



# Apache Flink Installation and Running Applications

## Apache Flink Installation

1. Install PyFlink using the below command

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

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

3. Unzip the downloaded file using the following tar utility or 7zip/winrar

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

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

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

Navigate to **Flink Home/bin/** directory and open the file named flink in a notepad or any text editor and search for the below **log\_setting** and update the configuration path according to your flink home path.

```
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 below after the update.

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

Repeat the above step 4 to update the flink-daemon.sh file which can be found in the same directory <flinkhome>/bin and make the same changes as above.

## Start/Run the Apache Flink Cluster

Now start the flink cluster by executing the **start-cluster.sh** file which can be found inside the **bin** directory. Note: Run it from the Bash in case of windows using either **Git-Bash** or **Cygwin** (if you don't have **Cygwin** installed in your machine then download it from the following location <https://cygwin.com/install.html>)

If you are using the Windows git shell, Cygwin can produce a failure similar to this one:

```
c:/flink/bin/start-cluster.sh: line 30: $'\r': command not found
```

This error occurs because git is automatically transforming UNIX line endings to Windows style line endings when running in Windows. The problem is that Cygwin can only deal with UNIX style line endings. The solution is to adjust the Cygwin settings to deal with the correct line endings by following these three steps:

### Start a Cygwin shell.

Determine your home directory by entering

```
cd; pwd
```

This will return a path under the Cygwin root path.

Using Notepad or a different text editor open the file **.bash\_profile** in the home directory and append the following: (If the file does not exist you will have to create it)

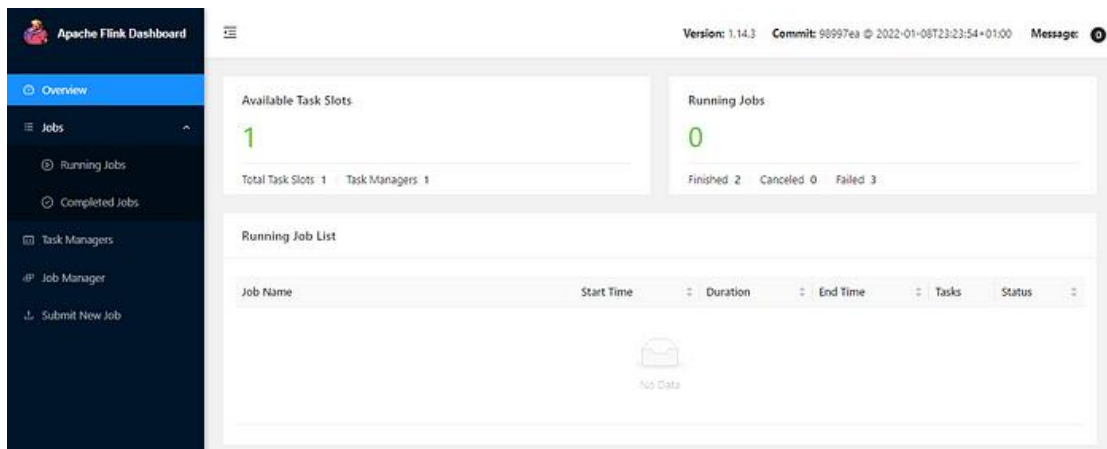
```
export SHELLOPTS
```

```
set -o igncr
```

Save the file and open a new Cygwin bash shell and navigate to the Flink home folder and then execute the below command.

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

Check if the cluster is up and running by opening a browser and type the following url in the browser address bar <http://localhost:8081>, after press go you should be able to see the flink dashboard like shown below in the screenshot.



**Figure 1:** Flink Dashboard

Run where python (Windows) / python (Linux/ Mac) to get the path to your python venv environment which has apache-flink installed and submit the stream to run the Apache Flink job.

```
.<drive><flink-home-folder>/bin>flink run -py <filename.py> -pyclientexec "path_to\python.exe" -  
pyexec "path_to\python.exe" --input "path_to\input.txt" --output "path_to\output.txt"
```

Now finally check the output.txt and the <http://localhost:8081> flink dashboard to verify everything is looking good. This concludes the Apache Flink installation and application execution.