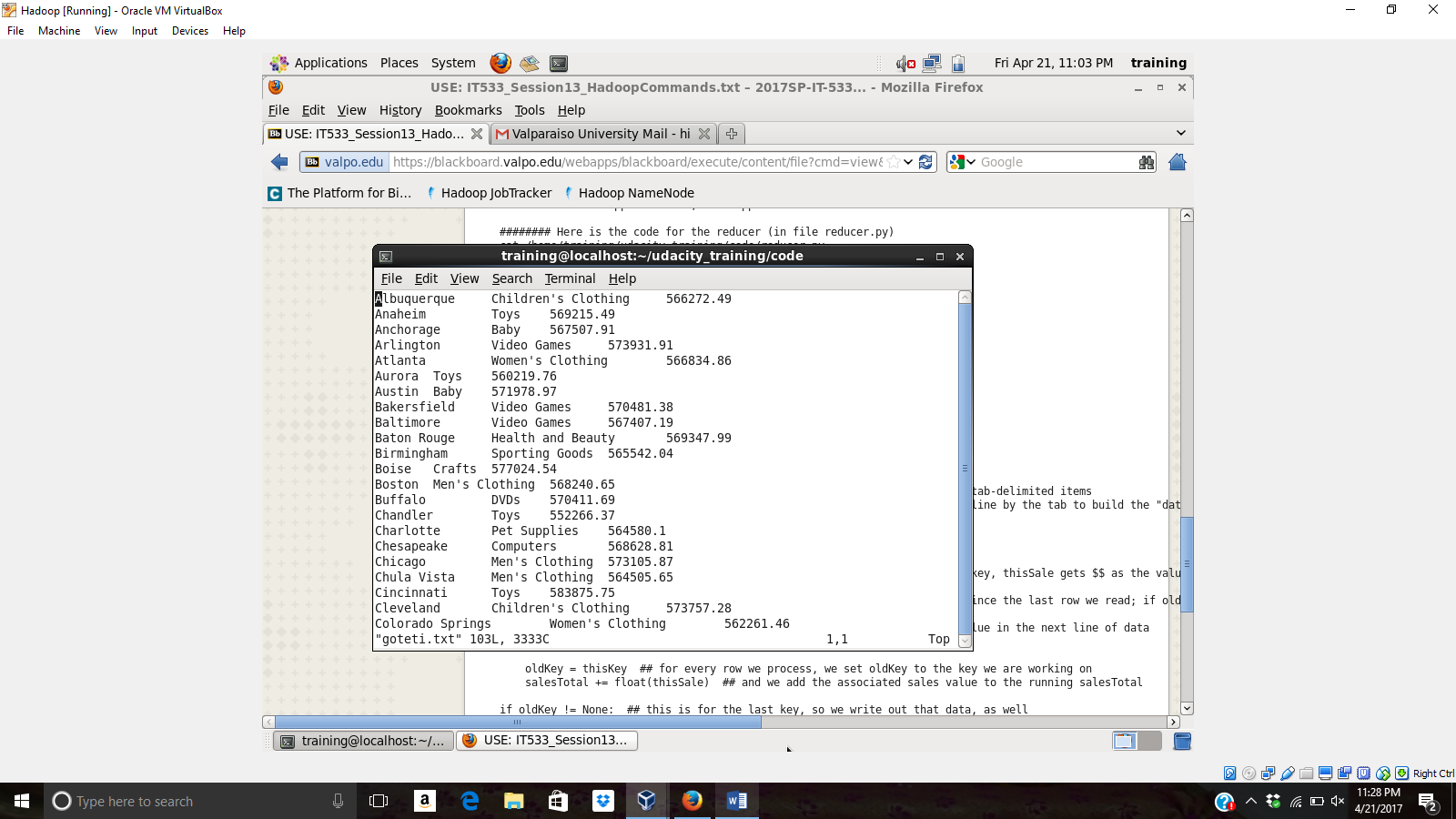
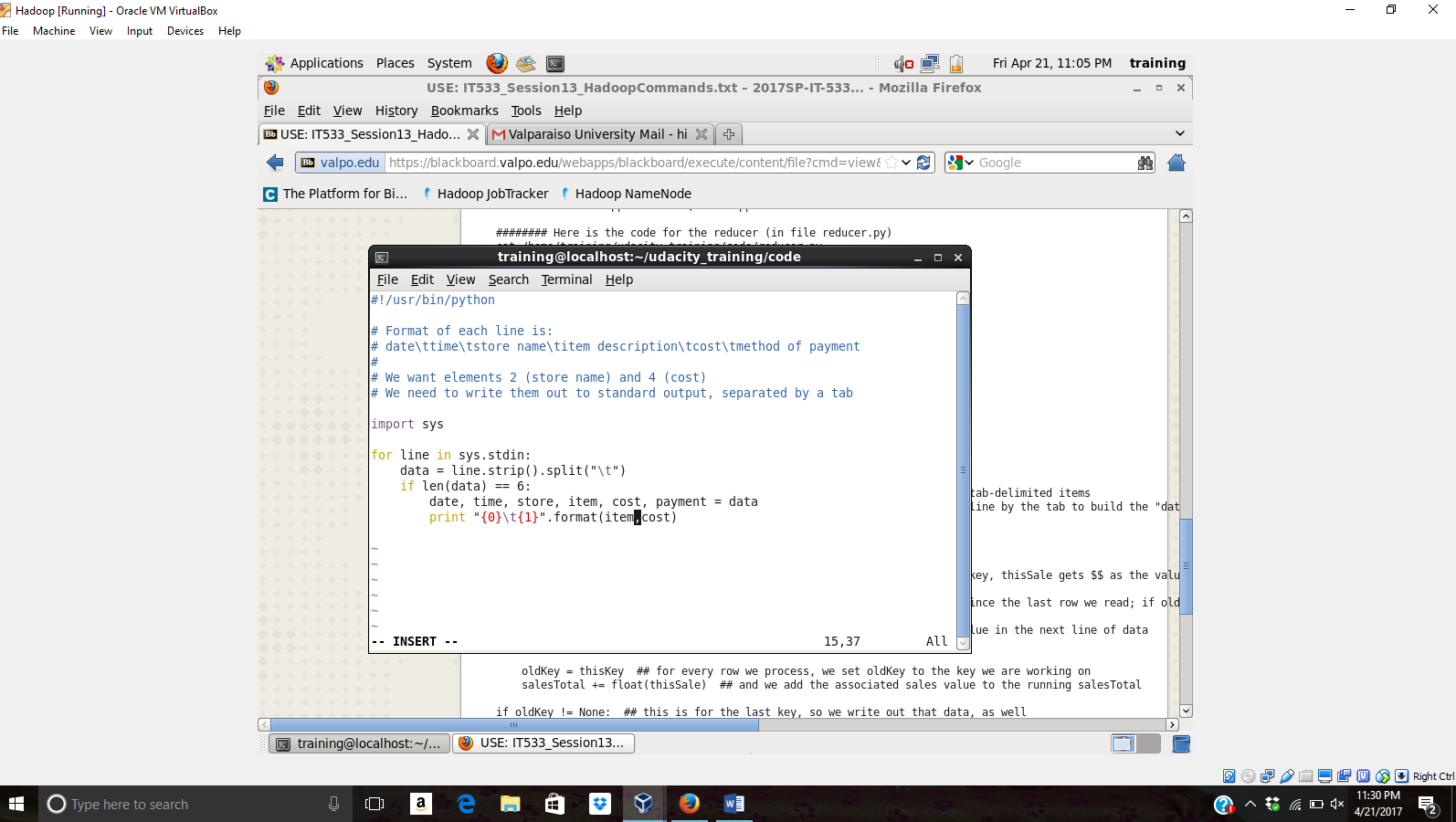
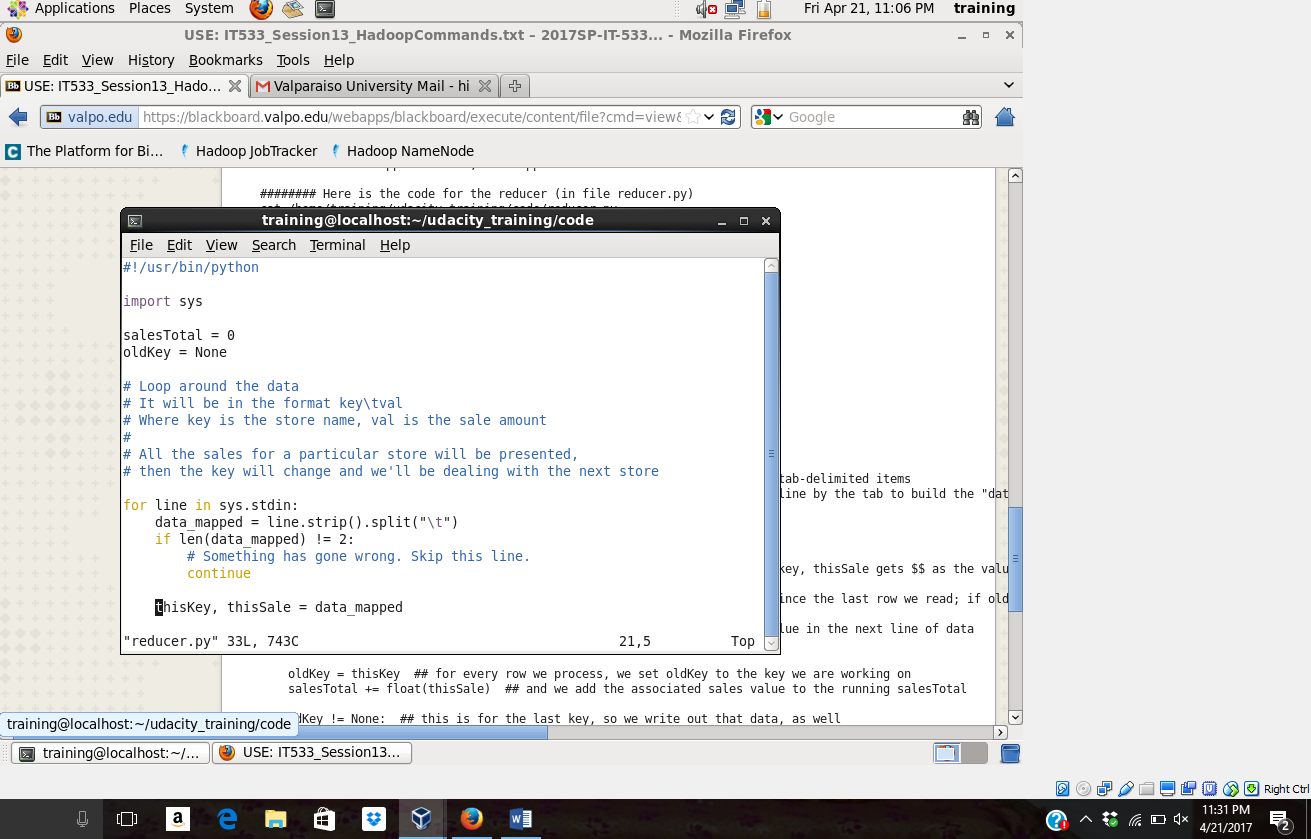
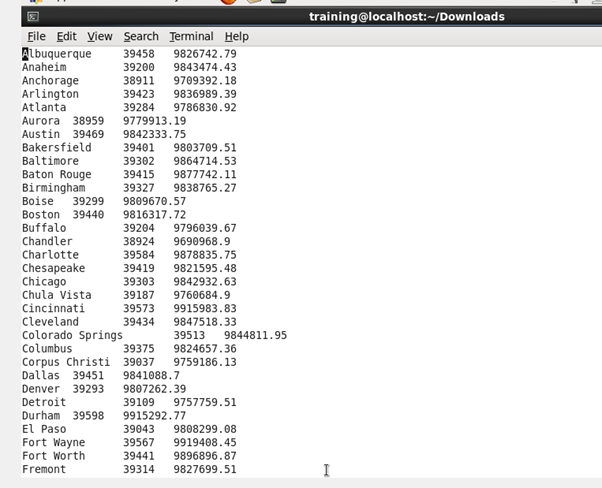
Using the purchases.txt data set that came with your Hadoop training VM, write a mapper.py and a reducer.py function

1. The three questions that you have to answer about this data set with your function output are:
2. Instead of breaking the sales down by store, instead provide a sales breakdown by product category across all of the store locations. Find the monetary value for the highest individual sale for each separate location.

Output of the running file :   
  
  
  
  
  
Mapper .py:   
  
  
  
  
  
  
  
  
  
Reducer.py:  
  
  
  
  
  
  
  
  
Value of identifying of each store:  
  
  
Mapper.py:  
  
#!/usr/bin/python  
  
import sys  
  
salesTotal = 0  
oldKey = None  
  
# Loop around the data  
# It will be in the format key\tval  
# Where key is the store name, val is the sale amount  
#  
# All the sales for a particular store will be presented,  
# then the key will change and we'll be dealing with the next store  
  
reducer.py  
  
  
for line in sys.stdin:  
    data\_mapped = line.strip().split("\t")  
    if len(data\_mapped) != 2:  
        # Something has gone wrong. Skip this line.  
        continue  
  
    thisKey, thisSale = data\_mapped  
  
"reducer.py" 33L, 743C                                        22,0-1        Top  
  
  
  
  
  
#!/usr/bin/python  
  
# Format of each line is:  
# date\ttime\tstore name\titem description\tcost\tmethod of payment  
#  
# We want elements 2 (store name) and 4 (cost)  
# We need to write them out to standard output, separated by a tab  
  
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(item,cost)  
  
~  
~  
~  
~  
~  
~  
                                                              15,36         All

1. Find the total sales value across all the locations, and the total number of sales.



1. Display your name at the beginning and the end of your output files.

Upload your mapper.py and reducer.py file and your output files.

