

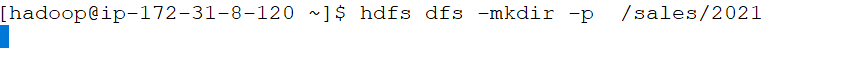
**Solution**:

|  |
| --- |
|   A business wants to know the ways to count number of different types of sports items  sold in last one year. Please refer the input file “SportItems.txt”  **Solution**: We need to mention InputFormat as TextInputFormat as given file is a  csv file in the driver class . In the Mapper class’s map() we need to define the key as  LongWritable and value as Text.  We have to do following steps in to achieve our objective   1. Covert the Text value object to String object 2. Split or tokenize String object by delimiter ‘,’ to obtain String[] 3. Make first field of String[0] as key(Text) and second field String[1] as   value(IntWritable)   1. Write these key and value to context.write(key,value) 2. In the Reducer class reduce() , iterate over the list of IntWritable and get the total 3. Finally write the key and total as value by calling context.write(key,total) |
|   A researcher wants to know the frequency of domestic air travel for various age groups.  let’s assume there are three age group are 18-30,30-50 and 50 & above.  Please refer the input file “Airtravel.csv”  **Solution**: We need to mention InputFormat as TextInputFormat as given file is a  csv file in the driver class . In the Mapper class’s map() we need to define the key as  LongWritable and value as Text.  We have to do following steps in to achieve our objective   1. Covert the Text value object to String object 2. Split or tokenize String object by delimiter ‘,’ to obtain String[] 3. Take the second field String[1] which is an age of the customer and apply logic   such as if age is between 18-30 then make key as “Young” , if 30-50 the key as  “Middle” and if age >50 then key as “Old”   1. Make value as IntWritable(1) Write these key and value to context.write(key,value) 2. In the Reducer class reduce() , iterate over the list of IntWritable and get the total 3. Finally write the key and total as value by calling context.write(key,total) |
|   A call center wants to know the number of occurrences of words “satisfied” or “good”  in their C-SAT survey. Please refer the input file “Survery.csv”  **Solution**: We need to mention InputFormat as TextInputFormat as given file is a  csv file in the driver class . In the Mapper class’s map() we need to define the key as  LongWritable and value as Text.  We have to do following steps in to achieve our objective   1. Covert the Text value object to String object 2. Split or tokenize String object by delimiter ‘,’ to obtain String[] 3. Take the second field String[1] which is a feedback of the customer and make   it as key and set value(IntWritable) as 1   1. Write these key and value to context.write(key,value) 2. In the Reducer class reduce() , iterate over the list of IntWritable and get the total 3. Finally write the key and total as value by calling context.write(key,total) |
|  |

**Hadoop command exercise**

1. Create a directory in HDFS whose path should be /sales/2020

hdfs dfs -mkdir -p /sales/2020



1. Create a Linux file and write some data into it as shown in the below screenshot.

Chart

Description automatically generated with medium confidence

Graphical user interface, text, application, chat or text message

Description automatically generated

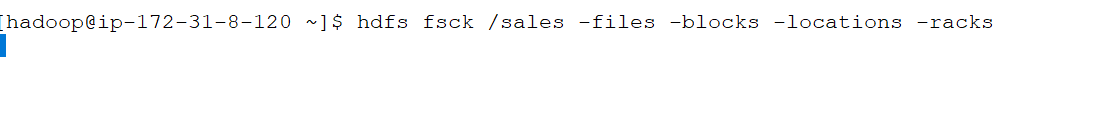
3. You save this Linux file into HDFS in the folder “/sales/2020” which you create in the step no 1.

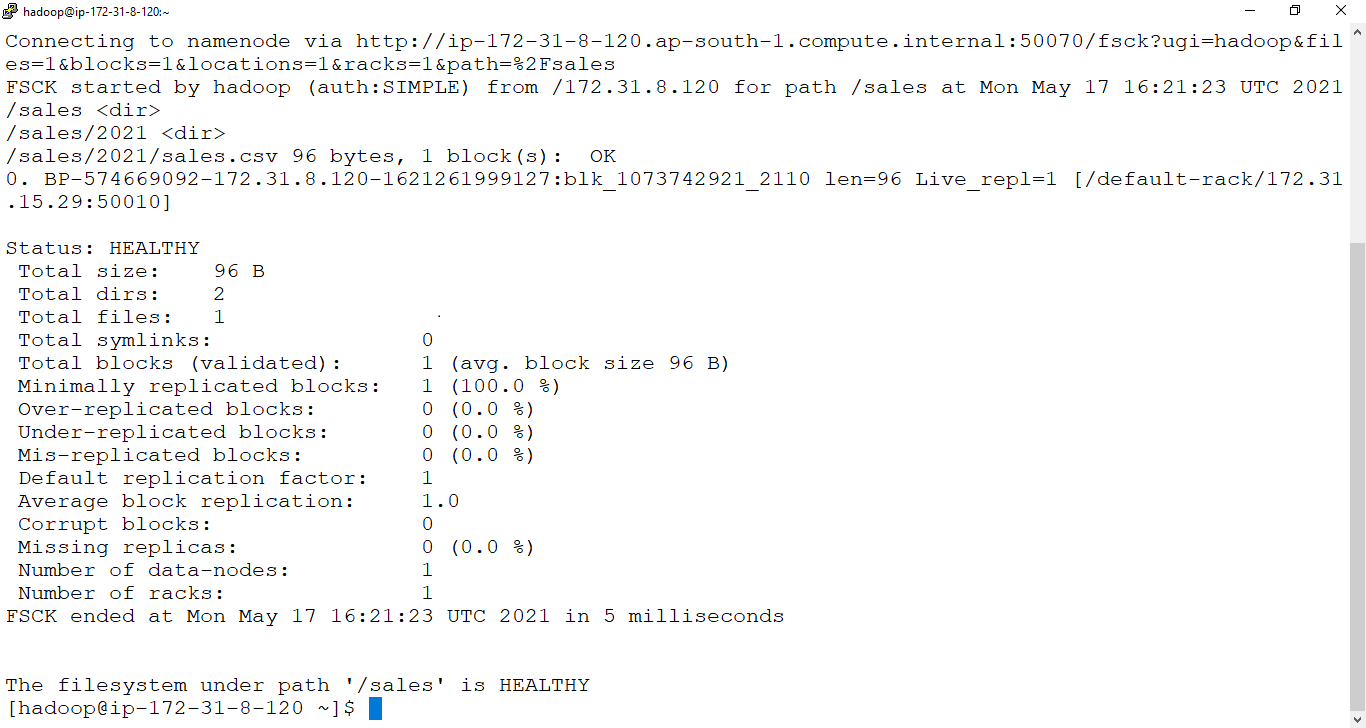
hdfs dfs -put sales.csv /sales/2020



4. Execute a command to verify number of files, block for each file, rack information of this folder “/sales”

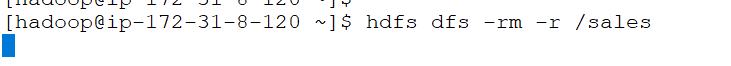
hdfs fsck /sales -files -blocks -locations -racks





5.

hdfs dfs -rm -r /sales



6

hdfs dfs -ls /sales

