Apache Flink Getting Started

Experiment 0: Setup

1.1 Steps to setup your Experiment Account

1.1.1 Use Launch an Ubuntu 22 environment in t3.large with 80GB hard drive. Turn off auto hibernation, but remember that you don't want to leave this run for days or weeks without doing something with it, be frugal. Ensure that you generate a new PEM file for accessing the environment during your launch.

1.1.2 Grab your SSH identity pem file created in the instance launch, IP address, user (ubuntu), and pem file.

1.1.3 Ensure the pem file is in the same folder that you do the ssh from following and that you've changed the permissions for the file

In Windows

attrib -x ansible.pem

In Mac

chmod 600 ansible.pem

1.1.4 From a command prompt on your jumpbox machine SSH to the Unbuntu server
ssh -o ServerAliveInterval=180 -o ServerAliveCountMax=2 -i ansible.pem ubuntu@ip_address_of VM

1.1.5 Change the Server RDP user password, (make sure to keep this for access to the Flink WebUI through Remote Desktop Protocol)

sudo passwd xrdp

1.1.6 Run the following commands

```
sudo apt update
sudo apt install ubuntu-desktop
sudo apt install xrdp
sudo systemctl status xrdp
sudo adduser xrdp ssl-cert
sudo ufw allow 3389
wget -q -O
      - https://dlssl.google.com/linux/linux_signing_key.pub
      | gpg --dearmor | sudo tee
/etc/apt/trusted.gpg.d/chrome.gpg
sudo apt update
sudo apt install google-chrome-stable
```

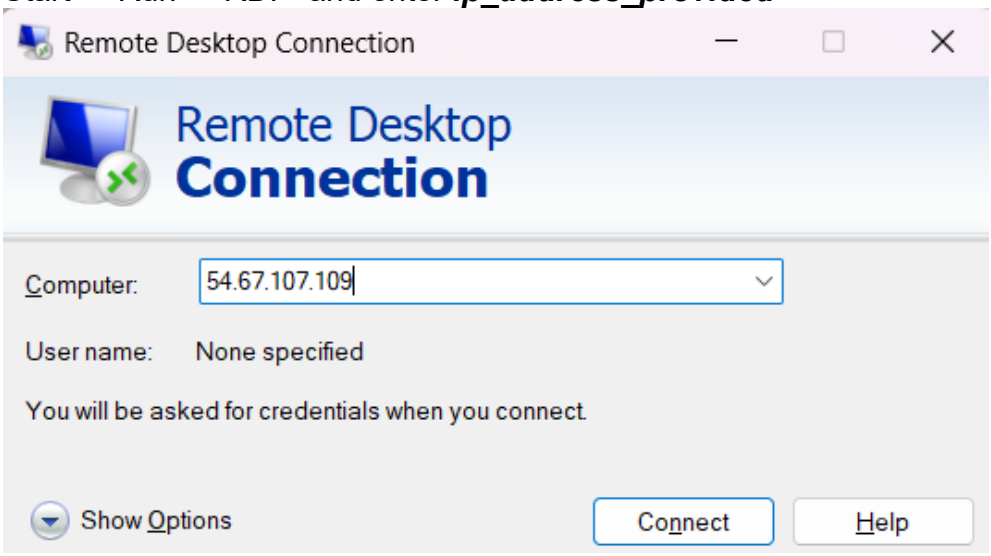
1.1.7 Retrieve the Flink download we need with
wget https://dlcdn.apache.org/flink/flink-2.0.0/flink-2.0.0-bin-scala_2.12.tgz

1.1.8 Unpack that in your ubuntu user home folder
cd ~
tar -xzf flink-2.0.0-bin-scala_2.12.tgz

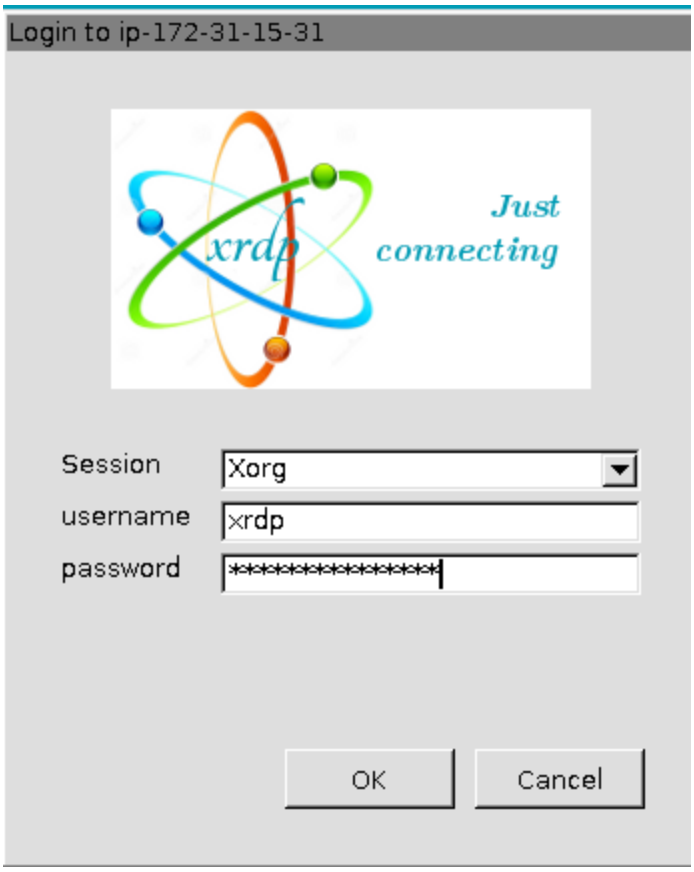
1.1.9 Download the GitHub repo on the dev server
cd ~
git clone https://github.com/GeorgeNiece/flink-data-processing-2day

1.1.10 Change to the flink folder, verify Flink isn't started, start the Flink dev cluster, and verify that it started
ps -ef | grep flink
cd ~/flink-2.0.0
./bin/start-cluster.sh
ps -ef | grep flink

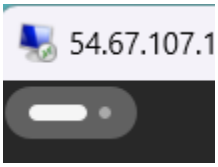
1.1.11 From the Windows Jump Host, RDP to the dev server
Start -> Run -> RDP and enter ***ip_address_provided***



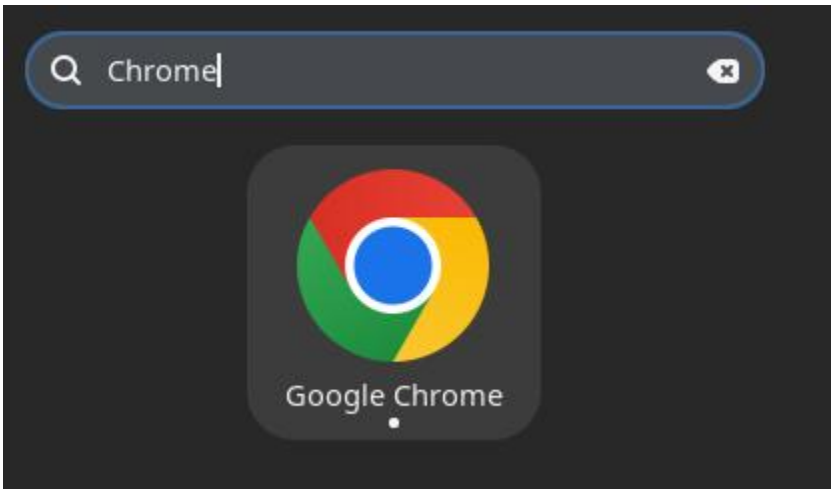
1.1.12 Login to the ubuntu dev sandbox using Windows RDP with the xrdp user and the password that you set in Step 1.14



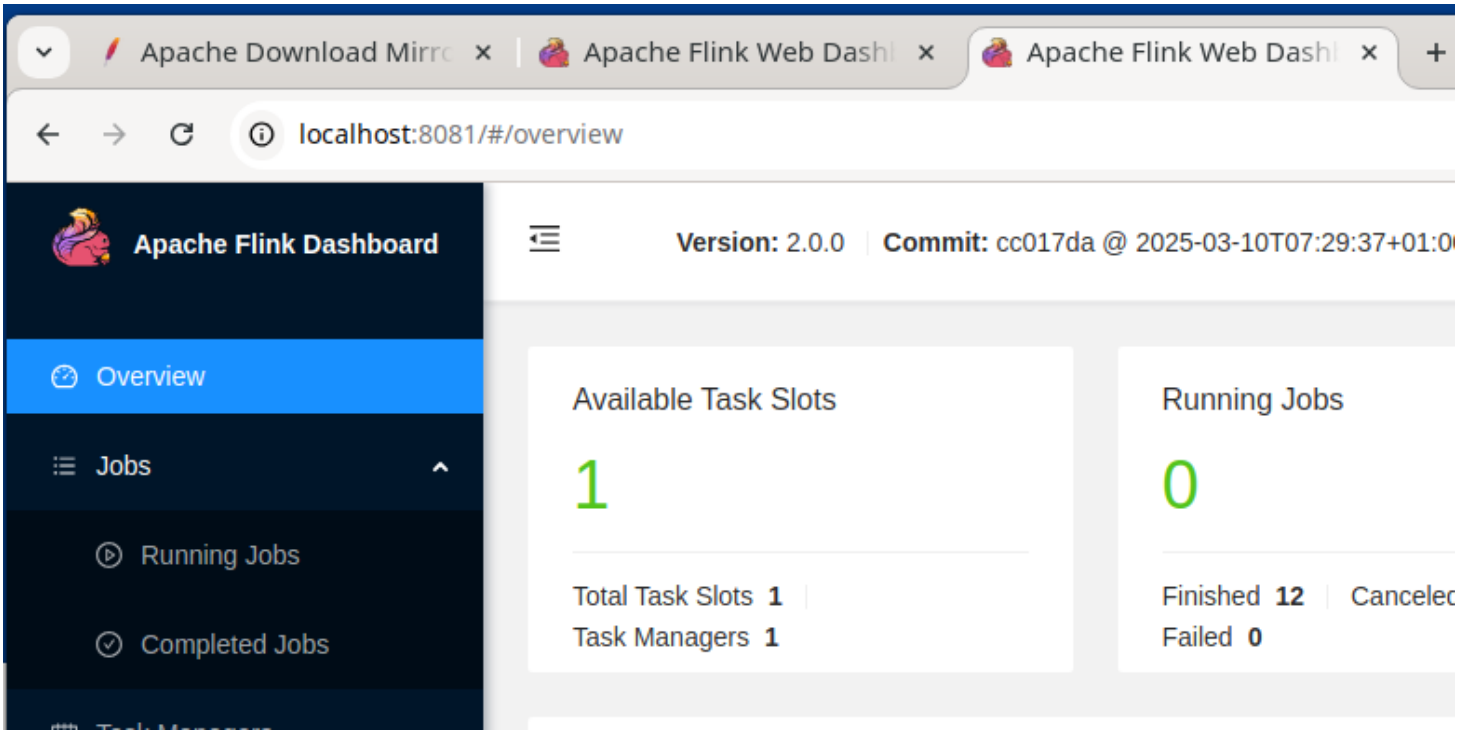
1.1.13 Click the Activities button in the top left corner of the Ubuntu Desktop



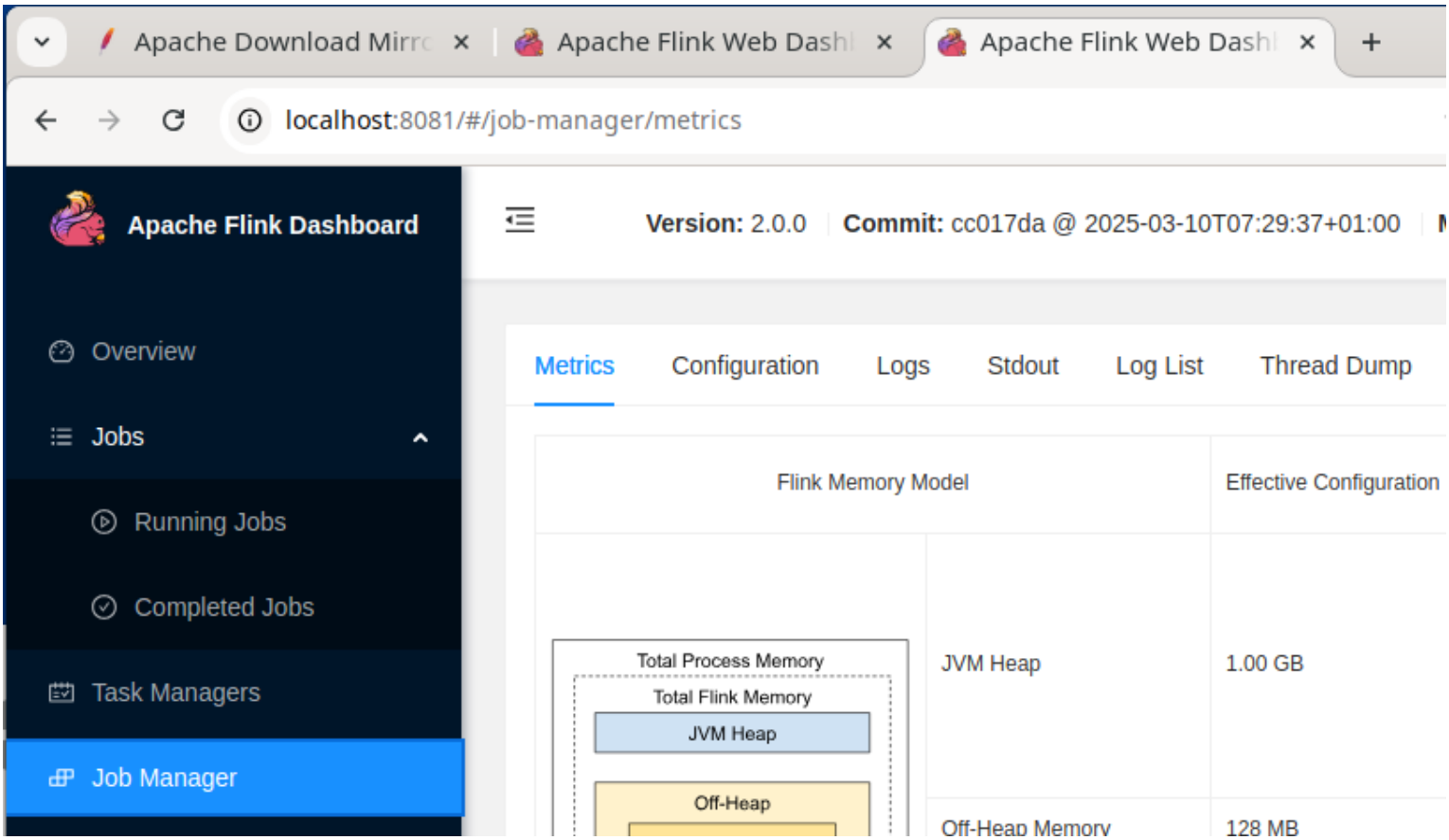
1.1.14 Wait for the Search Box at the top of the Ubuntu Desktop, and enter Chrome, click on the Launch Logo



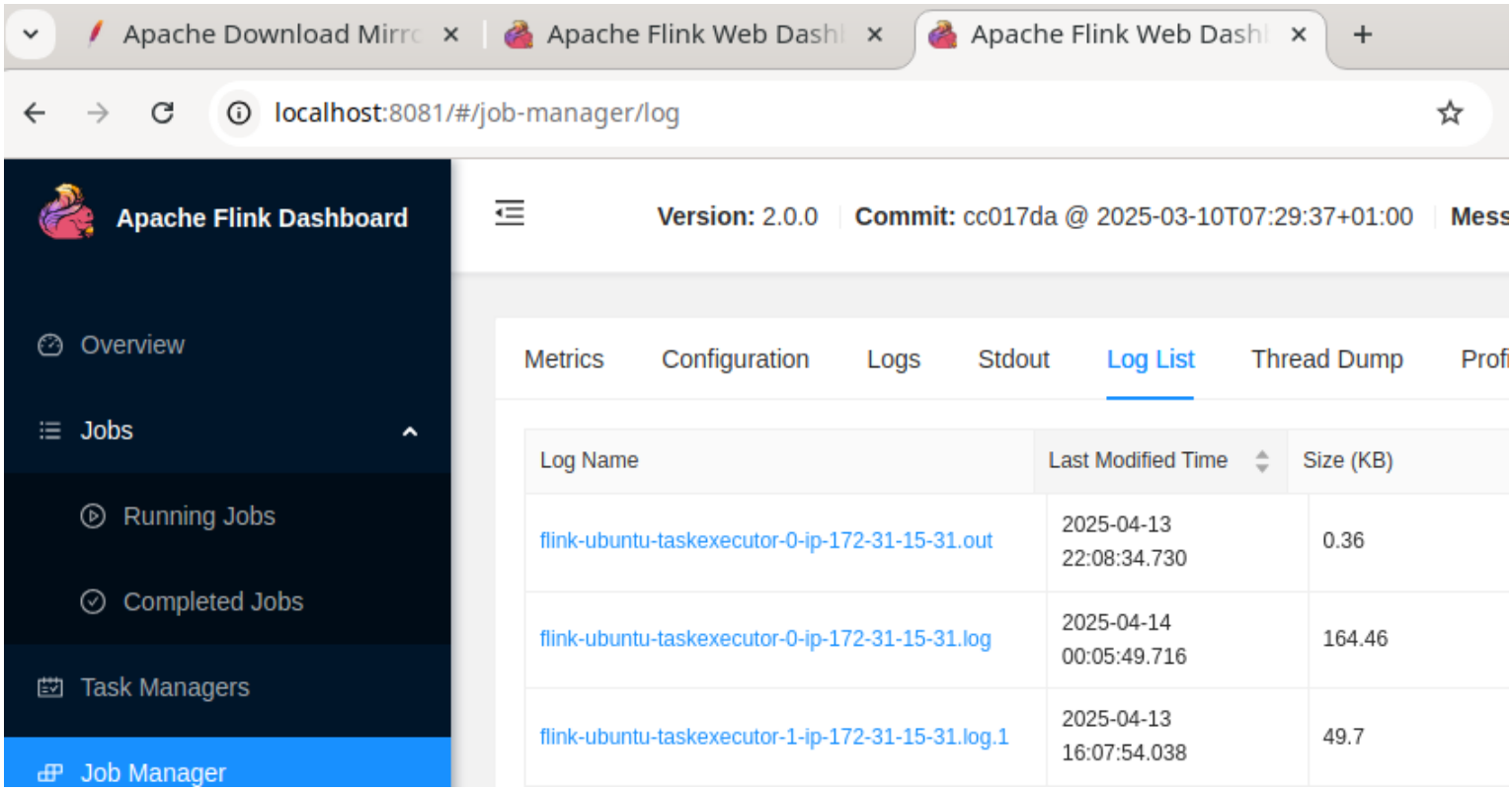
1.1.15 Load the Flink Web UI at localhost:8081



1.1.16 Click the Job Manager in the left hand navigation



1.1.17 Select Log List in the page navigation. These are the logs we'll monitor while we're running some of our experiments



1.1.18 The first flink-ubuntu-taskexecutor*.out file will be the one we spend the most time looking at.

1.1.19 Navigate back to the SSH terminal and open the Flink configuration

cd ~/flink-2.0.0
./vi conf/config.yaml

1.1.20 Review the file (if you're not VI proficient, use "**nano conf/config.yaml**")

1.1.21 **Congratulations, time to celebrate** finishing the setup of a Flink Developer Sandbox