

CSCI 330 ASSIGNMENT 8 (SPRING 2021)

GENERATING REPORTS WITH awk (100 PTS)

OVERVIEW

Write an awk script that computes and displays a summary of the performance of sales associates for a company based on data in the format specified below. If you are lost, remember that awk loads the input file record by record and does the action of any records that match one of the patterns. If you want to put data from multiple records together, you will need to use some sort of variable.

The awk script will be invoked from the command line as shown, with an input file that contains data on a company's sales department:

```
% awk -f zxxxxxx.awk inputfile
```

FORMAT SPECIFICATION

The records in the input file will use the semicolon, “:”, as their field separator, and the newline character as their record separator. The records will all be contained in the same file, and will be of one of the following three types:

- ① **Associates** – Each salesperson working with the company will have a record in the following format:

#	Field Name	Field Description
1	associateid	An decimal integer, uniquely identifying this salesperson
2	name	Alphanumeric text containing the name of the salesperson, first name first.
3	position	Alphanumeric text containing the position/rank of this salesperson

- ② **Products** – Every product that is available to be sold will have a record in the following format:

#	Field Name	Field Description
1	productid	A decimal integer uniquely identifying this product
2	category	A string containing the name or code for the category this item falls into
3	name	A string containing the name of the product
4	price	A decimal number with two digits after the decimal point containing the price of the product per item in US dollars

- ③ **Transactions** – Every time a product is sold, there will be at least one record in the following format. If there are multiple records with the same transactionid, the associateid should be the same and the productid should be different. These records would denote that several products were sold as part of the same order.

#	Field Name	Field Description
1	transactionid	An integer number identifying the current transaction
2	productid	An integer that will match the productid in the record for the product sold
3	associateid	An integer number identifying which associate made the transaction
4	date	A date in the form MM/DD/YYYY, so 07/04/1776 would mean July 4, 1776
5	quantity	An integer number that will contain the quantity of this product sold in this transaction

REQUIREMENTS

Your script should have awk process the input file and print out a report that includes the following:

- ① A title, printed before anything else is read in.

- ② Labels for each of the fields of the table that follows, neatly-aligned.
- ③ A line drawn with “=” separating the labels from the table data.
- ④ A neatly-formatted table, sorted in order of total sales, highest first, containing
 - ▶ The name of each associate that made any sales during the year 2020, last name first, with a comma and a space between.
 - ▶ The ID number of each of those associates.
 - ▶ The total dollar value of all of the sales that they made in 2020.
- ⑤ A line drawn with “=”s separating the table data from the summary below.
- ⑥ A neatly-aligned summary of:
 - ▶ How many *Associate* records were read in from the input file.
 - ▶ How many *Product* records were read in from the input file.
 - ▶ How many *Transaction* records were read in from the input file.

Do not assume that the records will come in any particular order. They may be intermixed in any order within the data file. Your script should be able to recognize which type of record based on the number of fields.

EXAMPLE OUTPUT

An example of what the output should look like with the salesdb example input file that is provided on Blackboard. It will be tested with other data files as well, so make sure your awk script is doing the work.

```
% awk -f zxxxxxx.awk salesdb
Summary of Sales for 2020

Name                ID      Sales
=====
Smith, Matt         4    990999.99
Jones, Davy         1     2659.54
Smith, Samantha     3     1159.81
Davis, Ricky        2      299.98
=====
                        8 Associates
                        6 Products
                        23 Transactions
```

ERROR CHECKING

If any error is encountered, such as incorrect data in the input file, just skip the lines in question.

ADDITIONAL NOTES

- ▶ The order in which the associates are listed is not significant. They will be identified by their IDs.
- ▶ Make sure that your assignment is contained in a single file called zxxxxxx.awk (but based on *your* zid).
- ▶ Make sure that your awk script runs properly on the department Linux servers, turing/hopper.
- ▶ Make sure that your awk file is a regular Unix text file. If you made the file on Windows, you will likely need to use dos2unix to fix the file. Check the manpage if you need it.

WHAT TO TURN IN?

Submit your assignment as a single file, named as required above, via Blackboard before the deadline.