

```

[cloudera@quickstart ~]$ mkdir TP [cloudera@quickstart ~]$ cd TP [cloudera@quickstart ~]$
mkdir code [cloudera@quickstart ~]$ mkdir data //// deplacez vous dans le rep ~/TP/data puis,
emporter le fichier purchases.txt dans votre dossier /data ///Commencez avec Hadoop
[cloudera@quickstart ~]$ hadoop dfs -ls ou bien : [cloudera@quickstart ~]$ hdfs fs -ls crer un
rep myinput: [cloudera@quickstart code]$ hdfs dfs -mkdir myinput copier le fichier puchases.txt
sous hdfs dans le rep myinput: hadoop fs -put data.txt /input/files/data.txt afficher le contenu du
rep myinput: [cloudera@quickstart ~]$ hadoop fs -ls myinput puis, visualisez les dernieres lignes
de purchases.txt: [cloudera@quickstart code]$ head -50 ../data/purchases.txt | ./mapper.py il
vous donne on sait jamais permission denied !!!!! do this command: chmod +x
/home/cloudera/TP/code/mapper.py than tape the command : [cloudera@quickstart code]$
head -50 ../data/purchases.txt | ./mapper.py -----Mapper-----
tapez le code python suivant dans le dosssier /TP/code/ -----code-----
#!/usr/bin/python import sys for line in sys.stdin: data = line.strip().split("\t") if len(data) == 6:
date, time, store, item, cost, payment = data print "{0}\t{1}".format(store,cost) -----fin code---
----- expliquer ce que permet de faire ce code: tester ce mapper en local sur les 50
premieres lignes du fichier purchases.txt en tapant l'instruction suivante: [cloudera@quickstart
code]$ head -50 ../data/purchases.txt | ./mapper.py -----Reducer-----
-----

```