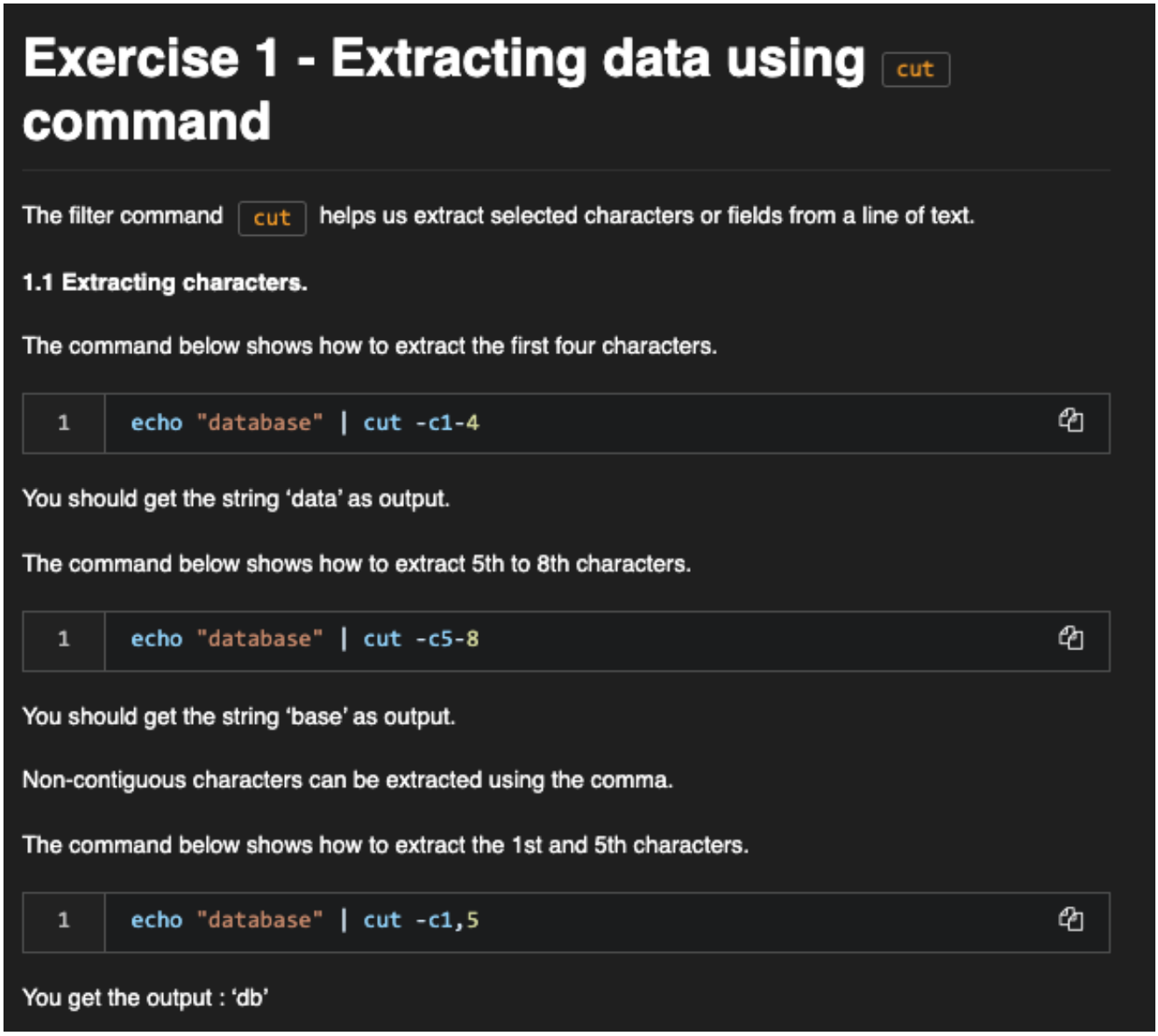
Hands-On Lab: ETL using Shell Scripts

Exercise 1 Extract and Slice Like Cut Operator for file Data



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.

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

You should get the string 'data' as output.

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

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

You should get the string 'base' as output.

Non-contiguous characters can be extracted using the comma.

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

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

You get the output : 'db'

```
theia@theiadocker-craigtrupp8:/home/project$ echo 'database'
database
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1-4
data
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c5-8
```

```
base
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1, 3
cut: byte/character positions are numbered from 1
Try 'cut --help' for more information.
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1,3
dt
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1,c5-8
cut: invalid byte/character position 'c5-8'
Try 'cut --help' for more information.
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1,3-8
dtabase
theia@theiadocker-craigtrupp8:/home/project$
```

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.

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

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

```
1  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.

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

```
theia@theiadocker-craigtrupp8:/home/project$ vim /etc/passwd
```

```
bash: vim: command not found
theia@theiadocker-craigtrupp8:/home/project$ vi /etc/passwd
bash: vi: command not found
theia@theiadocker-craigtrupp8:/home/project$ cat /etc/passwd
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
sys:x:3:3:sys:/dev:/usr/sbin/nologin
sync:x:4:65534:sync:/bin:/bin/sync
games:x:5:60:games:/usr/games:/usr/sbin/nologin
man:x:6:12:man:/var/cache/man:/usr/sbin/nologin
lp:x:7:7:lp:/var/spool/lpd:/usr/sbin/nologin
mail:x:8:8:mail:/var/mail:/usr/sbin/nologin
news:x:9:9:news:/var/spool/news:/usr/sbin/nologin
uucp:x:10:10:uucp:/var/spool/uucp:/usr/sbin/nologin
proxy:x:13:13:proxy:/bin:/usr/sbin/nologin
www-data:x:33:33:www-data:/var/www:/usr/sbin/nologin
backup:x:34:34:backup:/var/backups:/usr/sbin/nologin
list:x:38:38:Mailing List Manager:/var/list:/usr/sbin/nologin
irc:x:39:39:ircd:/var/run/ircd:/usr/sbin/nologin
gnats:x:41:41:Gnats Bug-Reporting System
(admin):/var/lib/gnats:/usr/sbin/nologin
nobody:x:65534:65534:nobody:/nonexistent:/usr/sbin/nologin
_apt:x:100:65534::/nonexistent:/usr/sbin/nologin
messagebus:x:101:102::/nonexistent:/usr/sbin/nologin
theia:x:1000:1000:,,,:/home/theia:/bin/bash
mongodb:x:102:106::/var/lib/mongodb:/usr/sbin/nologin
ntp:x:103:107::/nonexistent:/usr/sbin/nologin
cassandra:x:104:108:Cassandra
database,,,:/var/lib/cassandra:/usr/sbin/nologin
postgres:x:105:109:PostgreSQL
administrator,,,:/var/lib/postgresql:/bin/bash
theia@theiadocker-craigtrupp8:/home/project$ cut -d":" -f1 /etc/passwd
root
daemon
bin
sys
sync
games
man
lp
mail
news
```

```
uucp
proxy
www-data
backup
list
irc
gnats
nobody
_apt
messagebus
theia
mongodb
ntp
cassandra
postgres
theia@theiadocker-craigtrupp8:/home/project$ cut -d":" -f1,3,6 /etc/passwd
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
man:6:/var/cache/man
lp:7:/var/spool/lpd
mail:8:/var/mail
news:9:/var/spool/news
uucp:10:/var/spool/uucp
proxy:13:/bin
www-data:33:/var/www
backup:34:/var/backups
list:38:/var/list
irc:39:/var/run/ircd
gnats:41:/var/lib/gnats
nobody:65534:/nonexistent
_apt:100:/nonexistent
messagebus:101:/nonexistent
theia:1000:/home/theia
mongodb:102:/var/lib/mongodb
ntp:103:/nonexistent
cassandra:104:/var/lib/cassandra
postgres:105:/var/lib/postgresql
theia@theiadocker-craigtrupp8:/home/project$ cut -d":" -f 1,3,6 /etc/passwd
root:0:/root
```

```
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
man:6:/var/cache/man
lp:7:/var/spool/lpd
mail:8:/var/mail
news:9:/var/spool/news
uucp:10:/var/spool/uucp
proxy:13:/bin
www-data:33:/var/www
backup:34:/var/backups
list:38:/var/list
irc:39:/var/run/ircd
gnats:41:/var/lib/gnats
nobody:65534:/nonexistent
_apt:100:/nonexistent
messagebus:101:/nonexistent
theia:1000:/home/theia
mongodb:102:/var/lib/mongodb
ntp:103:/nonexistent
cassandra:104:/var/lib/cassandra
postgres:105:/var/lib/postgresql
theia@theiadocker-craigtrupp8:/home/project$ cut -d":" -f 3-6 /etc/passwd
0:0:root:/root
1:1:daemon:/usr/sbin
2:2:bin:/bin
3:3:sys:/dev
4:65534:sync:/bin
5:60:games:/usr/games
6:12:man:/var/cache/man
7:7:lp:/var/spool/lpd
8:8:mail:/var/mail
9:9:news:/var/spool/news
10:10:uucp:/var/spool/uucp
13:13:proxy:/bin
33:33:www-data:/var/www
34:34:backup:/var/backups
38:38:Mailing List Manager:/var/list
39:39:ircd:/var/run/ircd
41:41:Gnats Bug-Reporting System (admin):/var/lib/gnats
65534:65534:nobody:/nonexistent
```

```
100:65534::/nonexistent
101:102::/nonexistent
1000:1000:,,,:/home/theia
102:106::/var/lib/mongodb
103:107::/nonexistent
104:108:Cassandra database,,,:/var/lib/cassandra
105:109:PostgreSQL administrator,,,:/var/lib/postgresql
theia@theiadocker-craigtrupp8:/home/project$
```

Exercise 2 : Transform data using **tr**

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.

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

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

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

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

```
1  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'.

```
1  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.

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

The output will be : 'My login pin is'

```
theia@theiadocker-craigtrupp8:/home/project$ echo 'Shell Scripting' | tr
'[a-z]' '[A-Z]'
SHELL SCRIPTING
theia@theiadocker-craigtrupp8:/home/project$ echo "Shell Scripting" | tr
"[a-z]" "[A-Z]"
SHELL SCRIPTING
theia@theiadocker-craigtrupp8:/home/project$ echo 'Shell Scripting' | tr
'[:lower:]' '[:upper:]'
SHELL SCRIPTING
theia@theiadocker-craigtrupp8:/home/project$ echo "Shell Scripting" | tr
"[:lower:]" "[:upper:]"
SHELL SCRIPTING
theia@theiadocker-craigtrupp8:/home/project$ echo 'Shell Scripting' | tr
"[:upper:]" "[:lower:]"
shell scripting
theia@theiadocker-craigtrupp8:/home/project$ ps | tr -s " "
 PID TTY TIME CMD
 322 pts/0 00:00:00 bash
 652 pts/0 00:00:00 ps
 653 pts/0 00:00:00 tr
theia@theiadocker-craigtrupp8:/home/project$ ps
    PID TTY          TIME CMD
    322 pts/0    00:00:00 bash
    654 pts/0    00:00:00 ps
theia@theiadocker-craigtrupp8:/home/project$ echo "My login pin is 5634" |
tr -d '[:digit:]'
My login pin is
theia@theiadocker-craigtrupp8:/home/project$
```

Exercise 3 - Start the PostgreSQL Database

Exercise 3 - Start the PostgreSQL database.

On the terminal run the following command to start the PostgreSQL database.

```
1  start_postgres
```

Note down the access information presented towards the end of these messages, especially the **CommandLine:**.

A sample commandline displayed looks as given below.

```
1  `psql --username=postgres --host=localhost`
```

Running this command from the shell prompt will start the interactive `psql` client which connects to the PostgreSQL server.

```
theia@theiadocker-craigtrupp8:/home/project$ start_postgres
Unable to find image 'ubuntu:latest' locally
latest: Pulling from library/ubuntu
3153aa388d02: Pull complete
Digest:
sha256:0bced47fffa3361afa981854fcabcd4577cd43cebbb808cea2b1f33a3dd7f508
Status: Downloaded newer image for ubuntu:latest
Starting your Postgres database....
This process can take up to a minute.

Postgres database started, waiting for all services to be ready....

Your Postgres database is now ready to use and available with username:
postgres password: MTk2Ni1j******

You can access your Postgres database via:
 • The Browser with pgadmin
   • URL:
https://craigtrupp8-5050.theiadocker-0-labs-prod-theiak8s-4-tor01.proxy.cog
nitiveclass.ai/browser/
   • Database Password: MTk2Ni1j******
 • CommandLine: psql --username=postgres --host=localhost
```

```
theia@theiadocker-craigtrupp8:/home/project$ psql --username=postgres
--host=localhost
psql (15.2 (Ubuntu 15.2-1.pgdg18.04+1), server 13.2)
Type "help" for help.

postgres=#
```

Exercise 4 - Create a table

Exercise 4 - Create a table

In this exercise we will create a table called '**users**' in the PostgreSQL database. This table will hold the user account information.

The table 'users' will have the following columns:

1. uname
2. uid
3. home

Step 1: Connect to the database server

Use the connection string saved in the previous exercise to connect to the PostgreSQL server.

Run the command below to login to PostgreSQL server.

```
1  psql --username=postgres --host=localhost
```

You will get the psql prompt: 'postgres=#'

Step 2: Connect to a database.

We will use a database called **template1** which is already available by default.

To connect to this database, run the following command at the 'postgres=#' prompt.

```
1  \c template1
```

You will get the following message.

```
You are now connected to database "template1" as user "postgres".
```

Also, your prompt will change to 'template1=#'.

Step 3: Create the table

Run the following statement at the 'template1=#' prompt:

```
1  create table users(username varchar(50),userid int,homedirectory varch
```

If the table is created successfully, you will get the message below.

```
CREATE TABLE
```

Step 4: Quit the psql client

To exit the psql client and come back to the Linux shell, run the following command:

```
1  \q
```

```
theia@theiadocker-craigtrupp8:/home/project$ psql --username=postgres
--host=localhost
psql (15.2 (Ubuntu 15.2-1.pgdg18.04+1), server 13.2)
Type "help" for help.

postgres=# \dt
Did not find any relations.
postgres=# \c template1
psql (15.2 (Ubuntu 15.2-1.pgdg18.04+1), server 13.2)
You are now connected to database "template1" as user "postgres".
template1=# create table users(username varchar(50), userid int,
homedirectory varchar(100));
CREATE TABLE
template1=# \dt
         List of relations
 Schema | Name  | Type  |  Owner
--------+-------+-------+----------
 public | users | table | postgres
(1 row)

template1=# \q
theia@theiadocker-craigtrupp8:/home/project$
```

Exercise 5 - Loading data into a PostgreSQL table.

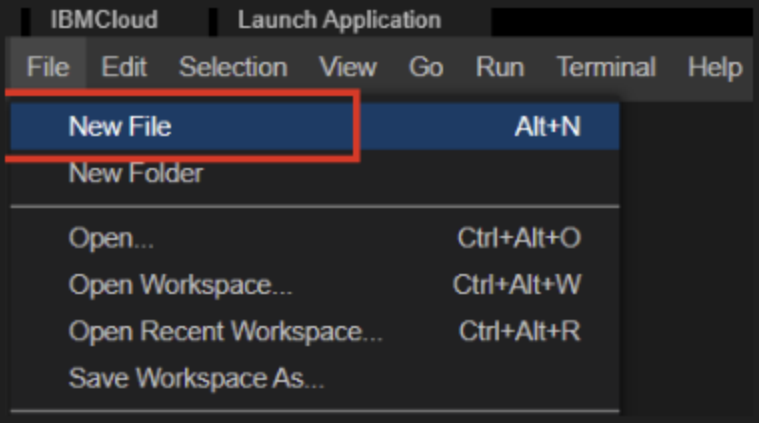
Exercise 5 - Loading data into a PostgreSQL table.

In this exercise, we will create a shell script which does the following.

- Extract the user name, user id, and home directory path of each user account defined in the /etc/passwd file.
- Save the data into a comma separated (CSV) format.
- Load the data in the csv file into a table in PostgreSQL database.

5.1. Create a shell script

Step 1: On the menu on the lab screen, use **File->New File** to create a new file.



Step 2: Give the name as 'csv2db.sh' and click 'OK'.

Step 3: State the objective of the script using comments.

Copy and paste the following lines into the newly created file.

```
# This script
# Extracts data from /etc/passwd file into a CSV file.

# The csv data file contains the user name, user id and
# home directory of each user account defined in /etc/passwd

# Transforms the text delimiter from ":" to ",".
# Loads the data from the CSV file into a table in PostgreSQL database.
```

Step 4: Save the file using the **File->Save** menu option.

5.2. Extract required user information from /etc/passwd

In this step, we will extract user name (field 1), user id (field 3), and home directory path (field 6) from /etc/passwd file using the `cut` command.

The /etc/passwd has ":" symbol as the column separator.

Copy the following lines and add them to the end of the script.

```
# Extract phase

echo "Extracting data"

# Extract the columns 1 (user name), 2 (user id) and
# 6 (home directory path) from /etc/passwd

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

Save the file.

Run the script.

```
bash csv2db.sh
```

Verify that the output contains the three fields, that we extracted.

5.3. Redirect the extracted output into a file.

In this step, we will redirect the extracted data into a file named `extracted-data.txt`

Replace the cut command at end of the script with the following command.

```
cut -d":" -f1,3,6 /etc/passwd > extracted-data.txt
```

Save the file.

Run the script.

```
bash csv2db.sh
```

Run the command below to verify that the file `extracted-data.txt` is created, and has the content.

```
cat extracted-data.txt
```

5.4. Transform the data into CSV format

The extracted columns are separated by the original ":" delimiter.

We need to convert this into a "," delimited file.

Add the below lines at the end of the script

```
# Transform phase
echo "Transforming data"
# read the extracted data and replace the colons with commas.

tr ":" "," < extracted-data.txt
```

Save the file.

Run the script.

```
bash csv2db.sh
```

Verify that the output contains ',' in place of ":".

Replace the tr command at end of the script with the command below.

```
tr ":" "," < extracted-data.txt > transformed-data.csv
```

Save the file.

Run the script.

```
bash csv2db.sh
```

Run the command below to verify that the file `transformed-data.csv` is created, and has the content.

```
cat transformed-data.csv
```

5.5. Load the data into the table 'users' in PostgreSQL

To load data from a shell script, we will use the `psql` client utility in a non-interactive manner.

This is done by sending the database commands through a command pipeline to `psql` with the help of `echo` command.

Step 1: Add the copy command

PostgreSQL command to copy data from a CSV file to a table is `COPY` .

The basic structure of the command which we will use in our script is,

```
1  COPY table_name FROM 'filename' DELIMITERS 'delimiter_character' FORMA
```

Now, add the lines below to the end of the script 'csv2db.sh'.

```
1  # Load phase
2  echo "Loading data"
3  # Send the instructions to connect to 'template1' and
4  # copy the file to the table 'users' through command pipeline.
5
6  echo "\c template1;\COPY users  FROM '/home/project/transformed-data.c DE
```

- Terminal Session

```
theia@theiadocker-craigtrupp8:/home/project$ echo 'database'
database
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1-4
data
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c5-8
base
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1, 3
cut: byte/character positions are numbered from 1
Try 'cut --help' for more information.
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1,3
dt
```

```
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1,c5-8
cut: invalid byte/character position 'c5-8'
Try 'cut --help' for more information.
theia@theiadocker-craigtrupp8:/home/project$ echo 'database' | cut -c1,3-8
dtabase
theia@theiadocker-craigtrupp8:/home/project$ pwd
/home/project
theia@theiadocker-craigtrupp8:/home/project$ ls
theia@theiadocker-craigtrupp8:/home/project$ vim /etc/passwd
bash: vim: command not found
theia@theiadocker-craigtrupp8:/home/project$ vi /etc/passwd
bash: vi: command not found
theia@theiadocker-craigtrupp8:/home/project$ cat /etc/passwd
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
sys:x:3:3:sys:/dev:/usr/sbin/nologin
sync:x:4:65534:sync:/bin:/bin/sync
games:x:5:60:games:/usr/games:/usr/sbin/nologin
man:x:6:12:man:/var/cache/man:/usr/sbin/nologin
lp:x:7:7:lp:/var/spool/lpd:/usr/sbin/nologin
mail:x:8:8:mail:/var/mail:/usr/sbin/nologin
news:x:9:9:news:/var/spool/news:/usr/sbin/nologin
uucp:x:10:10:uucp:/var/spool/uucp:/usr/sbin/nologin
proxy:x:13:13:proxy:/bin:/usr/sbin/nologin
www-data:x:33:33:www-data:/var/www:/usr/sbin/nologin
backup:x:34:34:backup:/var/backups:/usr/sbin/nologin
list:x:38:38:Mailing List Manager:/var/list:/usr/sbin/nologin
irc:x:39:39:ircd:/var/run/ircd:/usr/sbin/nologin
gnats:x:41:41:Gnats Bug-Reporting System
(admin):/var/lib/gnats:/usr/sbin/nologin
nobody:x:65534:65534:nobody:/nonexistent:/usr/sbin/nologin
_apt:x:100:65534::/nonexistent:/usr/sbin/nologin
messagebus:x:101:102::/nonexistent:/usr/sbin/nologin
theia:x:1000:1000:,,,:/home/theia:/bin/bash
mongodb:x:102:106::/var/lib/mongodb:/usr/sbin/nologin
ntp:x:103:107::/nonexistent:/usr/sbin/nologin
cassandra:x:104:108:Cassandra
database,,,:/var/lib/cassandra:/usr/sbin/nologin
postgres:x:105:109:PostgreSQL
administrator,,,:/var/lib/postgresql:/bin/bash
theia@theiadocker-craigtrupp8:/home/project$ cut -d":" -f1 /etc/passwd
root
```

```
daemon
bin
sys
sync
games
man
lp
mail
news
uucp
proxy
www-data
backup
list
irc
gnats
nobody
_apt
messagebus
theia
mongodb
ntp
cassandra
postgres
theia@theiadocker-craigtrupp8:/home/project$ cut -d":" -f1,3,6 /etc/passwd
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
man:6:/var/cache/man
lp:7:/var/spool/lpd
mail:8:/var/mail
news:9:/var/spool/news
uucp:10:/var/spool/uucp
proxy:13:/bin
www-data:33:/var/www
backup:34:/var/backups
list:38:/var/list
irc:39:/var/run/ircd
gnats:41:/var/lib/gnats
nobody:65534:/nonexistent
```

```
_apt:100:/nonexistent
messagebus:101:/nonexistent
theia:1000:/home/theia
mongodb:102:/var/lib/mongodb
ntp:103:/nonexistent
cassandra:104:/var/lib/cassandra
postgres:105:/var/lib/postgresql
theia@theiadocker-craigtrupp8:/home/project$ cut -d":" -f 1,3,6 /etc/passwd
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
man:6:/var/cache/man
lp:7:/var/spool/lpd
mail:8:/var/mail
news:9:/var/spool/news
uucp:10:/var/spool/uucp
proxy:13:/bin
www-data:33:/var/www
backup:34:/var/backups
theia@theiadocker-craigtrupp8:/home/project$ start_postgres
Unable to find image 'ubuntu:latest' locally
latest: Pulling from library/ubuntu
3153aa388d02: Pull complete
Digest:
sha256:0bced47fffa3361afa981854fcabcd4577cd43cebbb808cea2b1f33a3dd7f508
Status: Downloaded newer image for ubuntu:latest
Starting your Postgres database....
This process can take up to a minute.

Postgres database started, waiting for all services to be ready....

Your Postgres database is now ready to use and available with username:
postgres password: MTk2Ni1jcmFpZ3Ry

You can access your Postgres database via:
 • The Browser with pgadmin
   • URL:
https://craigtrupp8-5050.theiadocker-0-labs-prod-theiak8s-4-tor01.proxy.cog
nitiveclass.ai/browser/
   • Database Password: MTk2Ni1jcmFpZ3Ry
```

```
 • CommandLine: psql --username=postgres --host=localhost
theia@theiadocker-craigtrupp8:/home/project$ ls
csv2db.sh  postgres
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
man:6:/var/cache/man
lp:7:/var/spool/lpd
mail:8:/var/mail
news:9:/var/spool/news
uucp:10:/var/spool/uucp
proxy:13:/bin
www-data:33:/var/www
backup:34:/var/backups
list:38:/var/list
irc:39:/var/run/ircd
gnats:41:/var/lib/gnats
nobody:65534:/nonexistent
_apt:100:/nonexistent
messagebus:101:/nonexistent
theia:1000:/home/theia
mongodb:102:/var/lib/mongodb
```

```
ntp:103:/nonexistent
cassandra:104:/var/lib/cassandra
postgres:105:/var/lib/postgresql
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
man:6:/var/cache/man
lp:7:/var/spool/lpd
mail:8:/var/mail
news:9:/var/spool/news
uucp:10:/var/spool/uucp
proxy:13:/bin
www-data:33:/var/www
backup:34:/var/backups
list:38:/var/list
irc:39:/var/run/ircd
gnats:41:/var/lib/gnats
nobody:65534:/nonexistent
_apt:100:/nonexistent
messagebus:101:/nonexistent
theia:1000:/home/theia
mongodb:102:/var/lib/mongodb
ntp:103:/nonexistent
cassandra:104:/var/lib/cassandra
postgres:105:/var/lib/postgresql
theia@theiadocker-craigtrupp8:/home/project$ ls
csv2db.sh  postgres
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
cut -d: -f 1,3,6 /etc/passwd
```

```
theia@theiadocker-craigtrupp8:/home/project$ ls
csv2db.sh  extracted-data.txt  postgres
theia@theiadocker-craigtrupp8:/home/project$ cat extracted-data.txt
root:0:/root
daemon:1:/usr/sbin
bin:2:/bin
sys:3:/dev
sync:4:/bin
games:5:/usr/games
man:6:/var/cache/man
lp:7:/var/spool/lpd
mail:8:/var/mail
news:9:/var/spool/news
uucp:10:/var/spool/uucp
proxy:13:/bin
www-data:33:/var/www
backup:34:/var/backups
list:38:/var/list
irc:39:/var/run/ircd
gnats:41:/var/lib/gnats
nobody:65534:/nonexistent
_apt:100:/nonexistent
messagebus:101:/nonexistent
theia:1000:/home/theia
mongodb:102:/var/lib/mongodb
ntp:103:/nonexistent
cassandra:104:/var/lib/cassandra
postgres:105:/var/lib/postgresql
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
cut -d: -f 1,3,6 /etc/passwd
Transforming Data
tr : ,
root,0,/root
daemon,1,/usr/sbin
bin,2,/bin
sys,3,/dev
sync,4,/bin
games,5,/usr/games
man,6,/var/cache/man
lp,7,/var/spool/lpd
mail,8,/var/mail
news,9,/var/spool/news
```

```
uucp,10,/var/spool/uucp
proxy,13,/bin
www-data,33,/var/www
backup,34,/var/backups
list,38,/var/list
irc,39,/var/run/ircd
gnats,41,/var/lib/gnats
nobody,65534,/nonexistent
_apt,100,/nonexistent
messagebus,101,/nonexistent
theia,1000,/home/theia
mongodb,102,/var/lib/mongodb
ntp,103,/nonexistent
cassandra,104,/var/lib/cassandra
postgres,105,/var/lib/postgresql
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
cut -d: -f 1,3,6 /etc/passwd
Transforming Data
theia@theiadocker-craigtrupp8:/home/project$ ls
csv2db.sh  extracted-data.txt  postgres  transformed-data.csv
theia@theiadocker-craigtrupp8:/home/project$ bash csv2db.sh
Extracting data
cut -d: -f 1,3,6 /etc/passwd
Transforming Data
Loading data
You are now connected to database "template1" as user "postgres".
COPY 25
theia@theiadocker-craigtrupp8:/home/project$ echo '\c template1; \\SELECT *
from users;' | psql --username=postgres --host=localhost
You are now connected to database "template1" as user "postgres".
  username  | userid |    homedirectory
------------+--------+---------------------
 root       |      0 | /root
 daemon     |      1 | /usr/sbin
 bin        |      2 | /bin
 sys        |      3 | /dev
 sync       |      4 | /bin
 games      |      5 | /usr/games
 man        |      6 | /var/cache/man
 lp         |      7 | /var/spool/lpd
 mail       |      8 | /var/mail
 news       |      9 | /var/spool/news
```

```
 uucp       |     10 | /var/spool/uucp
 proxy      |     13 | /bin
 www-data   |     33 | /var/www
 backup     |     34 | /var/backups
 list       |     38 | /var/list
 irc        |     39 | /var/run/ircd
 gnats      |     41 | /var/lib/gnats
 nobody     |  65534 | /nonexistent
 _apt       |    100 | /nonexistent
 messagebus |    101 | /nonexistent
 theia      |   1000 | /home/theia
 mongodb    |    102 | /var/lib/mongodb
 ntp        |    103 | /nonexistent
 cassandra  |    104 | /var/lib/cassandra
 postgres   |    105 | /var/lib/postgresql
(25 rows)

theia@theiadocker-craigtrupp8:/home/project$
```