

Multi-Node Cluster Installation

Partek® Flow® is a genomics data analysis and visualization software product designed to run on compute clusters. The following instructions assume the most basic setup of Partek Flow and must only be attempted by system administrators who are familiar with Linux-based commands. These instructions are not intended to be comprehensive. Cluster environments are largely variable, thus there are no 'one size fits all' instructions. The installation procedure on a computer cluster is highly dependent on the type of computer cluster and the environment it is located. We can support a large array of Linux distributions and configurations. In all cases, Partek Technical Support will be available to assist with cluster installation and maintenance to ensure compatibility with any cluster environment. Please consult with Partek Licensing Support (licensing@partek.com) for additional information.

Prior to installation, make sure you have the license key related to the host-ID of the compute cluster the software will be installed in. Contact licensing@partek.com for key generation.

- [Installation on a Computer Cluster](#)
- [Integration with your queueing system](#)
- [Bringing up workers](#)
- [Shutting down workers](#)
- [Updating Partek Flow](#)

Installation on a Computer Cluster

Make a standard linux user account that will run the Partek Flow server and all associated processes. It is assumed this account is synced between the cluster head node and all compute nodes. For this guide, we name the account *flow*

1. Log into the flow account and proceed to the cd to the flow home directory

```
cd /home/flow
```

2. Download Partek Flow and the remote worker package

```
wget --content-disposition http://packages.partek.com/linux/flow-release  
wget --content-disposition http://packages.partek.com/linux/flow-worker-release
```

3. Unzip these files into the flow home directory */home/flow*. This yields two directories: *partek_flow* and *PartekFlowRemoteWorker*
4. Partek Flow can generate large amounts of data, so it needs to be configured to the bulk of this data in the largest shared data store available. For this guide we assume that the directory is located at */shared*. Adjust this path accordingly.
5. It is required that the Partek Flow server (which is running on the head node) and remote workers (which is running on the compute nodes) see identical file system paths for any directory Partek Flow has read or write access to. Thus */shared* and */home/flow* must be mounted on the Flow server and all compute nodes. Create the directory */shared/FlowData* and allow the *flow* linux account write access to it
6. It is assumed the head node is attached to at least two separate networks: (1) a public network that allows users to log in to the head node and (2) a private backend network that is used for communication between compute nodes and the head node. Clients connect to the Flow web server on port *8080* so adjust the firewall to allow inbound connections to *8080* over the public network of the head node. Partek Flow will connect to remote workers over your private network on port *2552* and *8443*, so make sure those ports are open to the private network on the flow server and workers.
7. Partek Flow needs to be informed of what private network to use for communication between the server and workers. It is possible that there are several private networks available (gigabit, infiniband, etc.) so select one to use. We recommend using the fastest network available. For this guide, let's assume that private network is *10.1.0.0/16*. Locate the headnode hostname that resolves to an address on the *10.1.0.0/16* network. This must resolve to the same address on all compute nodes.

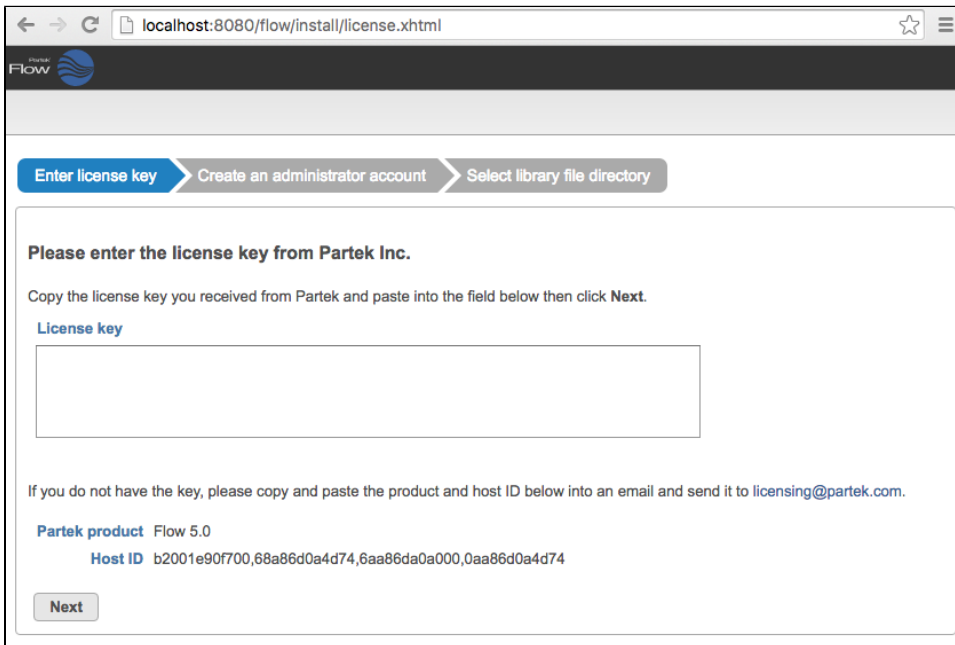
8. For example:
host head-node.local
yields
10.1.1.200

Open */home/flow/.bashrc* and add this as the last line:
export CATALINA_OPTS="\$CATALINA_OPTS -Djava.awt.headless=true
-DflowDispatcher.flow.command.hostname=head-node.local
-DflowDispatcher.akka.remote.netty.tcp.hostname=head-node.local"

Source *.bashrc* so the environment variable *CATALINA_OPTS* is accessible.

NOTE: If workers are unable to connect (below), then replace all hostnames with their respective IPs.

9. Start Partek Flow
`~/partek_flow/start_flow.sh`
10. You can monitor progress by tailing the log file `partek_flow/logs/catalina.out`. After a few minutes, the server should be up.
11. Make sure the correct ports are bound
`netstat -tulpn`
12. You should see `10.1.1.200:2552` and `:::8080` as *LISTENing*. Inspect `catalina.out` for additional error messages.
13. Open a browser and go to <http://localhost:8080> on the head node to configure the Partek Flow server.
14. Enter the license key provided (Figure 1)

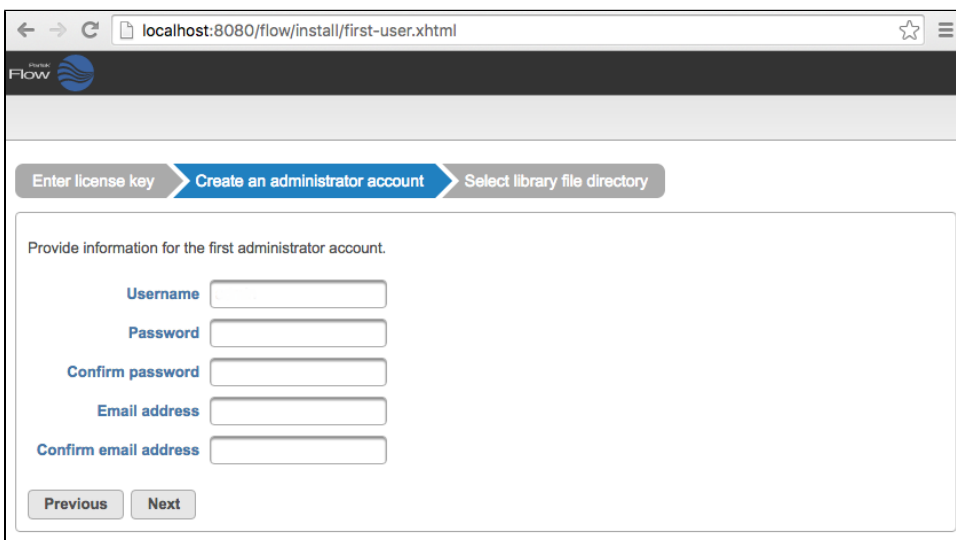


The screenshot shows a web browser window with the address bar displaying `localhost:8080/flow/install/license.xhtml`. The page has a dark header with the Partek Flow logo. Below the header, there are three navigation buttons: "Enter license key" (highlighted in blue), "Create an administrator account", and "Select library file directory". The main content area is titled "Please enter the license key from Partek Inc." and includes instructions to copy the license key from Partek and paste it into a text field. Below the text field, there is a link "License key". Further down, there is a message: "If you do not have the key, please copy and paste the product and host ID below into an email and send it to licensing@partek.com." Below this message, the "Partek product" is listed as "Flow 5.0" and the "Host ID" is listed as "b2001e90f700,68a86d0a4d74,6aa86da0a000,0aa86d0a4d74". At the bottom of the form, there is a "Next" button.

Figure 4. Setting up the Partek Flow license during installation

15. If there appears to be an issue with the license or there is a message about *'no workers attached'*, then restart Partek Flow. It may take 30 sec for the process to shut down. Make sure the process is terminated before starting the server back up.:
`~/partek_flow/stop_flow.sh`
Then run:
`~/partek_flow/start_flow.sh`

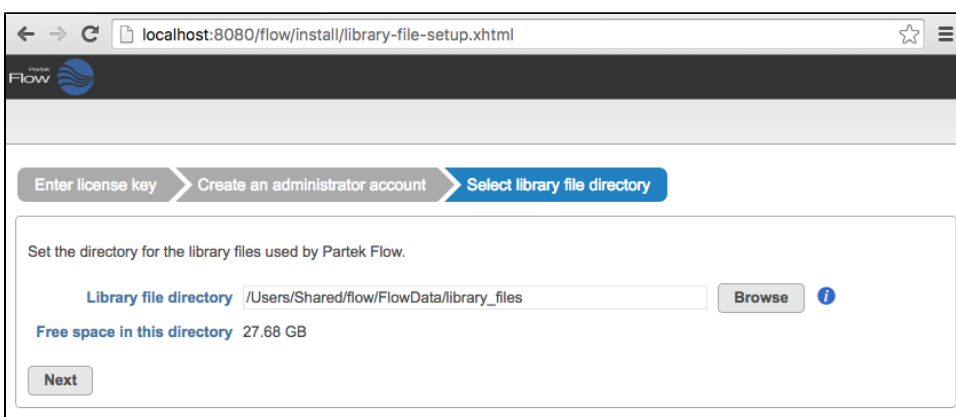
16. You will now be prompted to setup the Partek Flow *admin* user (Figure 2). Specify the username (admin), password and email address for the administrator account and click **Next**



The screenshot shows a web browser window at localhost:8080/flow/install/first-user.xhtml. The interface has a dark header with the Partek Flow logo. Below the header, there are three steps in a progress bar: 'Enter license key', 'Create an administrator account' (which is highlighted), and 'Select library file directory'. The main content area is titled 'Provide information for the first administrator account.' and contains five input fields: 'Username', 'Password', 'Confirm password', 'Email address', and 'Confirm email address'. At the bottom of the form are two buttons: 'Previous' and 'Next'.

Figure 5. Setting up the Partek Flow 'admin' account during installation

17. Select a directory folder to store the library files that will be downloaded or generated by Partek Flow (Figure 3). All Partek Flow users share library files and the size of the library folder can grow significantly. We recommend at least 100GB of free space should be allocated for library files. The free space in the selected library file directory is shown. Click **Next** to proceed. You can change this directory after installation by changing system preferences. For more information, see [Library file management](#).



The screenshot shows a web browser window at localhost:8080/flow/install/library-file-setup.xhtml. The interface has a dark header with the Partek Flow logo. Below the header, there are three steps in a progress bar: 'Enter license key', 'Create an administrator account', and 'Select library file directory' (which is highlighted). The main content area is titled 'Set the directory for the library files used by Partek Flow.' and contains a text input field for 'Library file directory' with the value '/Users/Shared/flow/FlowData/library_files'. To the right of the input field is a 'Browse' button and an information icon. Below the input field, it says 'Free space in this directory 27.68 GB'. At the bottom of the form is a 'Next' button.

Figure 6. Selecting the library file directory

18. To set up the Partek Flow data paths, click on **Settings** located on the top-right of the Flow server webpage. On the left, click on **Directory permissions** then **Permit access to a new directory**. Add `/shared/PartekFlow` and allow all users access.
19. Next click on **System preferences** on the left menu and change **data download directory** and **default project output directory** to `/shared/PartekFlow/downloads` and `/shared/PartekFlow/project_output` respectively
- Note: If you do not see the /shared folder listed, click on the **Refresh folder list** link that is toward the bottom of the download directory dialog*
20. Since you do not want to run any work on the head node, go to **Settings>System preferences>Task queue and job processing** and uncheck **Start internal worker at Partek Flow server startup**.
21. Restart the Flow server:
~/`partek_flow/stop_flow.sh`
After 30 seconds, run:
~/`partek_flow/start_flow.sh`
This is needed to disable the internal worker.

22. Test that remote workers can connect to the Flow server
23. Log in as the *flow* user to one of your compute nodes. Assume the hostname is *compute-0*. Since your home directory is exported to all compute nodes, you should be able to go to */home/flow/PartekFlowRemoteWorker/*
24. To start the remote worker:
`./partekFlowRemoteWorker.sh head-node.local compute-0`
25. These two addresses should both be in the *10.1.0.0/16* address space. The remote worker will output to stdout when you run it. Scan for any errors. You should see the message *woot! I'm online*.
26. A successfully connected worker will show up on the *Resource management page* on the Partek Flow server. This can be reached from the main homepage or by clicking **Resource management** from the *Settings* page. Once you have confirmed the worker can connect, kill the remote worker (CTRL-C) from the terminal in which you started it.
27. Once everything is working, return to library file management and add the genomes/indices required by your research team. If Partek hosts these genomes/indices, these will automatically be downloaded by Partek Flow

Integration with your queueing system

1. In effect, all you are doing is submitting the following command as a batch job to bring up remote workers: */home/flow/PartekFlowRemoteWorker/partekFlowRemoteWorker.sh head-node.local compute-0*
2. The second parameter for this script can be obtained automatically via:
`$(hostname -s)`

Bringing up workers

Bring up workers by running the command below. You only need to run one worker per node:
/home/flow/PartekFlowRemoteWorker/partekFlowRemoteWorker.sh head-node.local compute-0

Shutting down workers

Go to the *Resource management page* and click on the **Stop button (red square)** next to the worker you wish to shut down. The worker will shut down gracefully, as in it will wait for currently running work on that node to finish, then it will shut down.

Updating Partek Flow

For the cluster update, you will get a link of .zip file for Partek Flow and remote Flow worker respectively from Partek support, all of the following actions should be performed as the Linux user that runs Flow. Do NOT run Flow as root.

1. Go to the Flow installation directory. This is usually the home directory of the Linux user that runs Flow and it should contain a directory named "**partek_flow**". The location of the Flow install can also be obtained by running *ps aux | grep flow* and examining the path of the running Flow executable.
2. Shut down Flow: *./partek_flow/stop_flow.sh*
3. Download the new version of Flow and the Flow worker:


```
wget --content-disposition http://packages.partek.com/linux/flow-release
wget --content-disposition http://packages.partek.com/linux/flow-worker-release
```
4. Make sure Flow has exited:
`ps aux | grep flow`
 The flow process should no longer be listed.
5. Unpack the new version of Flow install and backup the old install:
`mv partek_flow partek_flow_prev`
`mv PartekFlowRemoteWorker PartekFlowRemoteWorker_prev`
6. Backup the Flow database folder. This should be located in the home directory of the user that runs Flow.
`tar -czvf partek-db-bkp-date.tgz ~/.partekflow`
7. Start the updated version of Flow:
`./partek_flow/start_flow.sh`
`tail -f partek_flow/logs/catalina.out` (make sure there is nothing of concern in this file when starting up Flow. You can stop the file tailing by typing: *CTRL-C*)

You may also want to examine the the main Flow log for errors:

~/partekflow/logs/flow.log

Additional Assistance

If you need additional assistance, please visit [our support page](#) to submit a help ticket or find phone numbers for regional support.

[« Single Node Amazon Web Services Deployment](#) [Creating Restricted User Folders within the Partek Flow Server](#) »



You are unable to rate this content due to the filter settings.

Results: ★★★★★ 0 rates