Answered by: _Jianying Li_____

Data Management Written

It is to be understood that these questions are not to be posted on web sites, etc., or disclosed to others and that all assistance utilized is to be disclosed. Although this should not have to be stated, from hard experience, candidates should please understand that compromising these suggestions does them no good as there are sufficient following actions to detect what candidates can do on their own.

Initial here to show the above was read:_____JYL_____

1. You have a very large file, named 'ColCheckMe', tab-delimited, that you are asked to process.

You are told that each line in 'ColCheckMe' has 7 columns, and that the values in the 5th column are integers.

Using shell functions (and standard LINUX/UNIX filters), indicate how you would verify that these conditions were satisfied in 'ColCheckMe'

Part I:

awk -F'\t' '{print NF;exit}' ColCheckMe

Should return 7 if there were seven columns

Part II:

cut -f 5 ColCheckMe | grep -v [0-9] | wc –l > row_not_integer

This file "row_not_integer" should be empty, otherwise, there is error.

cut -f 5 ColCheckMe | wc –l > total_number_row

2. In the same file, you are told that each value in column 1 is unique. How would you verify that?

Cut –f 1 ColCheckMe | Wc –l > original_line_count

Cut –f 1 ColCheckMe | sort | uniq | Wc –l > unique_line_count

Diff original_line_count unique_line_count

3. Write a shell function that counts the number of occurrences of the word "SpecStr" in the file 'ColCheckMe'.

Create a bash file called "sample_function.sh" as following, and invoke as "bash sample_function.sh ColCheckMe"

Answered by: _Jianying Li_____

```
#!/bin/bash
awk 'BEGIN{
 search="SpecStr"
 total=0
}
NR%10==0{
  c=gsub(search,"",s)
  total+=c
}
NR{ s=s $0 }
END{
 c=gsub(search,"",s)
 print "total count: "total+c
}'

[jl407@compute1 test_doc]$ more  ColCheckMe
SpecStr col2    SpecStr col4    5       col6    col7
SpecStr somehin SpecStr2        2       7       _SpecStr        what
some    2       8       4       2       SpecCstr        SpecStr

[jl407@compute1 test_doc]$ cat ColCheckMe |  bash count_occur_awk.sh
total count: 6
```

4. Please show how you would this in PERL, or related language.

Please see four perl scripts attached for each part:

Q1, part I: validate_7cols.pl
Q1, part II: validate_col5_int.pl
Q2: validate_col1_uniq.pl
Q3: count_str.pl

```
[jl407@compute1 test_doc]$ perl count_str.pl ColCheckMe
6
[jl407@compute1 test_doc]$ perl validate_7cols.pl ColCheckMe

[jl407@compute1 test_doc]$ perl validate_col5_int.pl ColCheckMe

[jl407@compute1 test_doc]$ perl validate_col1_uniq.pl ColCheckMe
Column one contains repeat entry
```

5. How would you log the output (regular output as well as error messages) of a unix program into a file?

```
./script.sh >     output.log
```

Answered by: _Jianying Li_____

./script.sh >& error.log

Or, I can always write standout to a log, and standerr to an error log file.

[Questions 6-8 for candidates familiar with SQL]
Given 2 tables:

Person (id integer, first char(20), last char(20), address char(50),
city char(30), state char(2), zip char(5))
Capital (state char(2), city char(30)

6. List all last names that start with 'AB' (it is preferred that the result is a list of last names only)

Select last from Person where last like 'AB%' ;

7. List all states and number of persons (from Person table) from each.

Select count(*), state from Person group by state;

8. List all people who live in the capital of their state.

Select * from person as p, capital as c where c.city = p.city;