/6
 - Container spark-notebook          Started           1.3s
 - Container spark-master            Running           0.0s
 - Container docker-hadoop-spark-workbench-master-hue-1 Running           0.0s
 - Container docker-hadoop-spark-workbench-master-spark-worker-1 Running           0.0s
 - Container namenode                Running           0.0s
 - Container docker-hadoop-spark-workbench-master-datanode-1 Running           0.0s
```

CONTAINER ID	IMAGE	NAMES	COMMAND	CREATED	STATUS	PORTS
521fd7f8211e	bde2020/spark-worker:2.1.0-hadoop2.8-hive-java8	docker-hadoop-spark-workbench-master-spark-worker-1	"entrypoint.sh /bin/..."	5 hours ago	Up 3 minutes (healthy)	0.0.0.0:8081->8081/tcp
2a41b995c132	bde2020/spark-master:2.1.0-hadoop2.8-hive-java8	spark-master	"entrypoint.sh /bin/..."	5 hours ago	Up 3 minutes (healthy)	0.0.0.0:7077->7077/tcp, 0.0.0.0:8080->8080/tcp
1e5e7524f7a8	bde2020/hadoop-datanode:1.1.0-hadoop2.8-java8	docker-hadoop-spark-workbench-master-datanode-1	"/entrypoint.sh /run..."	6 hours ago	Up 3 minutes (healthy)	0.0.0.0:50075->50075/tcp
b48fbf1949ea	bde2020/hadoop-namenode:1.1.0-hadoop2.8-java8	namenode	"/entrypoint.sh /run..."	6 hours ago	Up 3 minutes (healthy)	0.0.0.0:50070->50070/tcp
2bf852832c5d	bde2020/hdfs-filebrowser:3.11	docker-hadoop-spark-workbench-master-hue-1	"entrypoint.sh buil..."	6 hours ago	Up 3 minutes	0.0.0.0:8088->8088/tcp

```
PS C:\Users\marai\Documents\Data_Engineer\Docker_Hadoop_Spark_Image\docker-hadoop-spark-workbench-master> docker cp C:/Users/marai/Documents/Data_Engineer/annkarbook.txt 1e5e7524f7a8:/
PS C:\Users\marai\Documents\Data_Engineer\Docker_Hadoop_Spark_Image\docker-hadoop-spark-workbench-master> docker exec -it 1e5e7524f7a8 bash
Error: No such container: t
PS C:\Users\marai\Documents\Data_Engineer\Docker_Hadoop_Spark_Image\docker-hadoop-spark-workbench-master> docker exec -it 1e5e7524f7a8 bash
root@1e5e7524f7a8:/# ls
annkarbook.txt boot entrypoint.sh hadoop home lib64 mnt proc run sbin sys usr
bin dev etc hadoop-data lib media opt root run.sh srv tmp var
root@1e5e7524f7a8:/#
```

```
root@1e5e7524f7a8:/# hdfs dfs -copyFromLocal annkarbook.txt /user/cloudera
root@1e5e7524f7a8:/# hdfs dfs -ls /user/cloudera
Found 1 items
-rw-r--r-- 3 root cloudera 1853821 2022-11-19 15:58 /user/cloudera/annkarbook.txt
root@1e5e7524f7a8:/#
```

HUE

File Browser

Search for file name				⚙ Actions	✕ Move to trash
Home / user / cloudera					
<input type="checkbox"/>	Name	Size	User	Group	
	.		root	supergroup	
<input type="checkbox"/>	.		cloudera	cloudera	
<input type="checkbox"/>	annkarbook.txt	1.8 MB	root	cloudera	

```

  /---/      /---/      /---/
  \V_  V_  \_  /---/  \_  /
 /---/  .---\  \_  /  /---\  \_  /
  /---/

```

version 2.1.2-SNAPSHOT

```
Using Scala version 2.11.8 (OpenJDK 64-Bit Server VM, Java 1.8.0_121)
Type in expressions to have them evaluated.
Type :help for more information.

scala> import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.SparkSession

scala> val spark = SparkSession.builder().master("local[1]").appName("1TSparkFirst").getOrCreate()
22/11/19 17:14:58 WARN sql.SparkSession$Builder: Using an existing SparkSession; some configuration may not take ef
spark: org.apache.spark.sql.SparkSession = org.apache.spark.sql.SparkSession@28d1b2f

scala> val vim = spark.read.textFile("/user/cloudera/annkarbook.txt")
vim: org.apache.spark.sql.Dataset[String] = [value: string]

scala> vim.head(10)
res0: Array[String] = Array("♦♦♦ ♦♦♦♦♦♦♦♦ ♦♦♦♦♦♦ ", ♦♦♦ ♦♦♦♦♦♦♦, "      ", "", "      ", ♦♦ ♦♦♦♦♦♦, ♦♦♦ ♦♦♦♦♦
♦♦♦♦", "      ")

scala> val wordCount = vim.flatMap(line => line.split(" ")).groupByKey(identity).count()
wordCount: org.apache.spark.sql.Dataset[(String, Long)] = [value: string, count(1): bigint]

scala> wordCount.show(10)
22/11/19 17:23:01 WARN executor.Executor: Managed memory leak detected; size = 4456448 bytes, TID = 2
+-----+-----+
|          value|count(1)|
+-----+-----+
|  ♦♦♦♦-♦♦♦♦♦♦♦♦,|      2|
|             XVII|      9|
|             tant|      3|
|  ♦♦♦?..|     12|
|      pluck,|      1|
|         some|      1|
|         ...|      4|
|  ♦♦♦♦♦♦♦♦♦133|      1|
```

```
root@1e5e7524f7a8:/# hdfs dfs -ls /user/cloudera
Found 1 items
-rw-r--r-- 3 root cloudera 1853821 2022-11-19 15:58 /user/cloudera/annkarbook.txt
root@1e5e7524f7a8:/# hdfs dfs -chmod 764 /user/cloudera/annkarbook.txt
root@1e5e7524f7a8:/# hdfs dfs -ls /user/cloudera
Found 1 items
-rwxrw-r-- 3 root cloudera 1853821 2022-11-19 15:58 /user/cloudera/annkarbook.txt
root@1e5e7524f7a8:/# hdfs dfs -du -h /user/cloudera/annkarbook.txt
1.8 M /user/cloudera/annkarbook.txt
root@1e5e7524f7a8:/#
```