



Installing Flink and Running:

1. Install PyFlink

```
python -m pip install apache-flink
```

2. Download Apache Flink binary: [here](#) (download the **binaries** and not the source code)

3. Unzip the downloaded file:

```
tar -zxvf flink-*.tgz
```

The unzipped directory should have a bin and conf folder. This is your Flink Home directory.

4. As we are running on a local system, we need to change some default configs, explained [here](#) — you need to:

Update Flink Home/bin/flink:

```
log_setting=(-Dlog.file="$log" -Dlog4j.configuration="path\\to\\flink_home\\conf\\log4j-cli.properties" -Dlog4j.configurationFile="path\\to\\flink_home\\conf\\log4j-cli.properties" -Dlogback.configurationFile="path\\to\\flink_home\\conf\\logback.xml")
```

Remove INTERNAL_HADOOP_CLASSPATHS from exec. It should look like this after update”

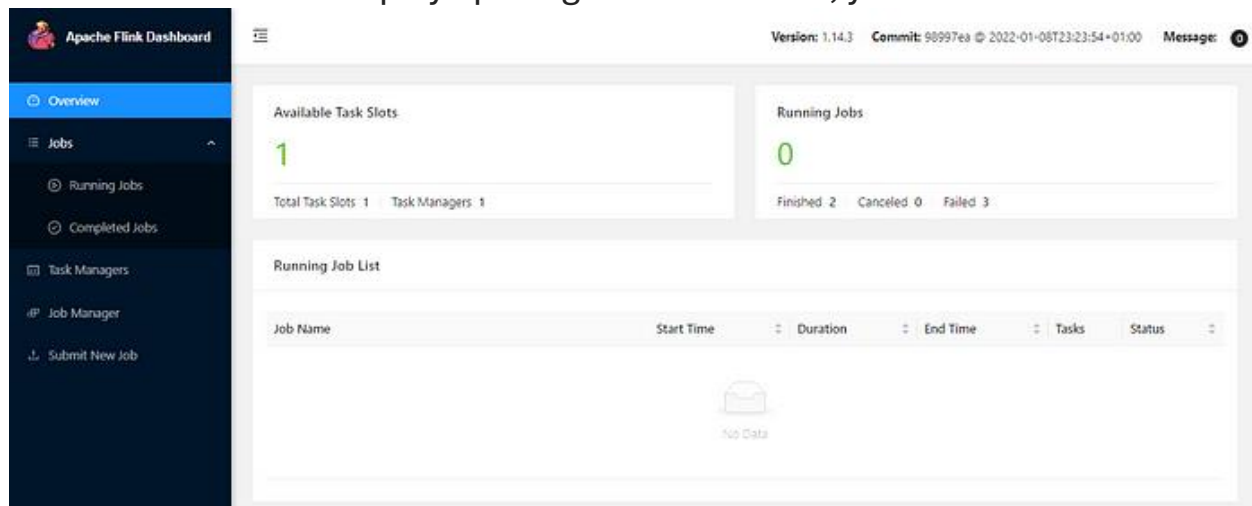
```
exec $JAVA_RUN $JVM_ARGS $FLINK_ENV_JAVA_OPTS "${log_setting[@]}" -classpath "`manglePathList "$CC_CLASSPATH" ` " org.apache.flink.client.cli.CliFrontend "$@"
```

Update Flink Home/bin/flink-daemon.sh and make the same changes as above.

5. Start the Flink Cluster

```
./bin/start-cluster.sh
```

Check if the cluster is up by opening localhost:8081, you should see:



Flink Dashboard at Startup

6. Run where python (Windows) / which python (Linux/ Mac) to get the path to your python venv which has apache-flink installed. Submit the stream with:

```
./flink-1.14.3/bin/flink run -py process_movies.py -pyclientexec "path_to\python.exe" -pyexec "path_to\python.exe" --output "path_to\out.txt" --input "path_to\input.txt"
```

7. Check out.txt and localhost:8081 for Success!