Apache Flink Getting Started

Experiment 0: Setup

1.1 Steps to setup your Experiment Account

1.1.1 Use the provided information to connect to your Windows Jump box environment. https://jruels.github.io/automation-dev/VM_access.html

1.1.2 Grab your flink sandbox credentials to connect, IP address, user (ubuntu), and pem file. This is in the GitHub Repo provided for this class (https://github.com/GeorgeNiece/flink-data-processing-2day), and the FlinkDev-2025-04-14 excel file shared by the Flink course leader.

1.1.3 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_provided*

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

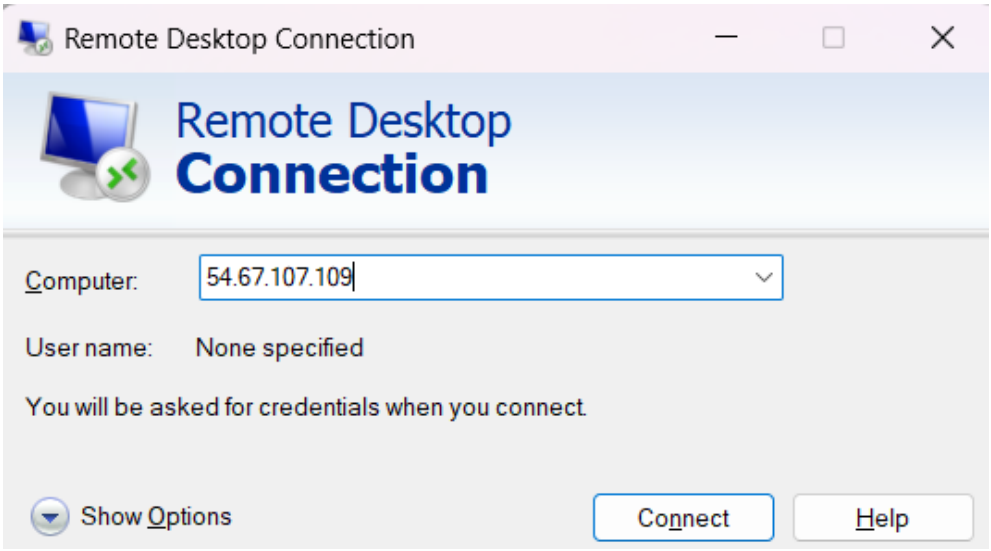
1.1.5 Retrieve the Flink download we need with
wget https://dlcdn.apache.org/flink/flink-2.0.0/flink-2.0.0-bin-scala_2.12.tgz

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

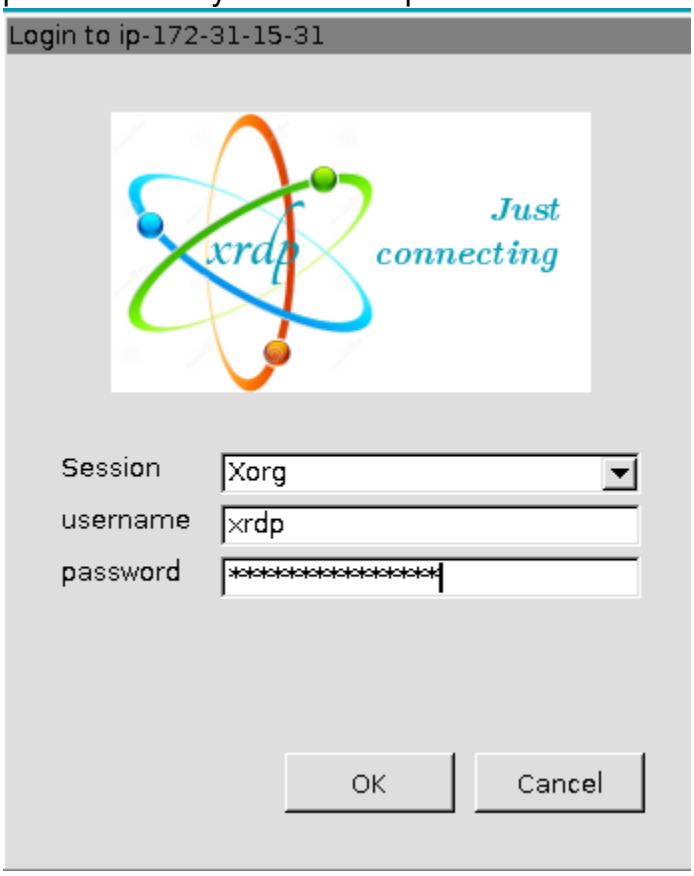
1.1.7 Download the GitHub repo on the dev server
cd ~
git clone https://github.com/GeorgeNiece/flink-data-processing-2day

1.1.8 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

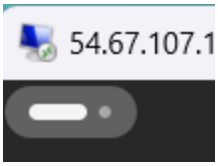
1.1.9 From the Windows Jump Host, RDP to the dev server
Start -> Run -> RDP and enter ***ip_address_provided***



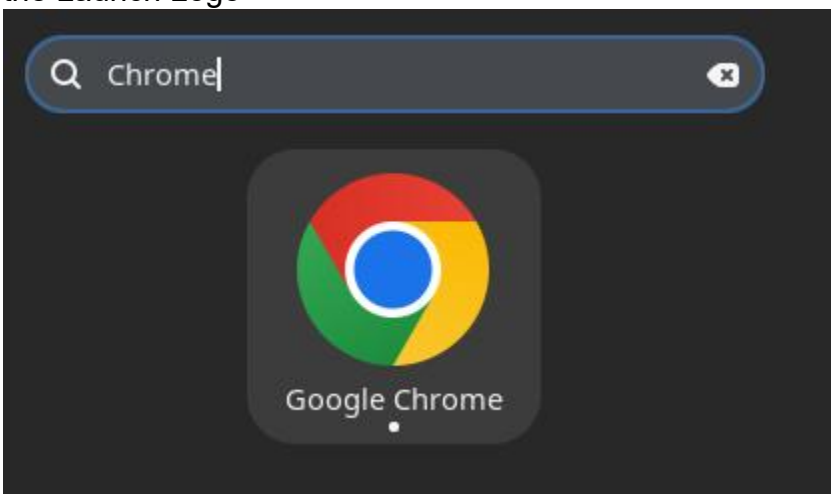
1.1.10 Login to the ubuntu dev sandbox using Windows RDP with the xrdp user and the password that you set in Step 1.14



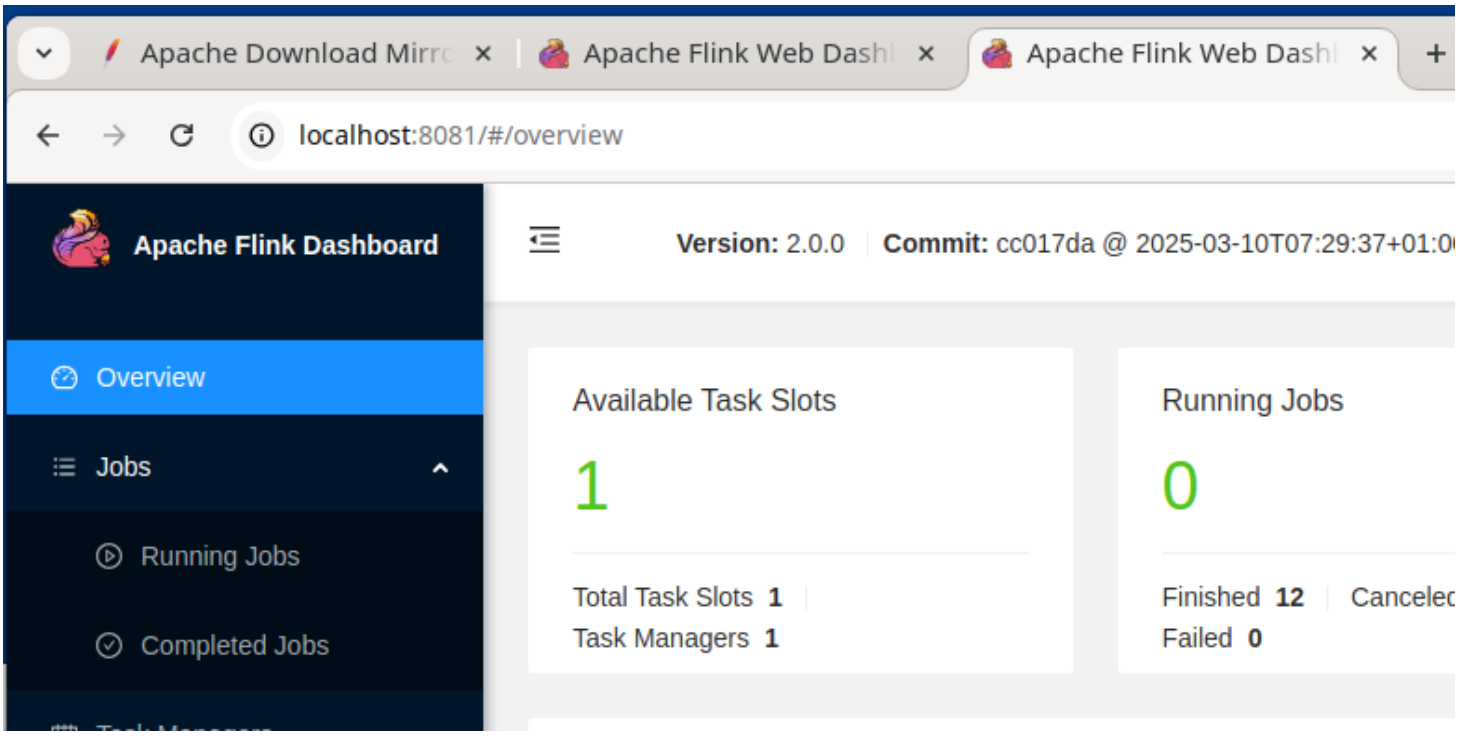
1.1.11 Click the Activities button in the top left corner of the Ubuntu Desktop



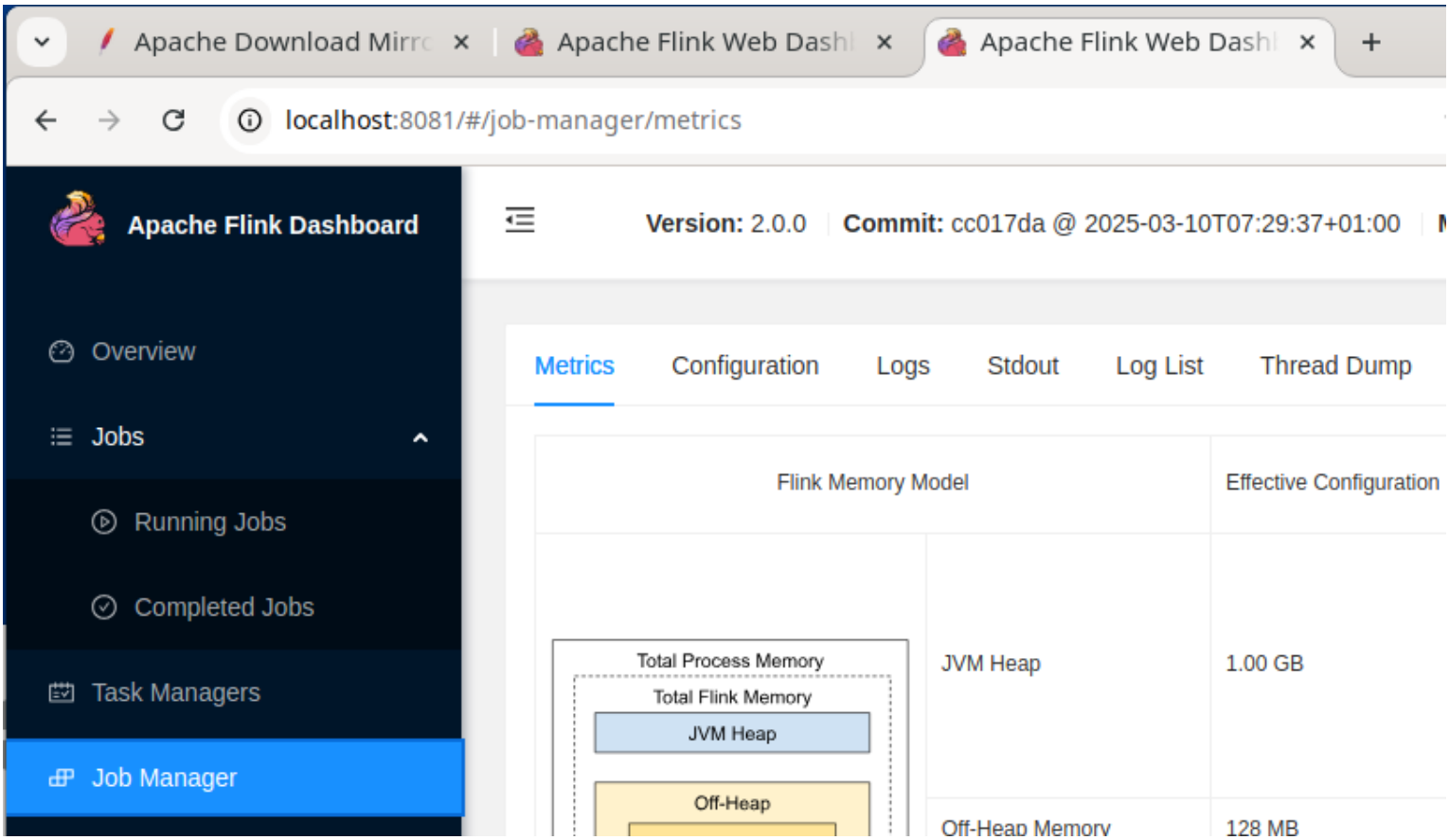
1.1.12 Wait for the Search Box at the top of the Ubuntu Desktop, and enter Chrome, click on the Launch Logo



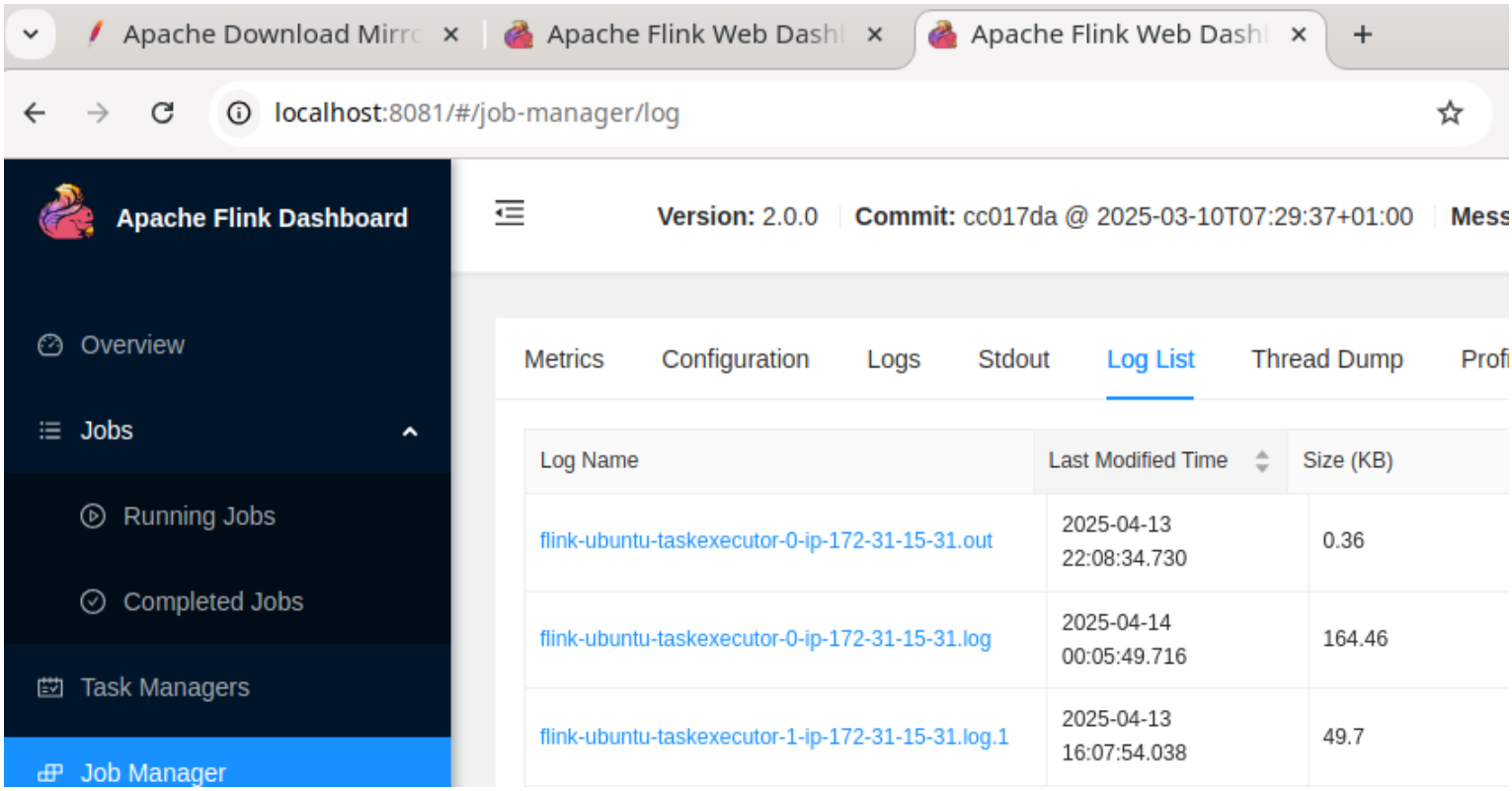
1.1.13 Load the Flink Web UI at localhost:8081



1.1.14 Click the Job Manager in the left hand navigation



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



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

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

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

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

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