Week2_DataPipelines

Exercise 1 - Extracting data using cut command

The filter command cut helps us extract selected characters or fields from a line of text.

1.1 Extracting characters.

The command below shows how to extract the first four characters.

```
echo "database" | cut -c1-4
```

The command below shows how to extract 5th to 8th characters.

```
echo "database" | cut -c5-8
```

Non-contiguous characters can be extracted using the comma.

The command below shows how to extract the 1st and 5th characters.

```
echo "database" | cut -c1,5
```

1.2. Extracting fields/columns

We can extract a specific column/field from a delimited text file, by mentioning

the delimiter using the -d option, or
the field number using the -f option.
The /etc/passwd is a ":" delimited file.

The command below extracts user names (the first field) from /etc/passwd.

```
cut -d":" -f1 /etc/passwd
```

The command below extracts multiple fields 1st, 3rd, and 6th (username, userid, and home directory) from /etc/passwd.

```
cut -d":" -f1,3,6 /etc/passwd
```

The command below extracts a range of fields 3rd to 6th (userid, groupid, user description and home directory) from /etc/passwd.

```
cut -d":" -f3-6 /etc/passwd
```

Exercise 2 - Transforming data using tr.

tr is a filter command used to translate, squeeze, and/or delete characters.

2.1. Translate from one character set to another

The command below translates all lower case alphabets to upper case.

```
echo "Shell Scripting" | tr "[a-z]" "[A-Z]"
```

You could also use the pre-defined character sets also for this purpose:

```
echo "Shell Scripting" | tr "[:lower:]" "[:upper:]"
```

The command below translates all upper case alphabets to lower case.

```
echo "Shell Scripting" | tr "[A-Z]" "[a-z]"
```

2.2. Squeeze repeating occurrences of characters

The -s option replaces a sequence of a repeated characters with a single occurrence of that character.

The command below replaces repeat occurrences of 'space' in the output of ps command with one 'space'.

```
ps | tr -s " "
```

In the above example, the space character within quotes can be replaced with the following : "[:space:]".

2.3. Delete characters

We can delete specified characters using the -d option.

The command below deletes all digits.

```
echo "My login pin is 5634" | tr -d "[:digit:]"
```

- The Browser with pgadmin
- URL: https://joaopaulonob-5050.theiadocker-1-labs-prod-theiak8s-4-tor01.proxy.cognitiveclass.ai/browser/
- Database Password: NTc1MC1qb2FvcGF1
- CommandLine: psql --username=postgres --host=localhost