

HyperData 데이터 수집

- NIFI를 이용한 S2S Ozone Storage 적재 가이드

1. HyperData, Ozone이 설치된 Namespace에서 Nifi를 설치합니다

- Install.sh 파일을 이용해서 설치 및 제거를 할 수 있습니다
 - 설치 - `./install.sh install Namespace명`
 - 제거 - `./install.sh uninstall Namespace명`
- 설치과정에서 배포된 Template이 제대로 업로드 되지 않았다면 다음 명령어로 Template을 업로드 시킬 수 있습니다
 - `./install.sh upload Namespace명`

NIFI 설치 및 설정방법



2. Nifi상단 메뉴에서 Template을 불러옵니다.

- Template명 – MinifiToOzone
- 해당 Template 존재하지 않는다면 설치 스크립트 `./install.sh upload` Namespace명 으로 템플릿을 다시 업로드할 수 있습니다
- 혹은 빈 팔레트에서 오른쪽 클릭 -> upload template으로 배포 된 `./conf-files/MinifiToOzone.xml`를 업로드 할 수 있습니다

NIFI 설치 및 설정방법

프로세스 그룹을 더블클릭하면 하위 프로세스로 들어갈 수 있습니다

MinifiToOzoneFlow	
Queued	0 (0 bytes)
In	0 (0 bytes) → 1 (1 remote) 5 min
Read/Write	0 bytes / 0 bytes 5 min
Out	0 → 0 (0 bytes) 5 min

NIFI 설치 및 설정방법

The screenshot displays the Apache NiFi web interface. The top status bar shows various system metrics. The left sidebar contains a 'Navigate' panel and an 'Operate' panel for the 'MinifiToOzoneFlow' process group. The main canvas shows a flow diagram with a 'From minifi' node connected to a 'PutHDFS' node, which is then connected to an 'UpdateAttribute' node. A context menu is open over the 'From minifi' node, with the 'Variables' option highlighted. The flow diagram also shows a 'Name failure' error on the 'PutHDFS' node. Performance metrics for 'Minifi Flow', 'Ozone Config Update', 'PutHDFS', and 'UpdateAttribute' are displayed as pop-up cards.

Minifi Flow

Queued	0 (0 bytes)
In	0 (0 bytes) → 0
Read/Write	0 bytes / 0 bytes
Out	0 → 0 (0 bytes)

Ozone Config Update

Queued	0 (0 bytes)
In	0 (0 bytes) → 0
Read/Write	0 bytes / 0 bytes
Out	0 → 0 (0 bytes)

PutHDFS

In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

UpdateAttribute

In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

3. 환경 변수 설정

- 빈 팔레트에서 오른쪽 클릭을 하면 전체 프로세스에 대한 설정을 할 수 있습니다
- Variables를 클릭하여 환경변수를 입력할 수 있습니다

NIFI 설치 및 설정방법

3. 환경 변수 설정

- Volume - 전송한 데이터를 저장할 Ozone의 Volume
- Bucket - 전송한 데이터를 저장할 Ozone의 Bucket
- Institution - 데이터를 전송하는 사이트를 구분할 기관명
(공백없이 입력해주세요)
- OzoneConfigUpdateID - **OzoneConfigUpdate** 프로세스를 클릭하여 ID를 확인할 수 있습니다

The screenshot displays the Apache NiFi web interface. The top navigation bar shows the 'nifi' logo and various status icons. The main workspace is divided into several panels. On the left, the 'Navigate' panel shows a tree view of the NiFi instance. The central panel displays the 'Variables' configuration for the 'MinifiToOzoneFlow' process group. A table lists the variables: 'Bucket' (value: bucket), 'Institution' (value: institution), 'OzoneConfigUpdateID' (value: 19aa466d-a102-3fff-8cd2-6e391965c0f2), and 'Volume' (value: volume). The 'OzoneConfigUpdateID' row is highlighted. To the right of the table, there are sections for 'Variables', 'Referencing Processors', 'Referencing Controller Services', and 'Unauthorized Referencing Components'. Below the table, there is a text input field containing the ID '19aa466d-a102-3fff-8cd2-6e391965c0f2'. At the bottom right, the 'Ozone Config Update' process group is shown in the 'Operate' tab, with the same ID circled in red. The bottom status bar shows the number of tasks and the time taken to execute the process group.

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

19aa466d-a102-3fff-8cd2-6e391965c0f2

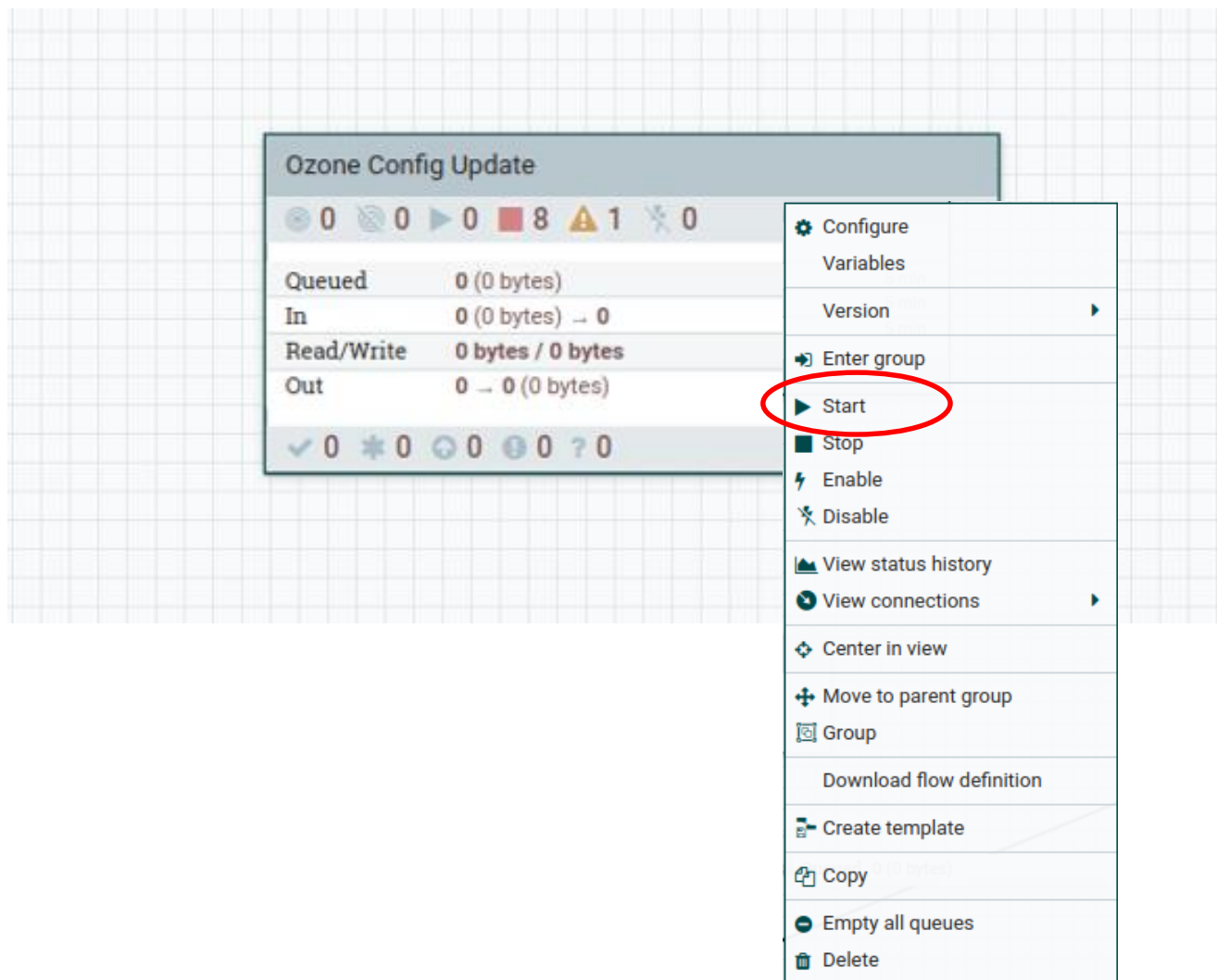
Ozone Config Update

19aa466d-a102-3fff-8cd2-6e391965c0f2

Out 0 (0 bytes) 5 min

Tasks/Time 0 / 00:00:00.000 5 min

NIFI 설치 및 설정방법



3. 환경 변수 설정

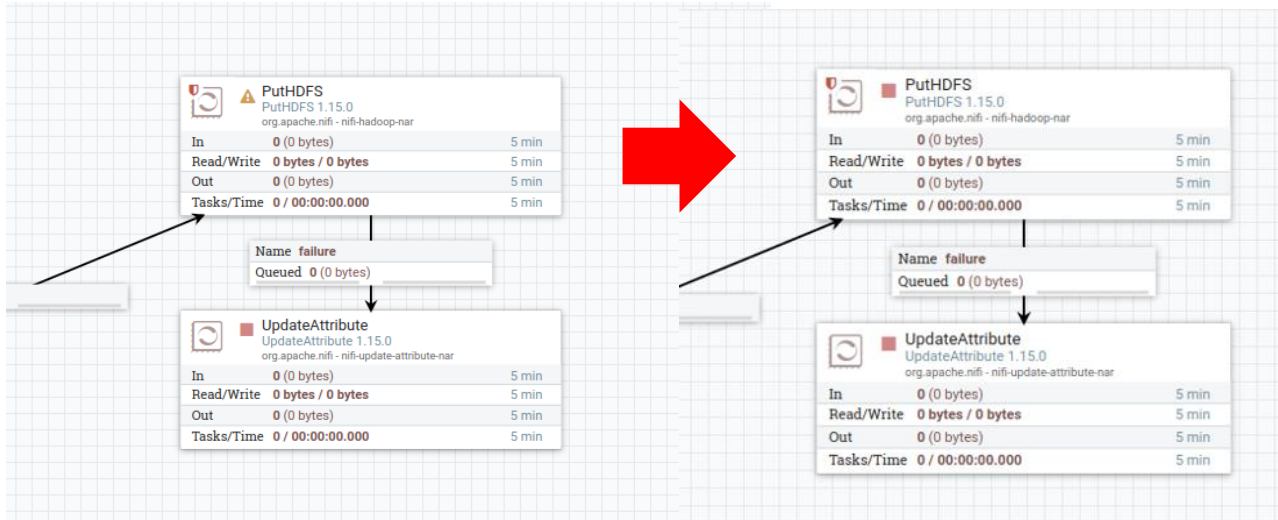
- 환경 변수를 입력하고 적용시킨 뒤 Ozone Config Update 프로세스 그룹을 오른쪽 클릭하여 해당 프로세스 그룹만 실행시킵니다

NIFI 설치 및 설정방법

Ozone Config Update		
Queued	0 (0 bytes)	
In	0 (0 bytes) → 0	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 → 0 (0 bytes)	5 min

Ozone Config Update

- 입력 받은 환경변수를 이용하여 NIFI가 Ozone과 통신하도록 설정해주는 프로세스입니다
- 프로세스가 완료되면 자동으로 중지됩니다
- 시간이 지나도 자동으로 중지되지 않는다면 입력한 환경변수를 다시 확인해주세요
- 프로세스가 완료되면 PutHDFS 프로세스의 상태가 아래 그림과 같이 변경됩니다



Minifi 설치 및 설정방법

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 Flow

0 1 0 1 1 0

Queued	0 (0 bytes)	
In	0 (0 bytes) → 0	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 → 0 (0 bytes)	5 min

✓ 0 * 0 0 0 ? 0

4. Minifi Template 설정 및 다운로드

- Minifi Flow 프로세스 그룹을 클릭하여 하위 프로세스로 들어갑니다
- GetFile의 설정을 변경해줍니다
 - GetFile – Input Directory에 적혀있는 폴더명이 원격지에서 데이터를 가져오는 폴더명입니다
 - /opt/minifi/{데이터를 전송할 폴더명}
 - 해당 폴더에 위치한 파일들은 자동으로 Ozone에 전송됩니다

Minifi 설치 및 설정방법

Configure Processor | PutFile 1.15.0

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Directory	/opt/minifi/HDBackup
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

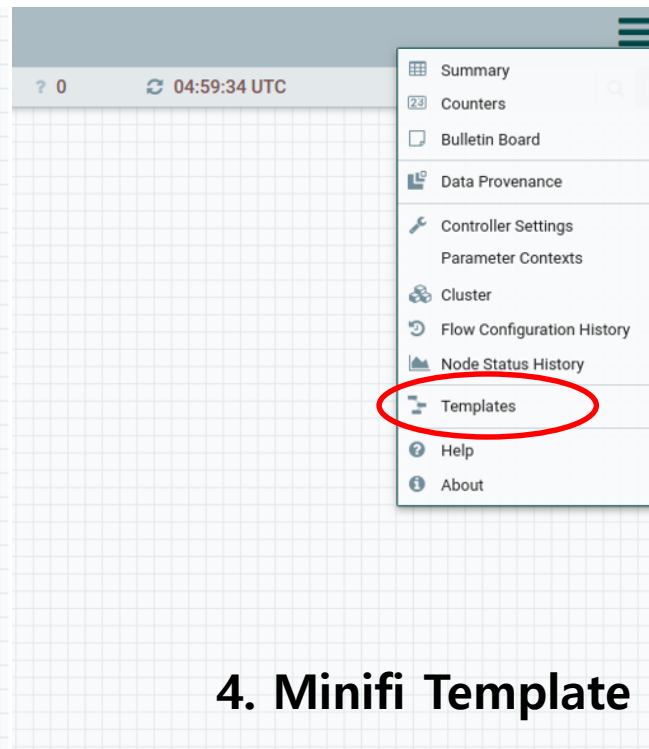
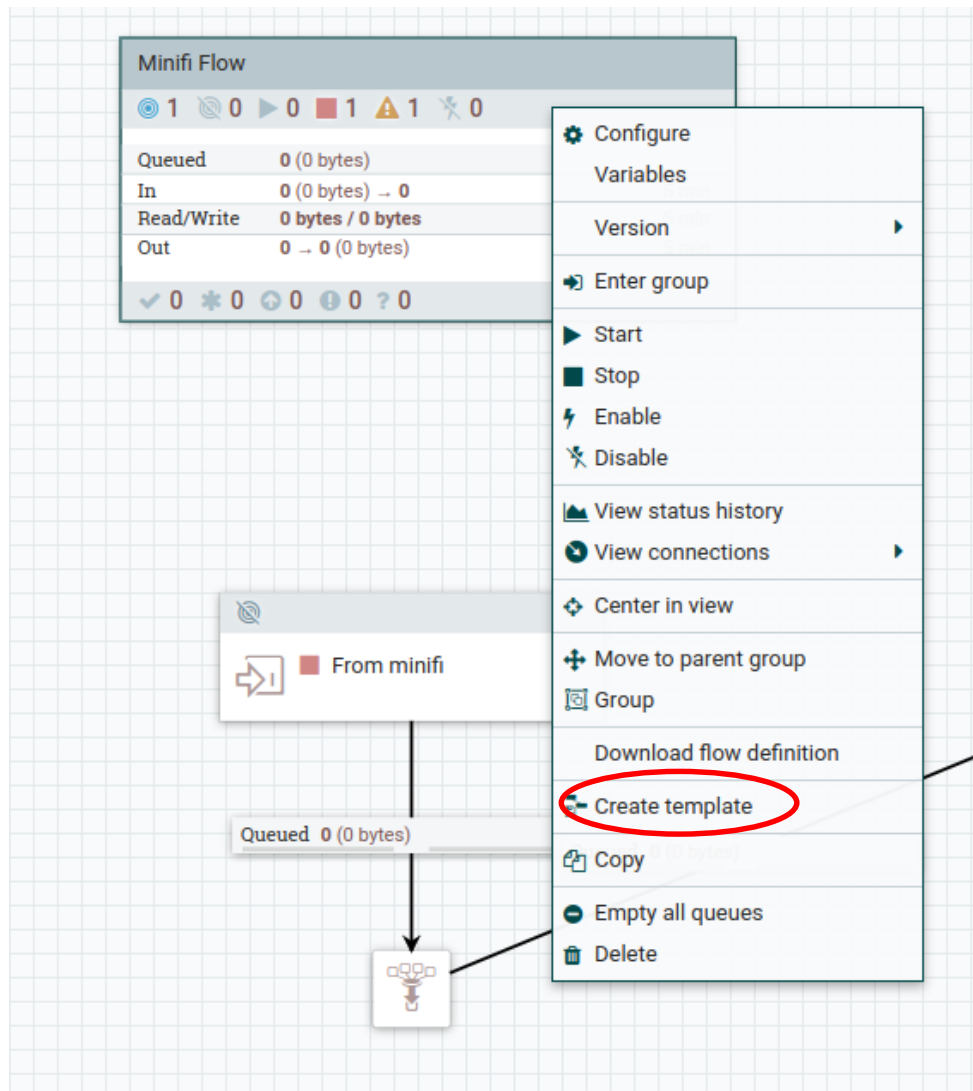
CANCEL APPLY

Minifi Flow			
0	1	0	1
0	0	1	0
Queued	0 (0 bytes)		
In	0 (0 bytes) → 0	5 min	
Read/Write	0 bytes / 0 bytes	5 min	
Out	0 → 0 (0 bytes)	5 min	
0	0	0	0

4. Minifi Template 설정 및 다운로드

- Minifi Flow 프로세스 그룹을 클릭하여 하위 프로세스로 들어갑니다
- PutFile의 설정을 변경해줍니다
 - PutFile – Input Directory에 적혀있는 폴더명이 전송하는 데이터를 백업하는 폴더입니다
 - /opt/minifi/{데이터를 전송할 폴더명}
 - 전송한 데이터를 로컬 시스템에서 백업 용으로 보관하는 용도입니다

NIFI 설치 및 설정방법



4. Minifi Template 설정 및 다운로드

- Minifi Flow를 Template화하고 원격지로 다운받습니다
- **Template명은 중복되지 않아야 합니다**
- NIFI 상단 오른쪽 메뉴에서 만들어진 Template 리스트 확인과 다운을 받을 수 있습니다

NIFI 설치 및 설정방법

NiFi Templates

Displaying 2 of 2

Filter

by name

Date/Time	Name	Description	Process Group Id	
03/29/2022 04:58:50 UTC	Minifi-Flow	Empty string set	d3fcc715-017f-1000-ffff-ffffe13033f0	
03/29/2022 04:40:27 UTC	MinifiToOzone	Empty string set	3c46eeb8-b2ab-4578-a262-ac790db206c6	

Download

```
hkl@hkl:~/Downloads/minifi-toolkit-1.15.3$ ./bin/config.sh transform ./flow/Minifi-Flow.xml ./flow/config.yml

Java home: /usr/lib/jvm/java-8-openjdk-amd64
MiNiFi Toolkit home: /home/hkl/Downloads/minifi-toolkit-1.15.3

No validation errors found in converted configuration.
```

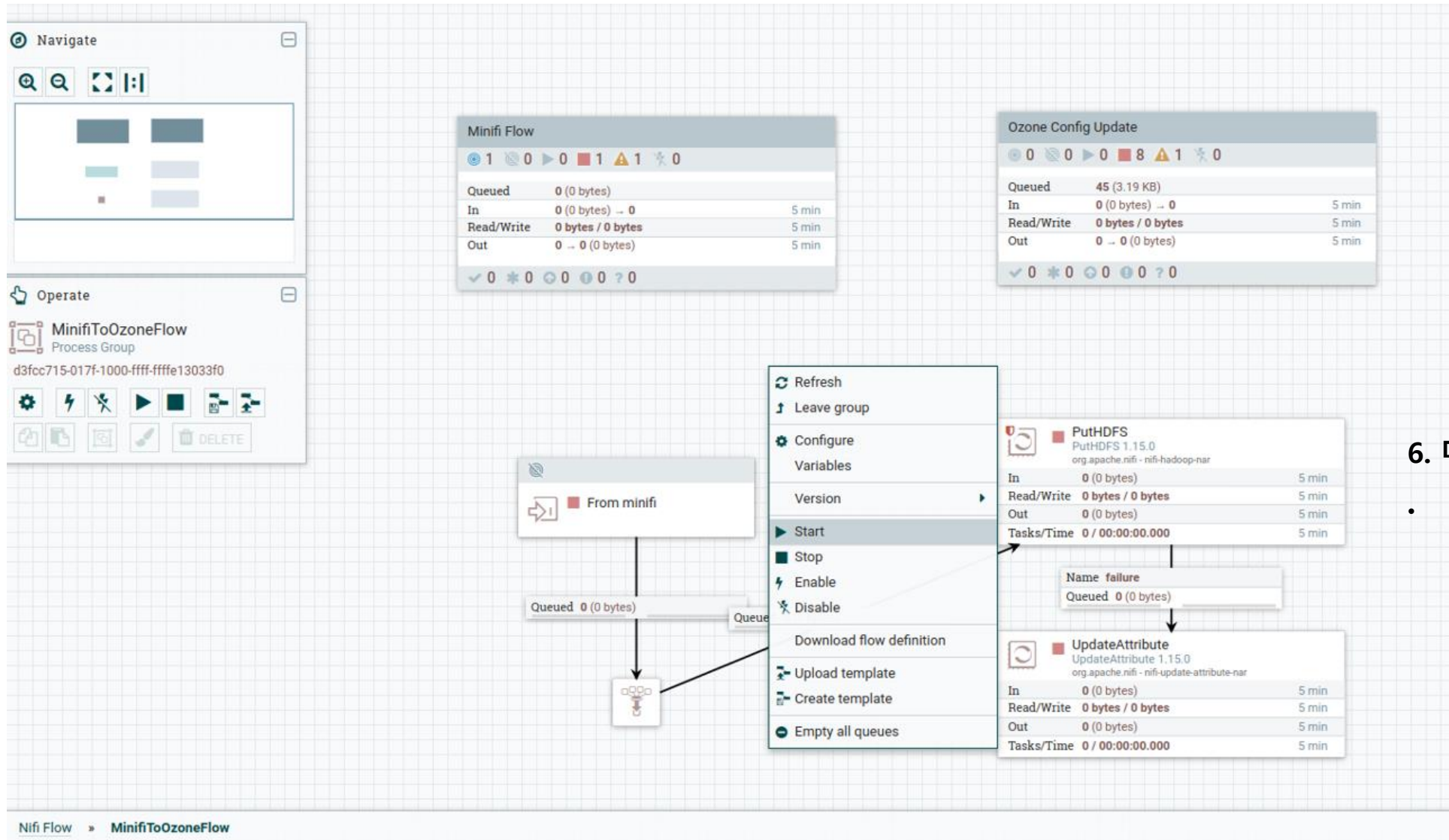
5. Minifi Template 적용 및 Minifi 실행

- 다운받은 Minifi template은 Minifi를 실행할 원격지에 위치해야 합니다
- 같이 배포된 minifi-toolkit을 이용해서 template(xml파일)을 config.yml파일로 변환시켜줍니다
 - `./bin/config.sh transform ./[minifi template].xml ./config.yml`
- “No validation errors found in converted configuration” 이외의 메시지가 표시되면,
 - Minifi Flow 프로세스 그룹만을 제대로 Template화 하였는지 확인 혹은 JAVA 버전을 확인해주세요 (JAVA 8 권장)

5. Minifi Template 적용 및 Minifi 실행

- 도커 환경이 필요합니다
- Minifi 실행을 위해서는 배포된 "bootstrap.conf" 그리고 수정을 완료한 "config.yml" 두개의 파일이 필요합니다
- **Minifi 구동 및 설정 방법에 대한 가이드는 minifi 폴더 안에 있는 MINIFI 설정 및 구동 가이드를 참고해주세요!**

NIFI 설치 및 설정방법



6. 데이터 전송

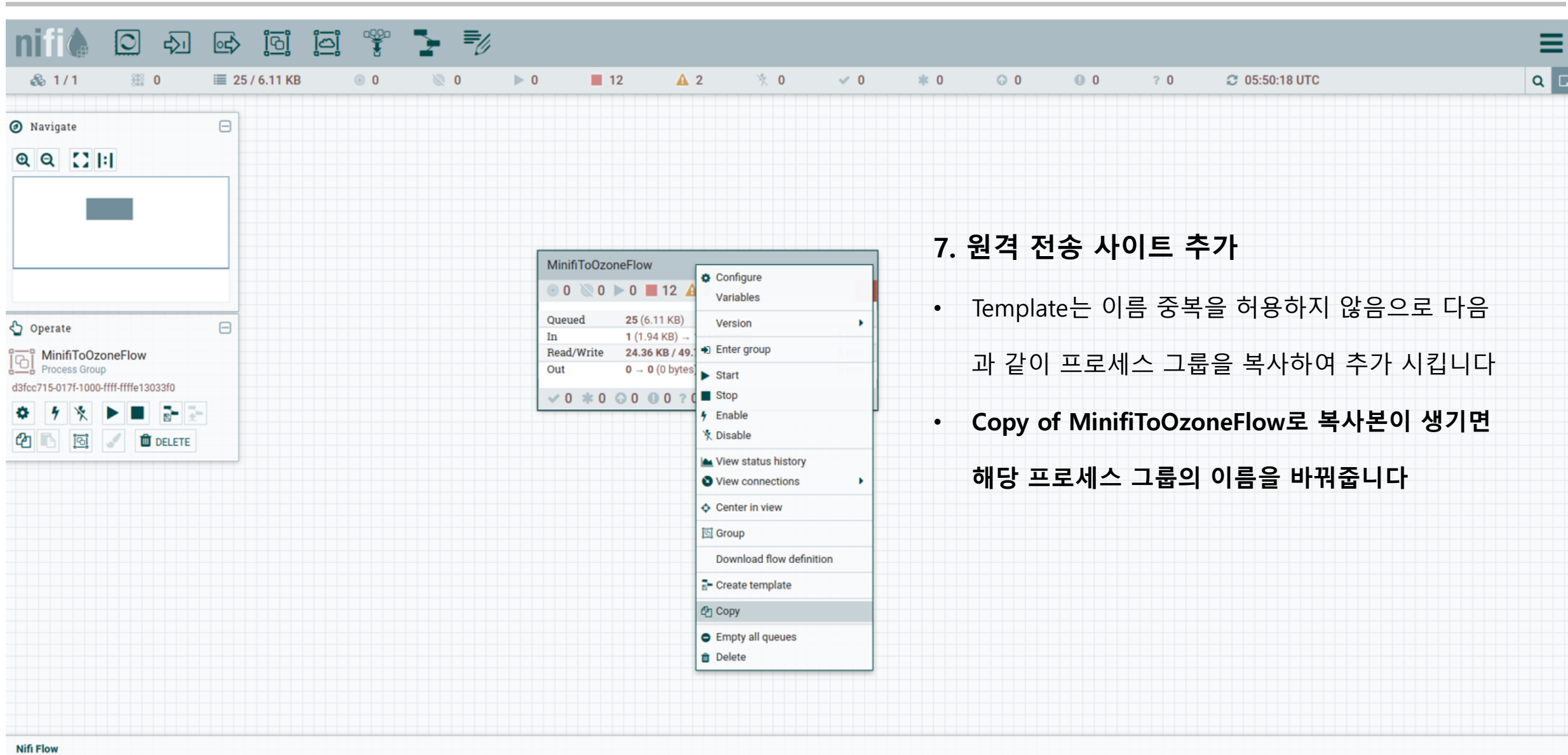
- 빈 팔레트에서 오른쪽 클릭 – Start로 From minifi -> PutHDFS 플로우를 실행시킵니다

```
{
  "volumeName" : "volume",
  "bucketName" : "bucket",
  "name" : "volume/bucket/test1.txt",
  "dataSize" : 1986,
  "creationTime" : "2022-03-29T05:46:24.181Z",
  "modificationTime" : "2022-03-29T05:46:27.184Z",
  "replicationType" : "RATIS",
  "replicationFactor" : 3
}
```

6. 데이터 전송

- 환경변수로 입력한 volume/bucket에 맞춰서 데이터가 오존에 적재됩니다
- 원격지에서 실행한 docker – minifi 컨테이너는 계속 실행중이어야 합니다
- NIFI UI 상에서 From Minifi – PutHDFS 플로우가 실행중이어야 합니다
- 원격 사이트 한 곳당 MinifiToOzone 프로세스 그룹 한 개가 있어야 합니다
- NIFI UI 상에서 Template을 추가하여 원격 전송 사이트를 추가할 수 있습니다

NIFI 설치 및 설정방법



The screenshot shows the Apache NiFi web interface. At the top, there's a navigation bar with the NiFi logo and various icons. Below it, a status bar shows metrics like '1 / 1', '0', '25 / 6.11 KB', and '05:50:18 UTC'. On the left, there's a 'Navigate' panel with search and view icons. The main area displays a 'Process Group' named 'MinifiToOzoneFlow' with a unique ID 'd3fcc715-017f-1000-ffff-ffffe13033f0'. A context menu is open over this process group, showing options like 'Configure', 'Variables', 'Version', 'Enter group', 'Start', 'Stop', 'Enable', 'Disable', 'View status history', 'View connections', 'Center in view', 'Group', 'Download flow definition', 'Create template', 'Copy' (highlighted), 'Empty all queues', and 'Delete'.

7. 원격 전송 사이트 추가

- Template는 이름 중복을 허용하지 않으므로 다음과 같이 프로세스 그룹을 복사하여 추가 시킵니다
- **Copy of MinifiToOzoneFlow**로 복사본이 생기면 해당 프로세스 그룹의 이름을 바꿔줍니다

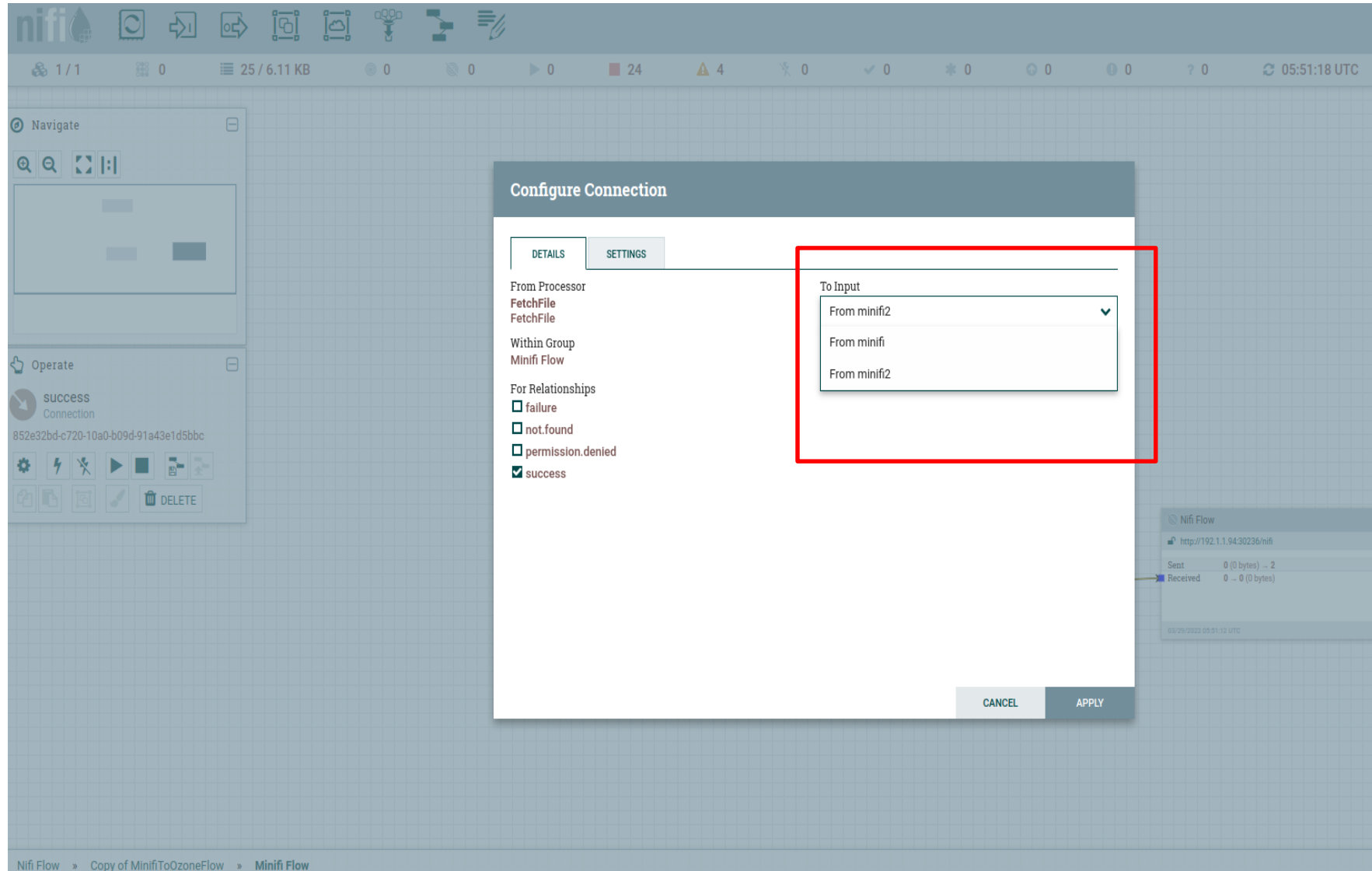
NIFI 설치 및 설정방법

The screenshot displays the Apache NiFi web interface. At the top, a status bar shows various metrics: 1/1, 0, 25/6.11 KB, 0, 0, 0, 24, 4, 0, 0, 0, 0, 0, 0. On the left, there are two panels: 'Navigate' and 'Operate'. The 'Operate' panel shows a process group named 'Copy of MinifiToOzoneFlow' with ID 'ed473446-a279-19c5-8ae6-ccc14780ff0f'. The main workspace shows a flow diagram. A red box highlights a 'Copy of From minifi' process node. This node is connected to a 'PutHDFS' node and an 'UpdateAttribute' node. The 'PutHDFS' node has a 'Name failure' error message. The 'UpdateAttribute' node is also connected to the 'PutHDFS' node. The flow diagram includes several status boxes: 'Minifi Flow', 'Ozone Config Update', and 'Copy of From minifi'. The 'Minifi Flow' box shows metrics: Queued 0 (0 bytes), In 0 (0 bytes) -- 0, Read/Write 0 bytes / 0 bytes, Out 0 -- 0 (0 bytes). The 'Ozone Config Update' box shows metrics: Queued 0 (0 bytes), In 0 (0 bytes) -- 0, Read/Write 0 bytes / 0 bytes, Out 0 -- 0 (0 bytes). The 'Copy of From minifi' box shows metrics: Queued 0 (0 bytes), In 0 (0 bytes) -- 0, Read/Write 0 bytes / 0 bytes, Out 0 -- 0 (0 bytes). The 'PutHDFS' box shows metrics: In 0 (0 bytes), Read/Write 0 bytes / 0 bytes, Out 0 (0 bytes), Tasks/Time 0 / 00:00:00.000. The 'UpdateAttribute' box shows metrics: In 0 (0 bytes), Read/Write 0 bytes / 0 bytes, Out 0 (0 bytes), Tasks/Time 0 / 00:00:00.000.

7. 원격 전송 사이트 추가

- Copy of From minifi 프로세스의 이름을 바꿔줍니다
- 다른 프로세스는 고유한 ID값으로 구분이 되기 때문에 이름을 바꾸지 않아도 됩니다

NIFI 설치 및 설정방법



7. 원격 전송 사이트 추가

- Minifi Flow 프로세스 그룹의 하위 프로세스로 들어갑니다
- **From Minifi** 커넥션을 클릭하여 **To input** 항목을 새롭게 추가한 **Input 커넥션**으로 바꿔줍니다
- 이후 과정은 처음에 진행한 과정과 같습니다
 - 환경변수를 설정하는 과정부터 다시 시작하면 됩니다 (5번 슬라이드 참조)