Installing Bitfusion FlexDirect

Bitfusion Guide

vmware®

Table of Contents

Prerequisites

Review *System Requirements* for running FlexDirect as remote and partial GPU configuration. CUDA libraries (plus CUDNN and CUBLAS libraries if you are using them) should be installed on CPU Client nodes, but the GPU driver only needs to be installed on the GPU Server nodes.

> ⚠ **NOTE ON HEALTH CHECK RESULTS**
>
> **FATAL** issues will have a direct impact on the virtualization functionality
>
> **MARGINAL** issues can affect performance and scaling

License Key

Contact us to get your license key from your Bitfusion sales or support representative.

You will need to have the license key ready during installation of FlexDirect.

Installing FlexDirect

To install, download our installation script at getfd.bitfusion.io. You will use the `binaries` only option on your client machines (CPU-only) and one of another set of options on your GPU nodes—the options for the GPU nodes are mutually exclusive. Here are the commands for a CPU-only server:

Client (CPU only) Server Installation

```
# Download the install script.
wget -O installfd https://getfd.bitfusion.io


# Run install with option for client mode (just the binaries).
sudo bash installfd -- -m binaries
# Answer 'y' on installing any dependencies.


# Initialize the license.
sudo flexdirect init
# Accept the EULA.
# Enter your license key at prompt (issued by your representative at Bitfusion).
```

And here are the commands for the SRS option (virtualization services) for a GPU server. If you use this option, do not then install the FDA or FDM options as they will simply replace the SRS option you previously installed.

GPU Server Basic Service Installation

```
# Download the install script.
wget -O installfd https://getfd.bitfusion.io


# Run install with option for FlexDirect service mode, a systemd service.
sudo bash installfd -- -m srs
# Answer 'y' on installing any dependencies, unless you have reason to install yourself.
# Answer 'y' on starting the service (background daemon listening for GPU requests) if
you already have a license on the server, otherwise 'n.'

# Initialize the license.
sudo flexdirect init
# Accept the EULA.
# Enter your license key at prompt (issued by your representative at Bitfusion).
```

Here are the commands for installing the FlexDirect Manager (FDM) option for a GPU server. If you use this option, do not then install the SRS or FDA options as they will simply replace the the FDM option you previously installed. FlexDirect Manager provides management services in addition to all the virtualization services in the SRS option.

GPU Server FlexDirect Manager Installation

```
# Download the install script.
wget -O installfd https://getfd.bitfusion.io

# Run install with option for FlexDirect manager mode, a systemd service.
sudo bash installfd -- -m fdm
# Answer 'y' on installing any dependencies, unless you have reason to install yourself.
# Answer 'y' on starting the service (background daemon listening for GPU requests) if
you already have a license on the server, otherwise 'n.'

# Initialize the license.
sudo flexdirect init
# Accept the EULA.
# Enter your license key at prompt(issued by your representative at Bitfusion).
```

Here are the commands for installing the FlexDirect Analytics (FDA) option for a GPU server. If you use this option, do not then install the SRS or FDM options as they will simply replace the the FDA option you previously installed. FlexDirect Analytics is a service that gathers and plots GPU utilization statistics. It does not offer GPU virtualization services; instead, it helps you to evaluate if such services would be valuable.

GPU Server FlexDirect Analytics Installation

```
# Download the install script.
wget -O installfd https://getfd.bitfusion.io

# Run install with option for flexdirect analytics mode, a systemd service.
sudo bash installfd -- -m fda
# Answer 'y' on installing any dependencies, unless you have reason to install yourself.
# Answer 'y' on starting the service (background daemon analyzing GPU use).

# FDA does not require a license.
```

In these examples, we have named the local copy of the install script, `installfd`. This scripts retrieves a tar file for your OS, extracts its contents and then runs a second-stage install file which was one of files extracted. The second-stage install file can also take arguments, and you have the option of giving them to `installfd` to be passed through. Just list them after the `--` separator.

In the examples above, we have passed through a 'mode' argument. In the case of client servers we pass, `-m binaries`, because they have no GPUs and do not need to launch a FlexDirect service (under systemd) to respond to GPU requests. In the case of GPU servers with basic FlexDirect service we pass `-m sys`, which does launch a FlexDirect service to respond to requests to use the GPUs. In the case of GPU servers with FlexDirect Manager service, we pass `-m fdm`, which provides management services in addition to the basic FlexDirect service.

If you do not supply the mode argument, you will be prompted to choose one during the install.

You may also pass the `-s` option through to the second-stage installer for a silent install. It will not prompt you about installing dependencies nor about the starting a service, but assumes your answer is 'yes.'

To see the help menu for other options, pass through the `-h` option.

Finishing the Installation

Ensure you have a valid license (except for FDA) and that the service is running.

License and Starting the Service

```
# Verify your license. If all is well, it will report the number of nodes you have
licensed and how many days remain. Example output is shown.
# No license is needed for FDA.
$ flexdirect license
Getting license status...

5 out of 5 nodes in use.
25 days left in trial.

# Check the status of the service if it is not active, then start and recheck.
$ sudo systemctl start flexdirect

●    flexdirect-manager.service - Start FlexDirect Manager
     Loaded: loaded (/lib/systemd/system/flexdirect-manager.service; enabled; vendor
     preset: enabled)
     Active: inactive (dead) since Mon 2019-03-04 19:43:16 IST; 12s ago

$ systemctl status flexdirect

●    flexdirect-manager.service - Start FlexDirect Manager
     Loaded: loaded (/lib/systemd/system/flexdirect-manager.service; enabled; vendor
     preset: enabled)
     Active: active (running) since Thu 2019-02-28 16:20:20 IST; 4 days ago
```

Please contact us at *askbitfusion@vmware.com* if there are any issues.

Upgrading FlexDirect

After the initial installation, FlexDirect can be upgraded with the `upgrade` command. By default, the upgrade command will install the latest stable release. The upgrade command also accepts an argument to specify a specific release (whether earlier or later than your current version). We do not recommend using the upgrade command for releases prior to fd-1.11.4. Default and specific commands are shown below:

Upgrading FlexDirect

```
# Upgrade to the latest stable release.
sudo flexdirect upgrade

# Install release v 1.11.5.
sudo flexdirect upgrade -v 1.11.5
```

Command

After FlexDirect is successfully installed, you will find the command in `/usr/bin`.

FlexDirect Location

```
$ which flexdirect
/usr/bin/flexdirect
```

Please note that the FlexDirect service (including FlexDirect Manager) assumes it owns and can schedule the GPUs. Do not mix use-cases by having some end-users using FlexDirect Services and others running GPU applications without FlexDirect. This will lead to conflicts. Please see the *Usage* guide if you want to launch the FlexDirect Service with ownership of a subset of the available physical GPUs.

vmware®

Installed Files and Configuration Files

DIRECTORY	CONTENTS	COMMENTS
/usr/bin	flexdirect flex-docker	FlexDirect is the single executable for low-level and high-level GPU virtualization functionality on both the client- and server-side.
/opt/bitfusionio/lib/ x86_64-linux-gnu/ bitfusion/bin	dispatcher cuda-server etc.	These are processes launched by FlexDirect to communicate with client applications. These sit at a core-level (below the low-level) functionality and will only be used by customers integrating Bitfusion technology into their own products.
~/.bitfusionio	servers.conf *.log	• This is the default location for FlexDirect log files. • This is the high-precedence location for servers. conf (which lists the addresses of the GPU servers in the pool).
/etc/bitfusionio	/lic servers.conf	• This is the default location for the FlexDirect license files (created when you run `flexdirect init`). • This is the low-precedence location for servers.conf (which lists the addresses of the GPU servers in the pool).