

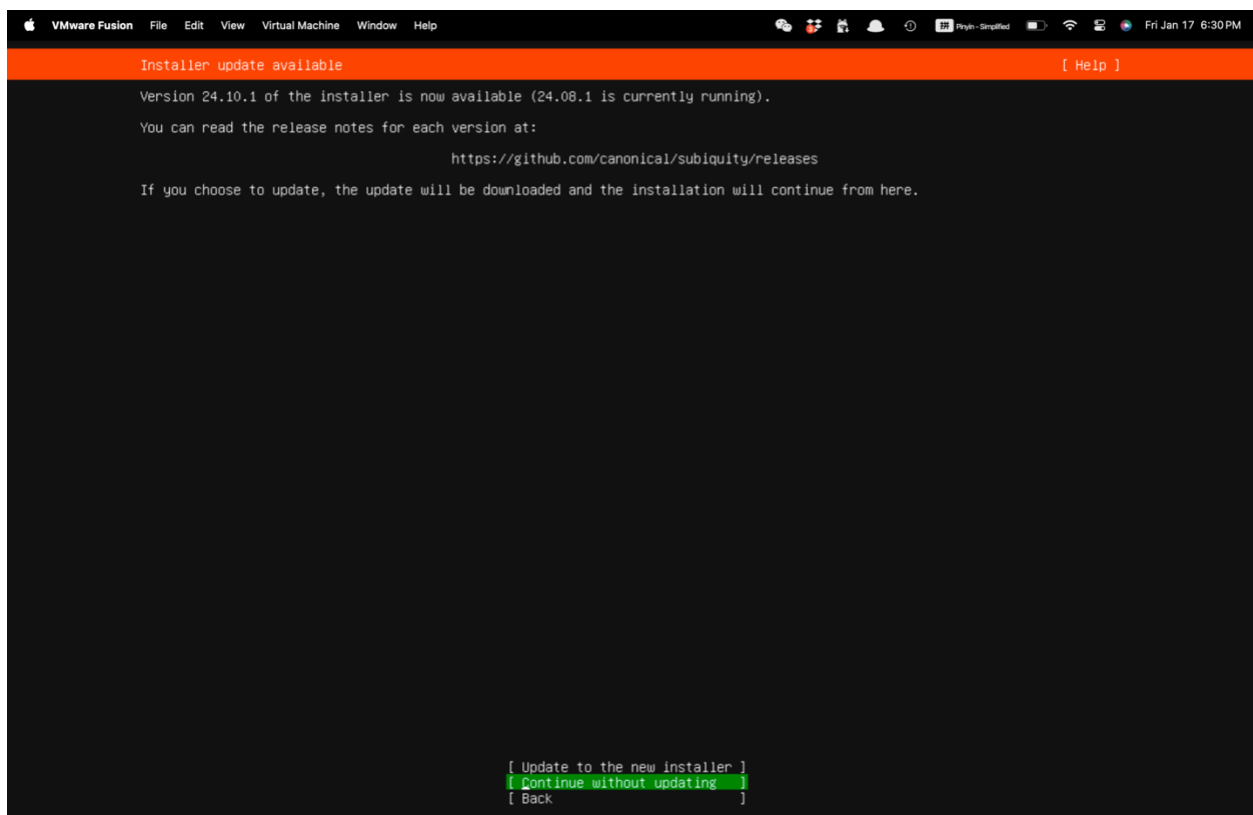
DSCI560 Lab1 Report

Name: Yuxuan Liu, USC ID: 4780355176

<https://github.com/paracosmgracee/usc-ds560-25sp/tree/main/lab1>

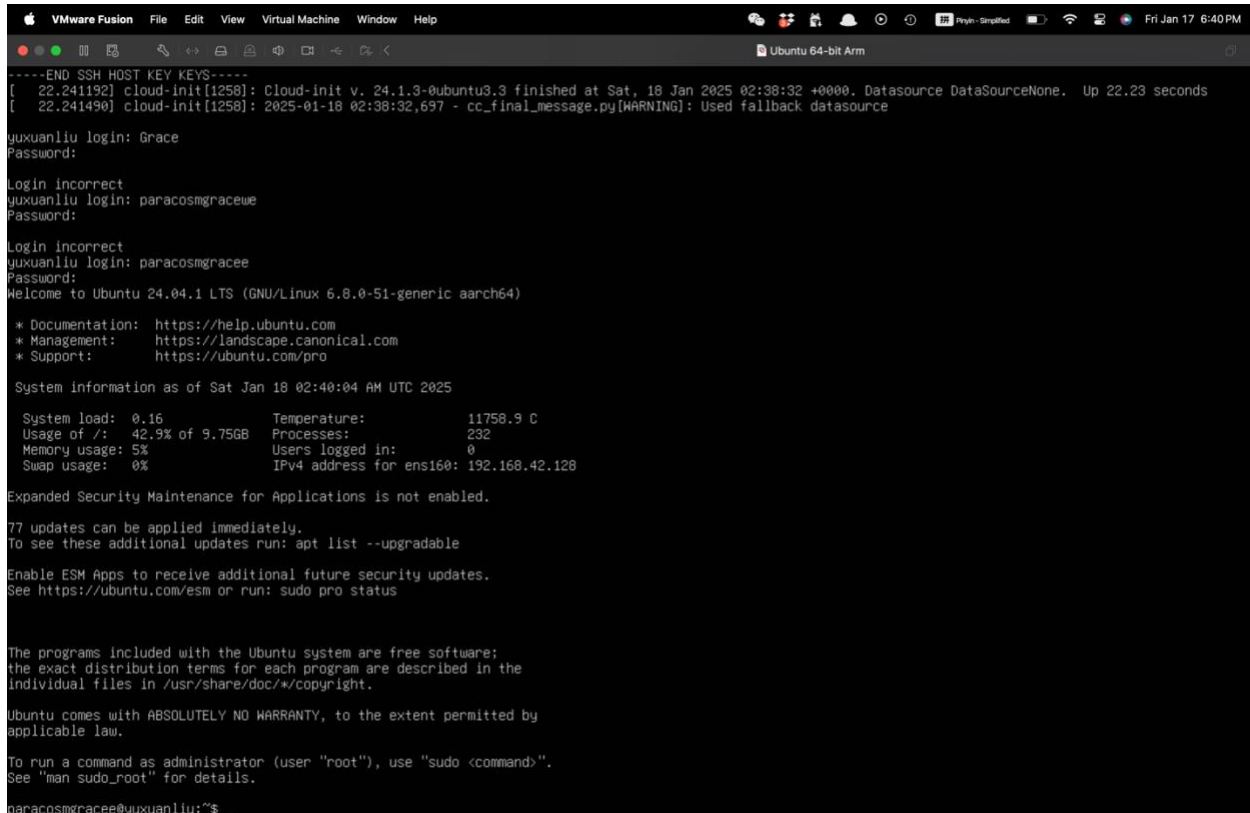
1. Installation and Setup

1.1. Install VMware



Shows the VMware installation screen, indicating the setup process for the virtual machine.

1.2. Download Ubuntu ISO Image



```
-----END SSH HOST KEY KEYS-----
[ 22.241192] cloud-init[1250]: Cloud-init v. 24.1.3-0ubuntu3.3 finished at Sat, 18 Jan 2025 02:38:32 +0000. Datasource DataSourceNone. Up 22.23 seconds
[ 22.241490] cloud-init[1250]: 2025-01-18 02:38:32,697 - cc_final_message.py[WARNING]: Used fallback datasource

yuxuanliu login: Grace
Password:

Login incorrect
yuxuanliu login: paracosmgracee
Password:

Login incorrect
yuxuanliu login: paracosmgracee
Password:
Welcome to Ubuntu 24.04.1 LTS (GNU/Linux 6.8.0-51-generic aarch64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/pro

System information as of Sat Jan 18 02:40:04 AM UTC 2025

System load:  0.16           Temperature:   11758.9 C
Usage of /:   42.9% of 9.75GB Processes:    232
Memory usage: 5%           Users logged in: 0
Swap usage:   0%           IPv4 address for ens160: 192.168.42.128

Expanded Security Maintenance for Applications is not enabled.

77 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

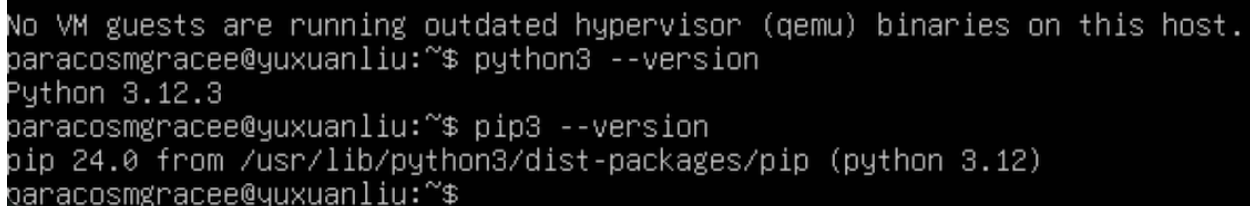
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

paracosmgracee@yuxuanliu:~$
```

Displays the Ubuntu ISO image download page, highlighting the steps to acquire the necessary operating system.

1.3. Install Python on Linux



```
No VM guests are running outdated hypervisor (qemu) binaries on this host.
paracosmgracee@yuxuanliu:~$ python3 --version
Python 3.12.3
paracosmgracee@yuxuanliu:~$ pip3 --version
pip 24.0 from /usr/lib/python3/dist-packages/pip (python 3.12)
paracosmgracee@yuxuanliu:~$
```

Shows the Linux terminal with commands to install Python 3 and verify the version.

2. Get Familiar with Linux and Python

2.1. Playing around with Linux Terminal

```
paracosmgracee@yuxuanliu:~$ mkdir YuxuanLiu_4780355176
paracosmgracee@yuxuanliu:~$ cd YuxuanLiu_4780355176
paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176$ mkdir data scripts
paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176$ cd scripts
paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176/scripts$ touch task_1.py
paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176/scripts$ cd ..
paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176$ cd ..
paracosmgracee@yuxuanliu:~$ ls -R
.:
YuxuanLiu_4780355176

./YuxuanLiu_4780355176:
data  scripts

./YuxuanLiu_4780355176/data:

./YuxuanLiu_4780355176/scripts:
task_1.py
paracosmgracee@yuxuanliu:~$
```

Depicts the Linux terminal, where the student has created a directory, subdirectories, and a Python file.

2.2. A basic Python Script

```
GNU nano 7.2
name = input("Enter your name: ")
print(f"Hello, {name}!")
```

Shows the Python script that prompts the user for their name and outputs a greeting.

```
paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176$ cd scripts
paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176/scripts$ python3 task_1.py
Enter your name: Yuxuan Liu
Hello, Yuxuan Liu!
paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176/scripts$
```

Displays the terminal output of the Python script running.

2.3. Python Web-scraping Task

```
(venv) paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176/scripts$ python3 web_scraper.py
Opening page: https://www.cnbc.com/world/?region=world
HTML data saved to: /home/paracosmgracee/YuxuanLiu_4780355176/data/raw_data/web_data.html
Printing first 10 lines of web_data.html:
<html itemscope="" itemtype="https://schema.org/WebPage" lang="en" prefix="og:https://ogp.me/ns#">
<head>
  <script src="https://nbcu.track.securedvisit.com/js/sv.js?sv_cid=5998_04719&sv_origin=cnbc.com" type="text/javascript">
  </script>
  <script async="" src="https://static.criteo.net/js/ld/publishertag.prebid.144.js" type="text/javascript">
  </script>
  <script async="" src="https://jssdk.cdn.mparticle.com/js/v2/us1-99f689156b80a84f8e3383020acfea25/mparticle.js" type="text/javascript">
  </script>
  <script async="" src="https://www.googletagmanager.com/gtag/js?id=AW-304008397&l=dataLayer&cx=c&gtm=45be51g0za200" type="text/javascript">
  </script>
```

Indicates the Python script being written to perform web scraping from the CNBC website.

2.4. Data Filtering Task

```
(venv) paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176/scripts$ python3 data_filter.py
Reading web_data.html...
Filtering market data fields...
Found 5 market items.
Successfully created /home/paracosmgracee/YuxuanLiu_4780355176/data/processed_data/market_data.csv
Filtering latest news fields...
Found 6 news items.
Successfully created /home/paracosmgracee/YuxuanLiu_4780355176/data/processed_data/news_data.csv
Data filtering complete!
```

```
(venv) paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176/scripts$ cat ../data/processed_data/market_data.csv
Symbol,StockPosition,ChangePct
RSX 200*, "8,408.20", +0.73%
NIKKEI*, "38,951.77", +0.13%
NIFTY 50*, "23,344.75", UNCH
HSI*, "20,201.09", +1.38%
SHANGHAI*, "3,249.54", +0.16%
(venv) paracosmgracee@yuxuanliu:~/YuxuanLiu_4780355176/scripts$ cat ../data/processed_data/news_data.csv
Timestamp,Title,Link
16 Min Ago,Here are the 4 things we're watching in the stock market in the week ahead,https://www.cnbc.com/2025/01/20/here-are-the-4-things-were-watching-in-the-stock-market-in-the-week-ahead.html
3 Hours Ago,Trump says he's thinking of imposing 25% tariffs on Canada and Mexico in February,https://www.cnbc.com/2025/01/20/trump-says-hes-thinking-of-imposing-25percent-tariffs-on-canada-and-mexico-in-february.html
3 Hours Ago,What to look out for as Trump 2.0 era starts,https://www.cnbc.com/2025/01/21/cnbc-daily-open-what-to-look-out-for-as-trump-2point0-era-starts.html
4 Hours Ago,Jefferies downgrades Apple to rare underperform rating,https://www.cnbc.com/2025/01/20/jefferies-downgrades-apple-to-rare-underperform-rating-says-r-evenue-may-disappoint.html
4 Hours Ago,Asia-Pacific markets trade mixed as investors await clarity on President Donald Trump's policies,https://www.cnbc.com/2025/01/21/asia-markets-live-updates.html
5 Hours Ago,S&P 500 futures are flat as Trump says he's considering tariffs on Canada and Mexico,https://www.cnbc.com/2025/01/20/stock-futures-rise-as-trumps-second-term-begins-live-updates.html
```

Shows the terminal output (market_data.csv and news_data.csv) when running the data filtering script.