Apache NiFi Tutorial: What is NiFi? Architecture & Installation

By David Taylor   🕐 Updated July 4, 2023

Apache NiFi Tutorial Summary

This apache NiFi tutorial covers all the basic to advanced topics from scratch. You will learn the concepts like NiFi definition, history, architecture, features, installation process, use cases. You will also learn why you need to use apache NiFi and the best practices of running apache NiFi.

What is Apache NiFi?

Apache NiFi is an open source software for automating and managing the data flow between systems. It is a powerful and reliable system to process and distribute data. It provides web-based User Interface to create, monitor, and control data flows. It has a highly configurable and modifiable data flow process to modify data at runtime.

Apache NiFi is easily extensible through the development of custom components.

In this Apache NiFi tutorial, you will learn-

- What is Apache NiFi?
- Why Use Apache NIfi?
- History of Apache NiFi
- NiFi Architecture
- Apache NiFi Features
- How to Install Apache NiFi
- Nifi Use Cases
- Best practices Running Apache NiFi
- Disadvantage of Nifi

Why Use Apache NIfi?

Here, are reasons for using Apache Nifi:

- Allows you to do data ingestion to pull data into NiFi, from numerous data sources and create flow files
- It offers real-time control which helps you to manage the movement of data between any source & destination
- Visualize DataFlow at the enterprise level
- Provide common tooling and extensions
- Allows you to take advantage of existing libraries and Java ecosystem functionality
- Helps organizations to integrate Nifi with their existing infrastructure
- NiFi is designed to scale-out in clusters which offer guaranteed delivery of data
- Visualize and Monitor performance, behavior in a flow bulletin which offers insight and inline documentation
- Helps you to start and stop components separately or at the group level
- It helps you to listen, fetch, split, aggregate, route, transform and drag & drop Dataflow

History of Apache NiFi

- Developed at NSA for over eight years
- 2014- It was donated to the Apache Software Foundation
- 2015- NiFi became an official part of the Apache Project Suite
- Since then every 6-8 weeks, Apache NiFi releases a new update

NiFi Architecture

Apache NiFi has a well-thought-out architecture. Once data is fetched from external sources, it is represented as FlowFile inside Apache NiFi architecture.
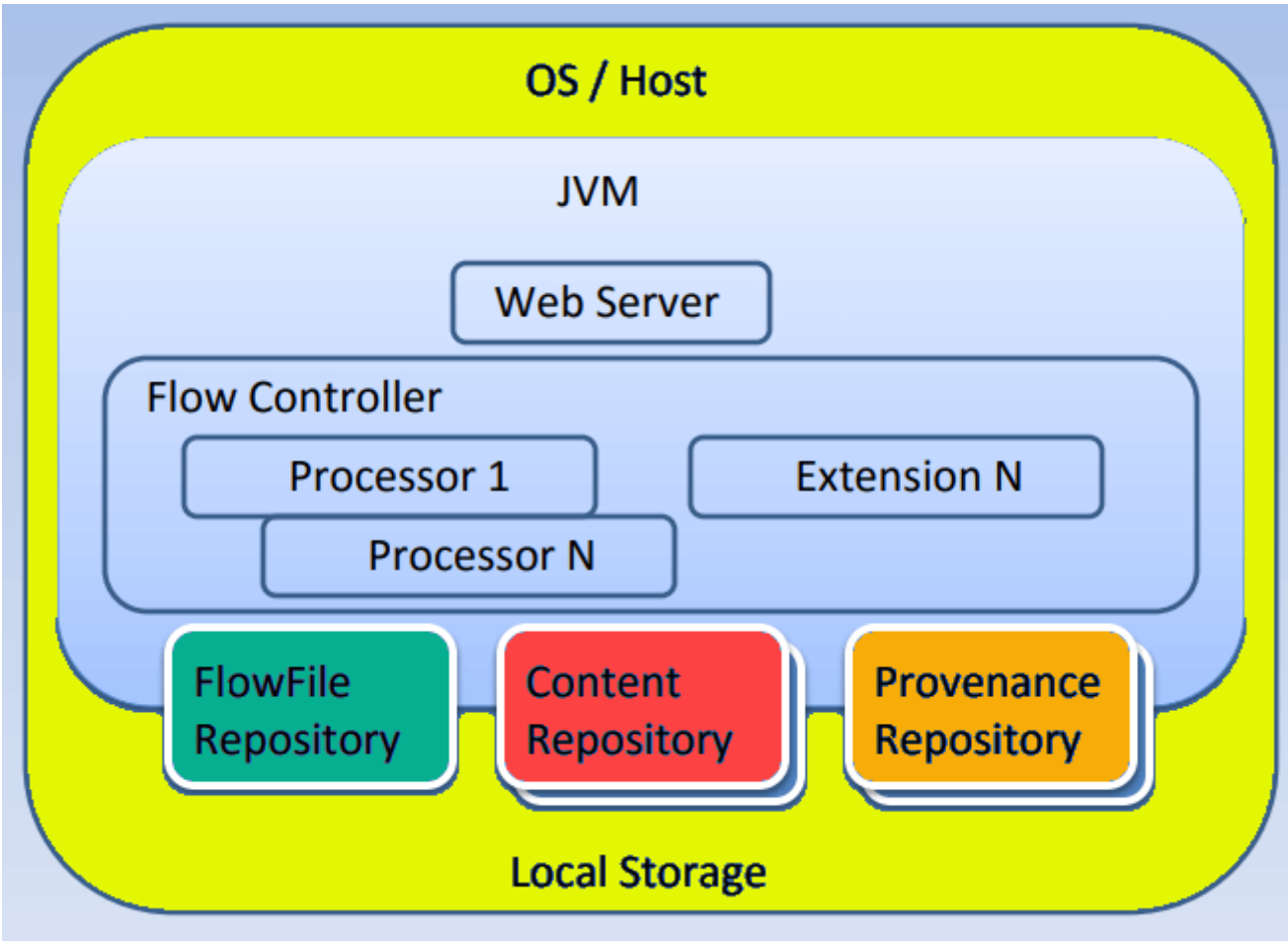
NiFi Architecture

Here are key components of NiFi architecture

Nifi Component	Description
FlowFile	FlowFile is original data with meta-information attached to it. It allows you to process not only CSV or other record-based data, but also pictures, videos, audio, or any other binary data.
Flowfile processor	Performs the work which acts as a building block of data flow in NiFi.
Flow controller	Keeps a record of how processes are connected. It manages the threads and allocations thereof which all processes use.
Web Server	Web server hosts NiFi's HTTP-based commands and API.
Extension	There are many types of NiFi extensions which operate and

execute within the JVM.

Connection	Acts as a linkage between processors that contain a queue and relationship(s) which affects where data is routed.
Back Pressure	Stop the system of becoming overrun by controlling the quantity or data size of flow files that can be stored in the queue.
Process Group	A process group is a set of processes and their connections, which receives and send data with the help of ports.
Flowfile Repository	In the FlowFile Repository, NiFi keeps track of the state of what details it has about a given FlowFile which is active in the flow.
Content Repository	The Content Repository is an area where the actual content bytes of a given FlowFile exist.
Provenance Repository	The Provenance Repository is an area where all provenance event data is gathered.
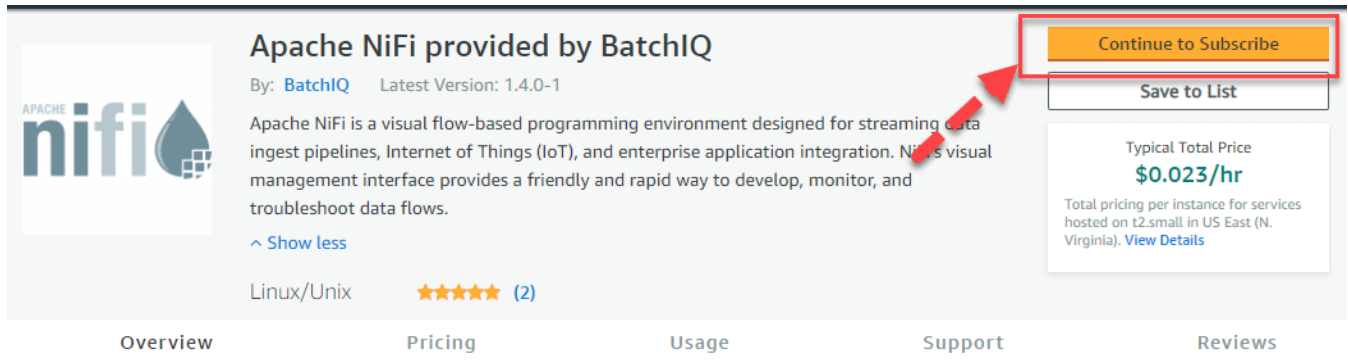
Apache NiFi Features

- NiFi supports buffering of all queued data and offers an ability of back pressure as those queues may reach specified limits
- NiFi allows the setting of one or more prioritization schemes
- Provides connection processors for many data sources
- Support any device which runs Java
- Ideal for limited connectivity places
- Support for troubleshooting and flow optimization
- Offers role-based authentication/authorization
- Allows download, recovery, and replay of individual files
- Build your processors, controller services, and more
- Provide content encryption, communication over secure protocols
- Enables rapid development and effective testing
- Allows for the development of simple single-function components that can be reused and combined to make more complex flows
- Allows classloader isolation for easier management of dependencies

How to Install Apache NiFi

Below is a step by step process for Apache NiFi installation

Step 1) Go to the link,

and click "Continue to Subscribe"

Step 2) On the next page,

Click "Accept Terms"

Step 3) You will see this page,

Thank you for subscribing to this product! We are processing your request.
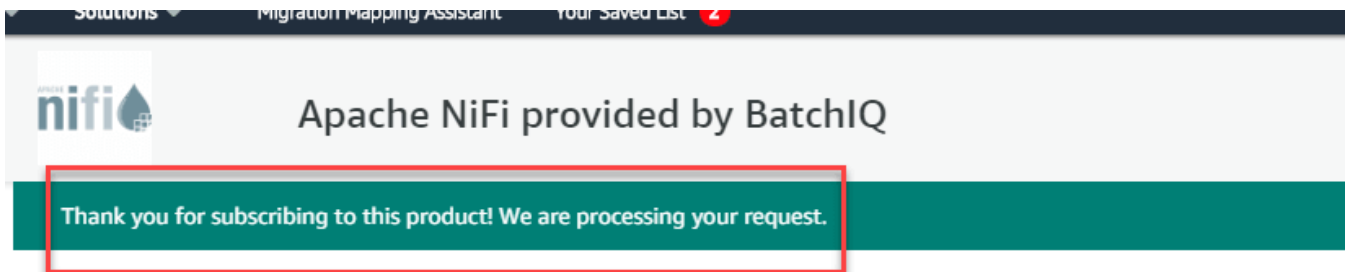
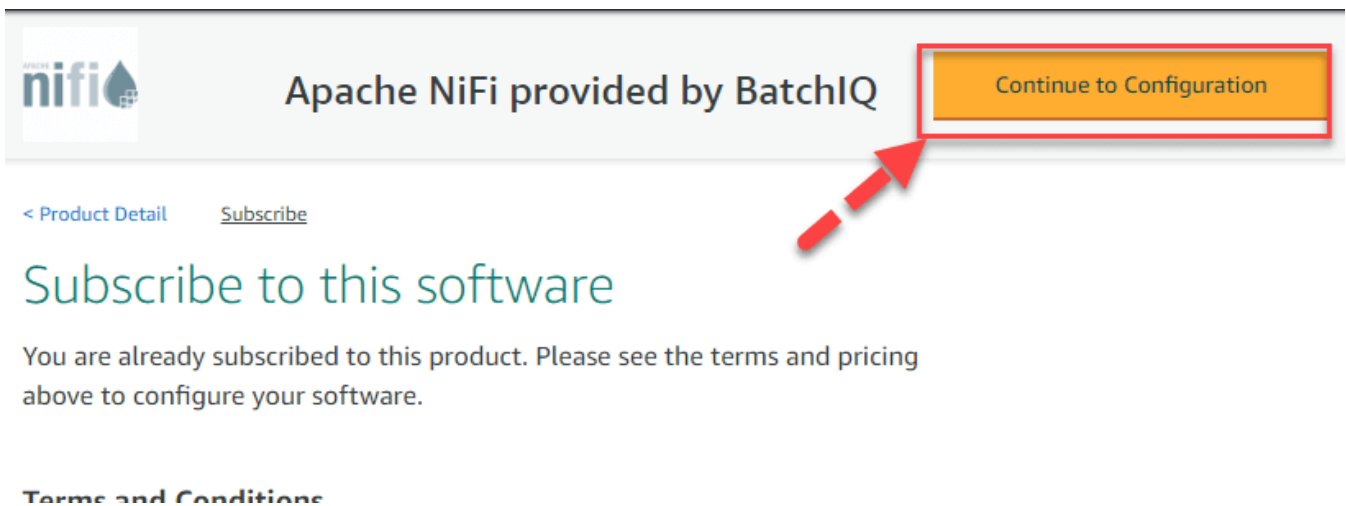**Step 4)** Refresh the page after 5 minutes.

Click on "Continue to Configuration"

Step 5) On the next page,

Keep settings default and click "Continue to Launch"

Step 6) On the next page,

Click on Launch. You may need to create a key

Step 7) You will see this success message.

Congratulations! An instance of this software is successfully deployed on EC2!

Apache NiFi provided by BatchIQ

Congratulations! An instance of this software is successfully deployed on EC2!

AMI ID: ami-fa13df80 (View Launch Configuration Details)
You can view this instance on EC2 Console. You can also view all instances on Your Software. Software and AWS hourly usage fees apply when the instance is running and will appear on your monthly bill.

Step 8) Note,

The instance id and the public DNS of the EC2 instance

	Name	Instance ID	Instance Type	Availability Zone	Instance State
☐					🔴 stopped
☐					🔴 stopped
☑		i-0507caeaf41ddb301	t2.small	us-east-1e	🟢 running
☐	SAS				🔴 stopped

Instance: i-0507caeaf41ddb301   Public DNS: ec2-100-26-156-57.compute-1.amazonaws.com

Step 9) In the security group,

Add all traffic rule to inbound and outbound

	Name	Group ID	Group Name	VPC ID	Description
☑		sg-058d8e2040472d6b1	launch-wizard-1	vpc-2b469d51	launch-wizard-1 created 2018-

Security Group: sg-058d8e2040472d6b1

Description	**Inbound** ①	Outbound	Tags

Edit

Type ⓘ ②	Protocol ⓘ	Port Range ⓘ	Source ⓘ
All traffic	All	All	0.0.0.0/0
All traffic	All	All	::/0

Step 10) To access Nifi,
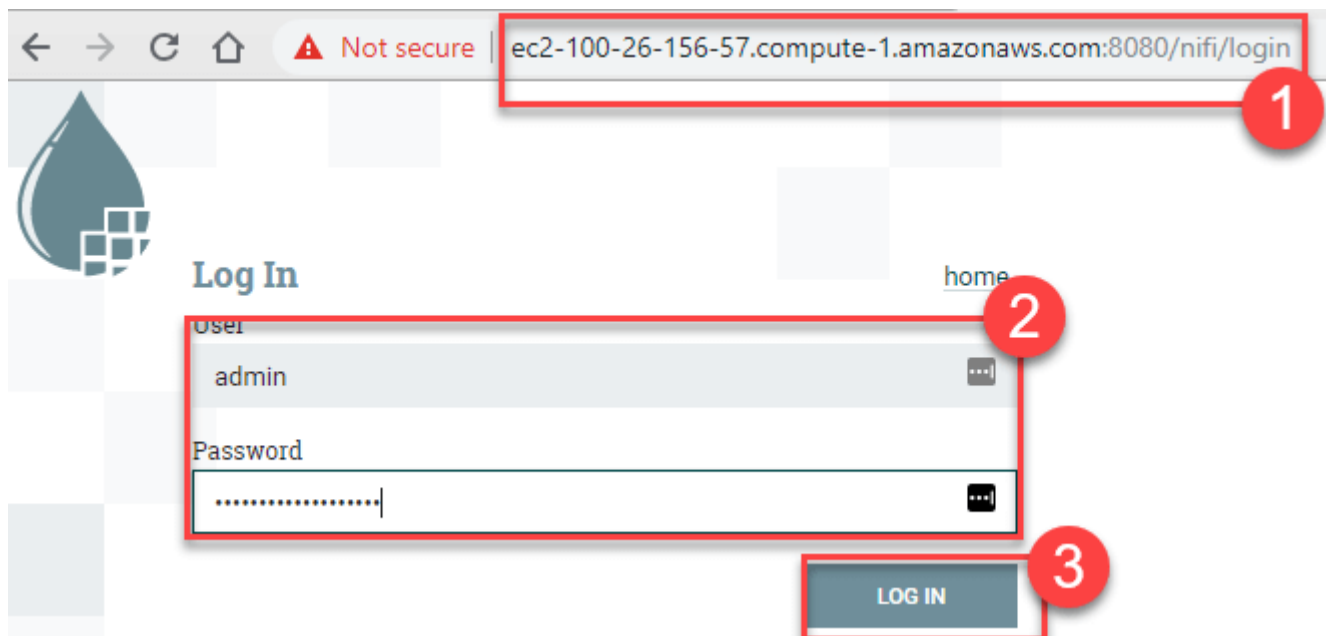
simply use the URL

```
http://publicdns:808/nifi
```

In our case it becomes

```
http://ec2-100-26-156-57.compute-1.amazonaws.com:8080/nifi/
```

User: admin

Password: Instance Id

Step 11) You will see,

NiFi home screen
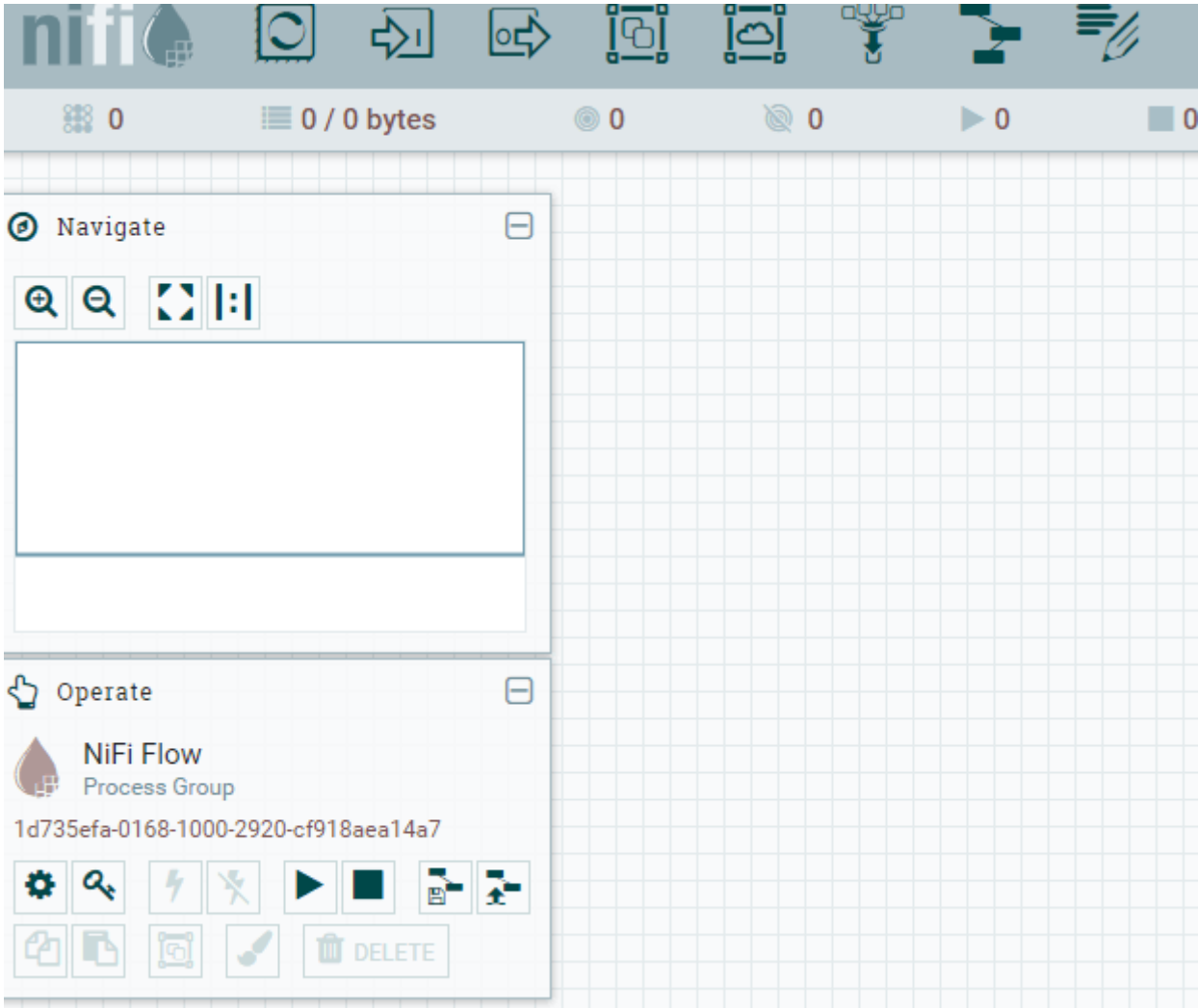
Nifi Use Cases

Below is a list of Apache NiFi use cases:

Industry	Usage
Insurance	- Risk & underwriting analysis - Claims Analytics - Usage-based Insurance - New product development
HealthCare	- Single view of Patient - Real-time vital sign monitoring - EMR optimization - Supply Chain Optimization

Telecommunication	- Single view of the customer - CDR analysis - Dynamic Bandwidth allocation
Manufacturing	- Preventative Maintenance - Supply Chain Optimization - Quality Control
Oil & Gas- Industry	- Real-time monitoring - Single view of the Operation - Predictive Maintenance - Archive & Analytics - Unstructured data classification
Financial Services	- Anti-money laundering - Fraud- Detection - Risk- data management

Best practices Running Apache NiFi

- Ideal to separate test/dev/production environments in NiFi
- You should break your flow into process groups
- Use a naming convention, use comments and labels
- Organize your projects into three parts ingestion, test & monitoring
- Use unique names for variable

Disadvantage of Nifi

- Need precise security and compliance controls
- You need to know the underlying system very well while working with Apache NiFi

- Must maintain chain of custody for data
- Transport / Messaging may not prove enough
- Data access needs exceed available resources to transport
- Not all data is created equally
- SSL and topic level authorization may not be sufficient

Summary

- Apache NiFi is an open source software for automating and managing the flow of data between systems
- NiFi is designed to scale-out in clusters which offer guaranteed delivery of data
- Nifi was developed at NSA for over eight years
- Once data is fetched from external sources, it is represented as FlowFile inside Apache NiFi architecture.
- FlowFile, processor, controller, web server, connection, back pressure, repository are important components of NiFi architecture
- NiFi expression language supports any device which runs Java
- You can easily install NiFi on AWS
- NiFi is used in varied industries such as healthcare, insurance, telecom, manufacturing, finance, oil and gas among others
- As a best practice, organize your projects into three parts ingestion, test & monitoring