

Hortonworks DataFlow

MiNiFi Quick Start

(February 28, 2018)

Hortonworks DataFlow: MiNiFi Quick Start

Copyright © 2012-2018 Hortonworks, Inc. Some rights reserved.



Except where otherwise noted, this document is licensed under
Creative Commons Attribution ShareAlike 4.0 License.
<http://creativecommons.org/licenses/by-sa/4.0/legalcode>

Table of Contents

1. MiNiFi Java Agent Quick Start	1
1.1. Overview	1
1.2. Before You Begin	1
1.3. Installing and Starting MiNiFi	2
1.3.1. Installing MiNiFi on Linux	2
1.3.2. Installing MiNiFi as a Service on Linux	2
1.3.3. Starting MiNiFi on Linux	2
1.3.4. Installing MiNiFi on Windows	3
1.3.5. Configuring the MiNiFi MSI	3
1.3.6. Starting MiNiFi on Windows	6
1.4. Working with Dataflows	6
1.4.1. Setting up Your Dataflow	6
1.4.2. Using Processors Not Packaged with MiNiFi	7
1.5. Securing your Dataflow	10
1.6. Managing MiNiFi	11
1.6.1. Monitoring Status	11
1.6.2. Loading a New Dataflow	12
1.6.3. Stopping MiNiFi	12
A. Supported Processors	14
A.1. Processors Supported by Java	14
A.2. Supported Processors Requiring NAR	16

List of Figures

- 1.1. HDF_MiNiFi_setup.png 4
- 1.2. Log_on_as_service.png 5
- 1.3. HDF_MiNiFi_setup.png 6

1. MiNiFi Java Agent Quick Start

This guide is intended to help you install and start using MiNiFi Java Agent quickly. For additional details, see the Administration Guide.

- [Overview \[1\]](#)
- [Before You Begin \[1\]](#)
- [Installing and Starting MiNiFi \[2\]](#)
- [Working with Dataflows \[6\]](#)
- [Securing your Dataflow \[10\]](#)
- [Managing MiNiFi \[11\]](#)

1.1. Overview

Apache NiFi, MiNiFi is an Apache NiFi project, designed to collect data at its source. MiNiFi was developed with the following objectives in mind:

- Small and lightweight footprint
- Central agent management
- Data provenance generation
- NiFi integration for follow-on dataflow management and chain of custody information

1.2. Before You Begin

MiNiFi is supported on the following operating systems:

- Red Hat Enterprise Linux / CentOS 6 (64-bit)
- Red Hat Enterprise Linux / CentOS 7 (64-bit)
- Ubuntu Trusty (14.04) (64-bit)
- Debian 7
- SUSE Linux Enterprise Server (SLES) 11 SP3 (64-bit)
- Windows

You can find download links for the following MiNiFi software in the [HDF Release Notes](#).

- MiNiFi Java Agent
- MiNiFi C++

- [MiNiFi Toolkit](#)

1.3. Installing and Starting MiNiFi

You have several options for installing and starting MiNiFi.

- [Installing MiNiFi on Linux \[2\]](#)
- [Installing MiNiFi as a Service on Linux \[2\]](#)
- [Starting MiNiFi on Linux \[2\]](#)
- [Installing MiNiFi on Windows \[3\]](#)
- [Configuring the MiNiFi MSI \[3\]](#)
- [Starting MiNiFi on Windows \[6\]](#)

1.3.1. Installing MiNiFi on Linux

To install MiNiFi on RHEL/CentOS, Ubuntu, Debian, SLES, complete the following steps:

1. Download MiNiFi.
2. Extract the file to the location from which you want to run the application.

1.3.2. Installing MiNiFi as a Service on Linux

You can also install MiNiFi as a service:

1. Navigate to the MiNiFi installation directory.
2. Enter:

```
bin/minifi.sh install
```

You can also specify a custom name for your MiNiFi installation, by specifying that name during your install command. For example, to install MiNiFi as a service and named dataflow, enter:

```
bin/minifi.sh install dataflow
```

1.3.3. Starting MiNiFi on Linux

Once you have downloaded and installed MiNiFi, you need to start MiNiFi

You can start MiNiFi in the foreground, background, or as a service on Linux.

Launching MiNiFi in the foreground:

1. From a terminal window, navigate to the MiNiFi installation directory.

2. Enter:

```
bin/minifi.sh run
```

Launching MiNiFi in the background:

1. From a terminal window, navigate to the MiNiFi installation directory.

2. Enter:

```
bin/minifi.sh start
```

Launching MiNiFi as a service:

1. From a terminal window, enter:

```
sudo service minifi start
```

1.3.4. Installing MiNiFi on Windows

Prerequisites

Before you begin your MiNiFi installation, be sure you meet the following requirements:

- Install JDK 8.0 64 bit.
- Install Java to `C:/java` instead of `C:/Program Files`.
Recent Windows versions mark everything in `C:\Program Files` as read only.
- Set the `JAVA_HOME` environment variable using the 8.3 style name conventions.
For example: `C:\Program\jdk1.8.0`.
- Ensure `JAVA_HOME` is pointing to a 64-bit JRE/JDK.
- Ensure the Domain user has administrator privilege.
- Ensure your system meets the minimum memory requirement for Windows which is 4GB.

You can install MiNiFi using windows MSI:

1. Extract the MiNiFi MSI files in the repo location at `http://public-repo-1.hortonworks.com/HDF/windows/3.x/updates/3.1.1.0/minifi-3.1.1.0-35.msi` to the location from which you want to run the application.
2. Execute the MSI.

1.3.5. Configuring the MiNiFi MSI

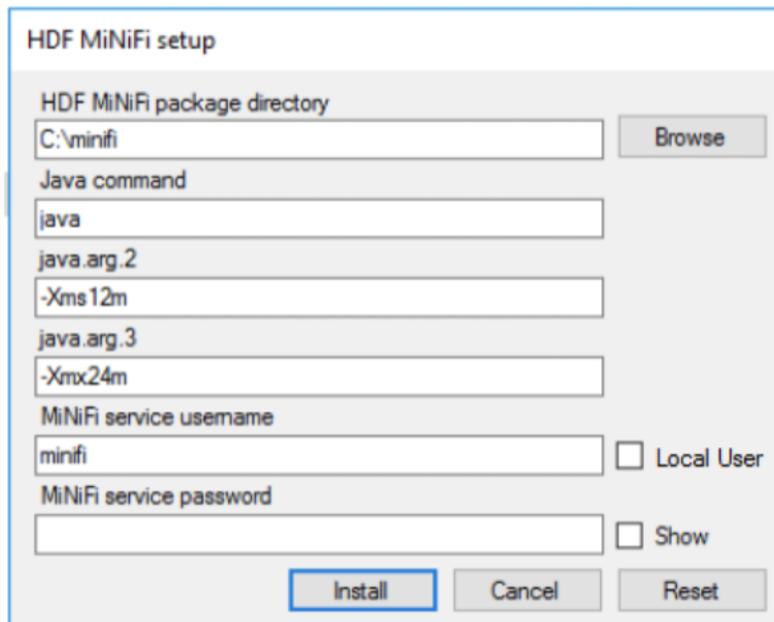
The MSI adds the Windows service for MiNiFi. The service is configured to be executed by either a local user in the computer, or a domain user in ActiveDirectory.

1.3.5.1. Using a Local User for MiNiFi Windows Service

There is no prerequisite to use a Local user for the Windows service. The installer automatically sets up the user.

1. If the computer is a part of a domain, then **Local User** checkbox appears in the **HDF NiFi setup** window. Check the **Local User** checkbox to specify that Local user is used to execute the installed service.

Figure 1.1. HDF_MiNiFi_setup.png



If a user specified at **MiNiFi service username** does not exist, the installer creates one with the specified **MiNiFi service password**. If the user already exists, the installer updates its password with the specified password.

The installer also grants the following privileges to the specified user:

- SeCreateSymbolicLinkPrivilege
- SeServiceLogonRight

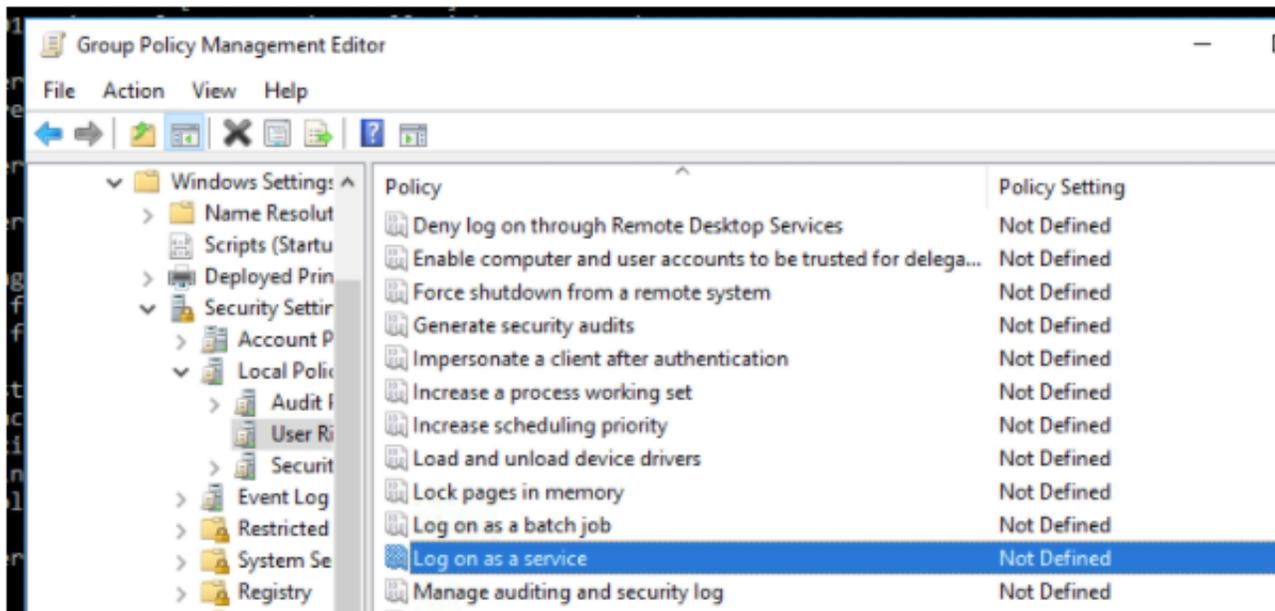
1.3.5.2. Using a Domain User for MiNiFi Windows Service

Prerequisites

- The computer must be part of the domain.
- The specified user must exist in the domain, and a correct password must be provided.
- ActiveDirectory PowerShell module must be available.

1. In the Group Policy Management Editor, set permission to 'Log on as a service.'

Figure 1.2. Log_on_as_service.png



2. Navigate to a machine on which MiNiFi is installed and enter the following command:

```
gpupdate
```

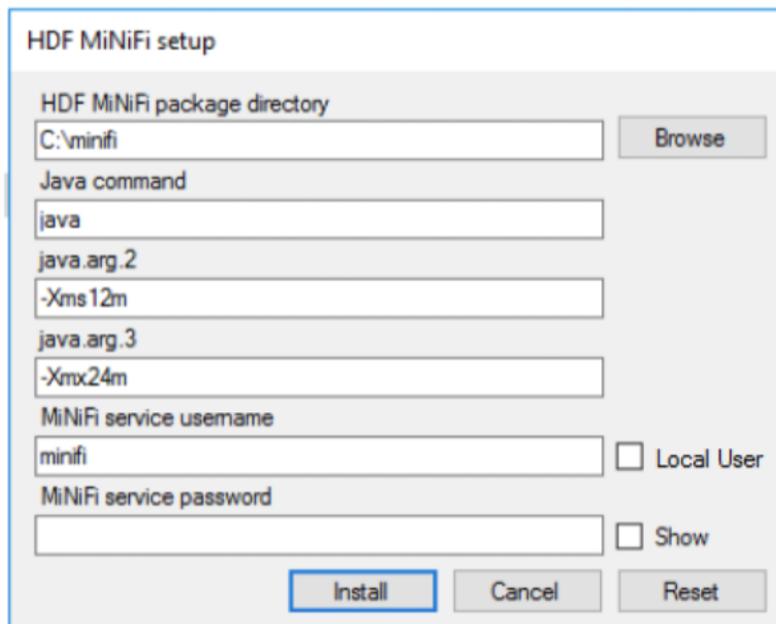
The **gpupdate** command is a machine-wide command and can be executed from any directory on the MiNiFi machine.

3. Install the ActiveDirectory PowerShell module by entering the following in the PowerShell console:

```
Add-WindowsFeature RSAT-AD-PowerShell
```

4. In the HDF NiFi setup window, uncheck 'Local User' checkbox then click **Install**.

Figure 1.3. HDF_MiNiFi_setup.png



HDF MiNiFi setup

HDF MiNiFi package directory
C:\minifi

Java command
java

java.arg.2
-Xms12m

java.arg.3
-Xmx24m

MiNiFi service username
minifi Local User

MiNiFi service password
 Show

After installation, you can update Java options at `nifi-install-dir\conf\bootstrap.conf` file. Repository locations are at `nifi-install-dir\conf\nifi.properties` file.

1.3.6. Starting MiNiFi on Windows

Once you have downloaded and installed MiNiFi, you can start MiNiFi in the foreground or as a service on Windows.

Launching MiNiFi in the foreground:

1. From a command prompt window, navigate to the MiNiFi installation directory.
2. Enter:

```
bin\run-minifi.bat
```

Launching MiNiFi as a service:

- You can start or stop the installed MiNiFi service from the Windows Service Manager.

1.4. Working with Dataflows

When you are working with a MiNiFi dataflow, you should design it, add any additional configuration your environment or use case requires, and then deploy your dataflow. MiNiFi is not designed to accommodate substantial mid-dataflow configuration.

1.4.1. Setting up Your Dataflow

Before you begin, you should be aware that the following NiFi components are not supported in MiNiFi dataflows:

- Funnels
- Multiple source relationships for a single connection
- Process groups

Additionally, each processor requires a distinct name.

You can use the MiNiFi Toolkit, located in your MiNiFi installation directory, and any NiFi instance to set up the dataflow you want MiNiFi to run:

1. Launch NiFi
2. Create a dataflow.
3. Convert your dataflow into a template.
4. Download your template as an `.xml` file.

For more information on working with templates, see the [Templates](#) section in the *User Guide*.

5. From the MiNiFi Toolkit, run the following command to turn your `.xml` file into a `.yaml` file:

```
config.sh transform input_file output_file
```

6. Move your new `.yaml` file to `minifi/conf`.
7. Rename your `.yaml` file `config.yaml`.



Note

You can use one template at a time, per MiNiFi instance.

Result: Once you have your `config.yaml` file in the `minifi/conf` directory, launch that instance of MiNiFi and your dataflow begins automatically.

1.4.2. Using Processors Not Packaged with MiNiFi

MiNiFi is able to use the following processors out of the box:

- UpdateAttribute
- AttributesToJSON
- Base64EncodeContent
- CompressContent
- ControlRate
- ConvertCharacterSet
- ConvertJSONToSQL

- DetectDuplicate
- DistributeLoad
- DuplicateFlowFile
- EncryptContent
- EvaluateJsonPath
- EvaluateRegularExpression
- EvaluateXPath
- EvaluateXQuery
- ExecuteProcess
- ExecuteSQL
- ExecuteStreamCommand
- ExtractText
- FetchDistributedMapCache
- FetchFile
- FetchSFTP
- GenerateFlowFile
- GetFTP
- GetFile
- GetHTTP
- GetJMSQueue
- GetJMSTopic
- GetSFTP
- HandleHttpRequest
- HandleHttpResponse
- HashAttribute
- HashContent
- IdentifyMimeType
- InvokeHTTP

- ListFile
- ListSFTP
- ListenHTTP
- ListenRELP
- ListenSyslog
- ListenTCP
- ListenUDP
- LogAttribute
- MergeContent
- ModifyBytes
- MonitorActivity
- ParseSyslog
- PostHTTP
- PutDistributedMapCache
- PutEmail
- PutFTP
- PutFile
- PutJMS
- PutSFTP
- PutSQL
- PutSyslog
- QueryDatabaseTable
- ReplaceText
- ReplaceTextWithMapping
- RouteOnAttribute
- RouteOnContent
- RouteText
- ScanAttribute
- ScanContent

- SegmentContent
- SplitContent
- SplitJson
- SplitText
- SplitXml
- TailFile
- TransformXml
- UnpackContent
- ValidateXml

If you want to create a dataflow with a processor not shipped with MiNiFi, you can do so.

1. Set up your dataflow as described above.
2. Copy the desired NAR file into the MiNiFi lib directory.
3. Restart your MiNiFi instance.



Note

Currently only the StandardSSLContextService is supported as a controller service. It is created automatically if the "Security Properties" section is set and can be referenced in the processor configuration using the ID "SSL-Context-Service".

1.5. Securing your Dataflow

You can secure your MiNiFi dataflow using keystore or trust store SSL protocols, however, this information is not automatically generated. You will need to generate your security configuration information yourself.

To run a MiNiFi dataflow securely, modify the Security Properties section of your `config.yml` file.

1. Create your dataflow template as discussed above.
2. Move it to `minifi.conf` and rename `config.yml`.
3. Manually modify the Security Properties section of `config.yml`.

```
Security Properties:
keystore:
keystore type:
keystore password:
key password:
```

```
truststore:  
truststore type:  
truststore password:  
ssl protocol: TLS  
Sensitive Props:  
key:  
algorithm: PBEWITHMD5AND256BITAES-CBC-OPENSSL  
provider: BC
```

1.6. Managing MiNiFi

You can also perform some management tasks using MiNiFi

1.6.1. Monitoring Status

You can use the `minifi.sh flowStatus` option to monitor a range of aspects of your MiNiFi operational and dataflow status. You can use the `flowStatus` option to get information dataflow component health and functionality, a MiNiFi instance, or system diagnostics.

FlowStatus accepts the following flags and options:

- processors
 - health
 - bulletins
 - status
- connections
 - health
 - stats
- remoteProcessGroups
 - health
 - bulletins
 - status
 - authorizationIssues
 - inputPorts
- controllerServices
 - health
 - bulletins
- provenancereporting

- health
- bulletins
- instance
 - health
 - bulletins
 - status
- systemdiagnostics
 - heap
 - processorstats
 - contentrepositoryusage
 - flowfilerepositoryusage
 - garbagecollection

For example, this query gets the health, stats, and bulletins for the TailFile processors

```
minifi.sh flowStatus processor:TailFile:health,stats,bulletins
```



Note

Currently the script only accepts one high level option at a time.

Any connections, remote process groups or processors names that contain ":", ";" or "," will cause parsing errors when querying.

For details on the flowStatus option, see the [FlowStatus Query Option](#) section of the *Administration Guide*.

1.6.2. Loading a New Dataflow

You can load a new dataflow for a MiNiFi instance to run:

1. Create a new `config.yml` file with the new dataflow.
2. Replace the existing `config.yml` in `minifi/conf` with the new file.
3. Restart MiNiFi.

1.6.3. Stopping MiNiFi

You can stop MiNiFi at any time.

Stopping MiNiFi:

1. From a terminal window, navigate to the MiNiFi installation directory.
2. Enter:

```
bin/minifi.sh stop
```

Stopping MiNiFi as a service:

1. From a terminal window, enter:

```
sudo service minifi stop
```

Appendix A. Supported Processors

When you first obtain the Java version of MiNiFi, several processors are supported without you taking any further action. Other processors are a part of the default distribution but require you to add a NAR for a controller service not packaged by default. T

A.1. Processors Supported by Java

- UpdateAttribute
- AttributesToJSON
- Base64EncodeContent
- CompressContent
- ControlRate
- ConvertCharacterSet
- ConvertJSONToSQL
- DetectDuplicate
- DistributeLoad
- DuplicateFlowFile
- EncryptContent
- EvaluateJsonPath
- EvaluateRegularExpression
- EvaluateXPath
- EvaluateXQuery
- ExecuteProcess
- ExecuteSQL
- ExecuteStreamCommand
- ExtractText
- FetchDistributedMapCache
- FetchFile
- FetchSFTP
- GenerateFlowFile

- GetFTP
- GetFile
- GetHTTP
- GetJMSQueue
- GetJMSTopic
- GetSFTP
- HandleHttpRequest
- HandleHttpResponse
- HashAttribute
- HashContent
- IdentifyMimeType
- InvokeHTTP
- ListFile
- ListSFTP
- ListenHTTP
- ListenRELP
- ListenSyslog
- ListenTCP
- ListenUDP
- LogAttribute
- MergeContent
- ModifyBytes
- MonitorActivity
- ParseSyslog
- PostHTTP
- PutDistributedMapCache
- PutEmail
- PutFTP
- PutFile

- PutJMS
- PutSFTP
- PutSQL
- PutSyslog
- QueryDatabaseTable
- ReplaceText
- ReplaceTextWithMapping
- RouteOnAttribute
- RouteOnContent
- RouteText
- ScanAttribute
- ScanContent
- SegmentContent
- SplitContent
- SplitJson
- SplitText
- SplitXml
- TailFile
- TransformXml
- UnpackContent
- ValidateXml

A.2. Supported Processors Requiring NAR

The following processors are also a part of the default distribution but require you to add a NAR for a controller service not packaged by default. The processors are grouped by the NAR that is required.

- nifi-dbc-service-nar
 - ConvertJSONToSQL
 - PutSQL
 - GenerateTableFetch

- ListDatabaseTable
- QueryDatabaseTable
- ExecuteSQL
- nifi-distributed-cache-services-nar
 - DetectDuplicate
 - FetchDistributedMapCache
 - PutDistributedMapCache
- nifi-http-context-map-nar
 - HandleHttpRequest
 - HandleHttpResponse