

HyperData 데이터 수집

- MINIFI 구동 및 설정 가이드

MINIFI 구동 및 설정 가이드

1. 해당 가이드는 원격 사이트에서 실행되는 MINIFI 구동 및 설정법에 대한 가이드입니다
 - NIFI가 제대로 동작하고 설정이 제대로 되었음을 가정합니다
2. NIFI에 대한 가이드는 NIFI 가이드 문서를 참고해주세요
3. 본 가이드에서 다루는 적재 시나리오는 두 가지 상황입니다
 - ① 1개의 원격지 1개 디렉토리 -> 1개의 오존 버킷
 - ② 다중 원격지 디렉토리 -> 1개의 오존 버킷

MINIFI 구동 및 설정 가이드

Prerequisites

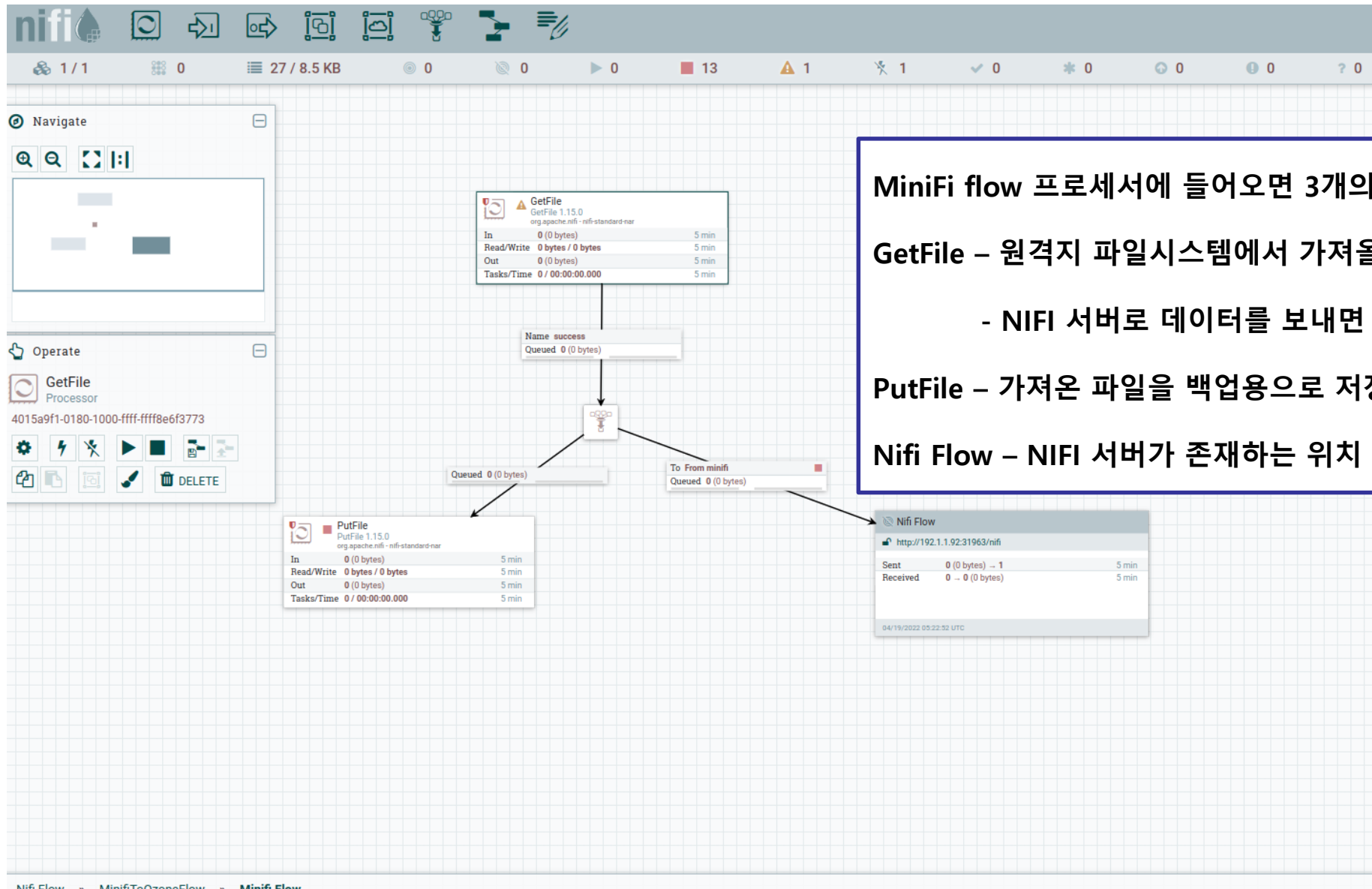
- 아래 변수를 입력하고 OzoneConfigUpdate 프로세스를 실행시켜서 PutHDFS가 제대로 인식하는지 참고해주세요 (NIFI 적재 가이드문서 6p~8p 참고)

The screenshot shows the Apache NiFi interface with a 'Variables' dialog box open. The dialog displays the following table of variables:

Scope	Name	Value
MinifiToOzoneFlow	Bucket	bucket
MinifiToOzoneFlow	Institution	institution
MinifiToOzoneFlow	OzoneConfigUpdateID	19aa466d-a102-3fff-8cd2-6e391965c0f2
MinifiToOzoneFlow	Volume	volume

Below the table, there is a text area containing the value '19aa466d-a102-3fff-8cd2-6e391965c0f2'. To the right of the table, there are sections for 'Referencing Processors' (ExecuteProcess) and 'Referencing Controller Services' (None). At the bottom of the dialog, there are buttons for 'CANCEL', 'OK', and 'APPLY'.

MINIFI 구동 및 설정 가이드



MINIFI 구동 및 설정 가이드

Nifi Flow – NIFI 서버가 존재하는 위치 정보를 올바르게 입력해주세요

- Apply를 누르고 Enable transmission을 클릭했을 때 해당 프로세스에서 노란색 느낌표 표시가 나온다면 제대로 입력이 안된 경우입니다.
- 해당 프로세스 수정이 안된다면 Disable transmission을 먼저 해주고 수정을 해주세요

Configure Remote Process Group

Name
Nifi Flow

Id
0fe2ce3a-5719-3f78-afaa-e4811c55a08e

URLs
http://마스터 IP:NIFI 파드의 NodePort/nifi

Transport Protocol ?
HTTP

Local Network Interface ?

HTTP Proxy Server Hostname ?
마스터 IP

HTTP Proxy Server Port ?
NIFI 파드의 NodePort

HTTP Proxy User ?

HTTP Proxy Password ?

Communications Timeout ?
30 sec

Yield Duration ?
10 sec

CANCEL APPLY

MINIFI 구동 및 설정 가이드

GetFile – Input Directory에 적혀있는 폴더명이
원격지에서 데이터를 가져오는 폴더명입니다.

- 기본값 (HDdata)
- /opt/minifi/ (해당 경로는 수정하지 마세요)
- 여기 입력되는 폴더명은 실제로 MINIFI 도커
를 구동할 때 마운트되는 폴더명과 같아야합니
다.

Configure Processor | GetFile 1.15.0

Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Input Directory	/opt/minifi/HDdata
File Filter	[^\\].*
Path Filter	No value set
Batch Size	10
Keep Source File	false
Recurse Subdirectories	true
Polling Interval	0 sec
Ignore Hidden Files	true
Minimum File Age	5 sec
Maximum File Age	No value set
Minimum File Size	0 B
Maximum File Size	No value set

CANCEL APPLY

MINIFI 구동 및 설정 가이드

PutFile – Input Directory에 적혀있는 폴더명이
전송하는 데이터를 백업하는 폴더입니다

- 기본값 (HDbbackup)
- /opt/minifi/ (해당 경로는 수정하지 마세요)
- 여기 입력되는 폴더명은 실제로 MINIFI 도커
를 구동할 때 마운트되는 폴더명과 같아야합니
다.

Configure Processor | PutFile 1.15.0

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Directory	/opt/minifi/HDbbackup
Conflict Resolution Strategy	fail
Create Missing Directories	true
Maximum File Count	No value set
Last Modified Time	No value set
Permissions	No value set
Owner	No value set
Group	No value set

CANCEL APPLY

Nifi Flow > MinifiToOzoneFlow > Minifi Flow

MINIFI 구동 및 설정 가이드

1. MINIFI 구동을 위한 명령어는 다음과 같습니다.

```
docker run -d -v $(pwd)/bootstrap.conf:/opt/minifi/minifi1.16.0/conf/bootstrap.conf
-v $(pwd)/config.yml:/opt/minifi/minifi-1.16.0/conf/config.yml
-v $(pwd)/HDdata:/opt/minifi/HDdata
-v $(pwd)/HDbackup:/opt/minifi/HDbackup
--name HDdata
apache/nifi-minifi:1.16.0
```

-v 옵션 왼쪽에 들어가는 파일 및 디렉토리의 경로는 **절대경로여야 합니다!**

--name 옵션은 여러 MINIFI 구동을 위한 컨테이너 관리용으로 Input 파일명을 입력해주세요

실시간 로그를 보고싶다면 -d 옵션을 빼주세요

다중 원격지 디렉토리 구동 및 설정 가이드

2. 다중 원격지 디렉토리 구동 및 설정 가이드

- 동일 Bucket으로 전송한다면 NIFI상에서 프로세스를 추가하거나 다시 템플릿을 다운받고 변환하는 과정없이 MINIFI 도커 컨테이너를 하나 더 구동시키면 됩니다.
- EX) 기존 MINIFI에서 HDdata 폴더 파일을 전송 중이고 HDdata2를 추가하고 싶을 때

```
Processors: []
Controller Services: []
Process Groups:
- id: 334bf91a-0166-3938-0000-000000000000
  name: Minifi Flow
  Processors:
  - id: ea17bc69-f4f4-341e-0000-000000000000
    name: GetFile
    class: org.apache.nifi.processors.standard.GetFile
    max concurrent tasks: 1
    scheduling strategy: TIMER_DRIVEN
    scheduling period: 0 sec
    penalization period: 30 sec
    yield period: 1 sec
    run duration nanos: 0
    auto-terminated relationships list: []
    Properties:
      Batch Size: '10'
      File Filter: '^[^\\.]*'
      Ignore Hidden Files: 'true'
      Input Directory: /opt/minifi/HDdata2
      Keep Source File: 'false'
      Maximum File Age:
      Maximum File Size:
      Minimum File Age: 5 sec
      Minimum File Size: 0 B
      Path Filter:
      Polling Interval: 0 sec
      Recurse Subdirectories: 'true'
  - id: 76e3a7bf-f725-333c-0000-000000000000
```

편집기로 config.yml 열어서 왼쪽 빨간 부분의 경로를 수정해
주세요

Ex) HDdata2

다중 원격지 디렉토리 구동 및 설정 가이드

2. 다중 원격지 디렉토리 구동 및 설정 가이드

- 만약 백업 폴더의 경로를 수정하고 싶다면 아래의 경로를 수정해주시면 됩니다.
- 다중 원격지 디렉토리의 경우 bootstrap.conf 와 config.yml이 각 MINIFI마다 존재해야 함으로 MINIFI 폴더를 만들어서 관리를 부탁드립니다.
- Bootstrap.conf는 같은 파일을 사용하시면 됩니다.

```
name: PutFile
class: org.apache.nifi.processors.standard.Put
max concurrent tasks: 1
scheduling strategy: TIMER_DRIVEN
scheduling period: 0 sec
penalization period: 30 sec
yield period: 1 sec
run duration nanos: 0
auto-terminated relationships list:
- failure
- success
Properties:
  Conflict Resolution Strategy: fail
  Create Missing Directories: 'true'
  Directory: /opt/minifi/HDbbackup
  Group:
  Last Modified Time:
  Maximum File Count:
  Owner:
  Permissions:
```

각 MINIFI 마다 bootstrap.conf 와
config.yml이 필요

```
hkl@hkl:~/Downloads/minifi/minifi1$ ls
bootstrap.conf  config.yml
```

다중 원격지 디렉토리 구동 및 설정 가이드

2. 다중 원격지 디렉토리 구동 및 설정 가이드

- config.yml 파일을 다 수정했다면, 실행명령어는 이전과 크게 다르지 않습니다

```
docker run -d -v $(pwd)/bootstrap.conf:/opt/minifi/minifi1.16.0/conf/bootstrap.conf
```

```
-v $(pwd)/config.yml:/opt/minifi/minifi-1.16.0/conf/config.yml
```

```
-v $(절대경로)/HDdata2:/opt/minifi/HDdata2
```

```
-v $(절대경로)/HDbackup:/opt/minifi/HDbackup
```

```
--name HDdata2
```

```
apache/nifi-minifi:1.16.0
```

```
hkl@hkl:~/Downloads/minifi/HDbackup$ docker ps
```

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NAMES
8564b31691d7	apache/nifi-minifi:1.16.0	"/bin/sh -c \${MINIFI..."	5 seconds ago	Up 5 seconds		HDdata
a9885a426b94	apache/nifi-minifi:1.16.0	"/bin/sh -c \${MINIFI..."	About a minute ago	Up About a minute		HDdata2