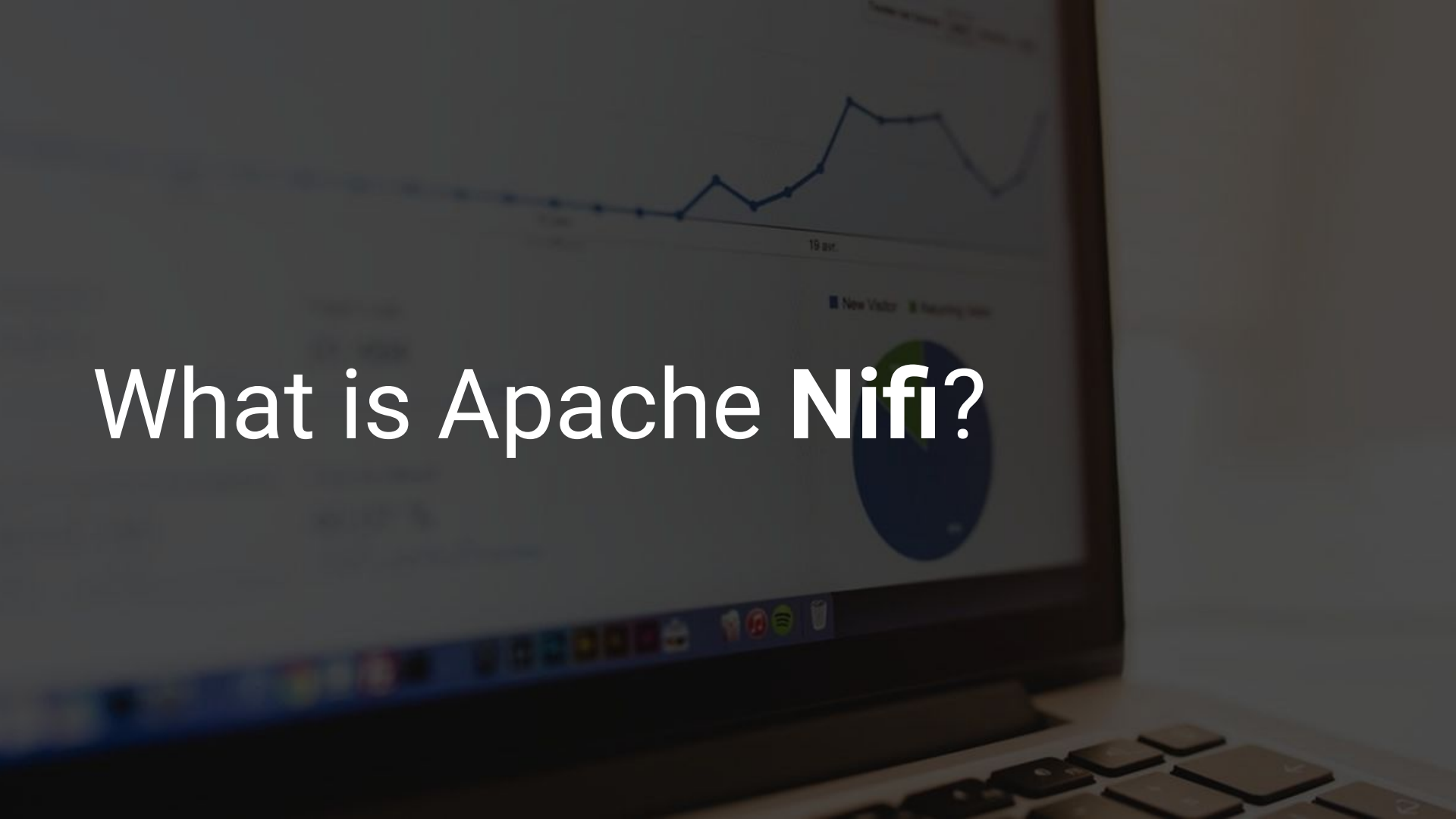

nifi



What is Apache **Nifi**?

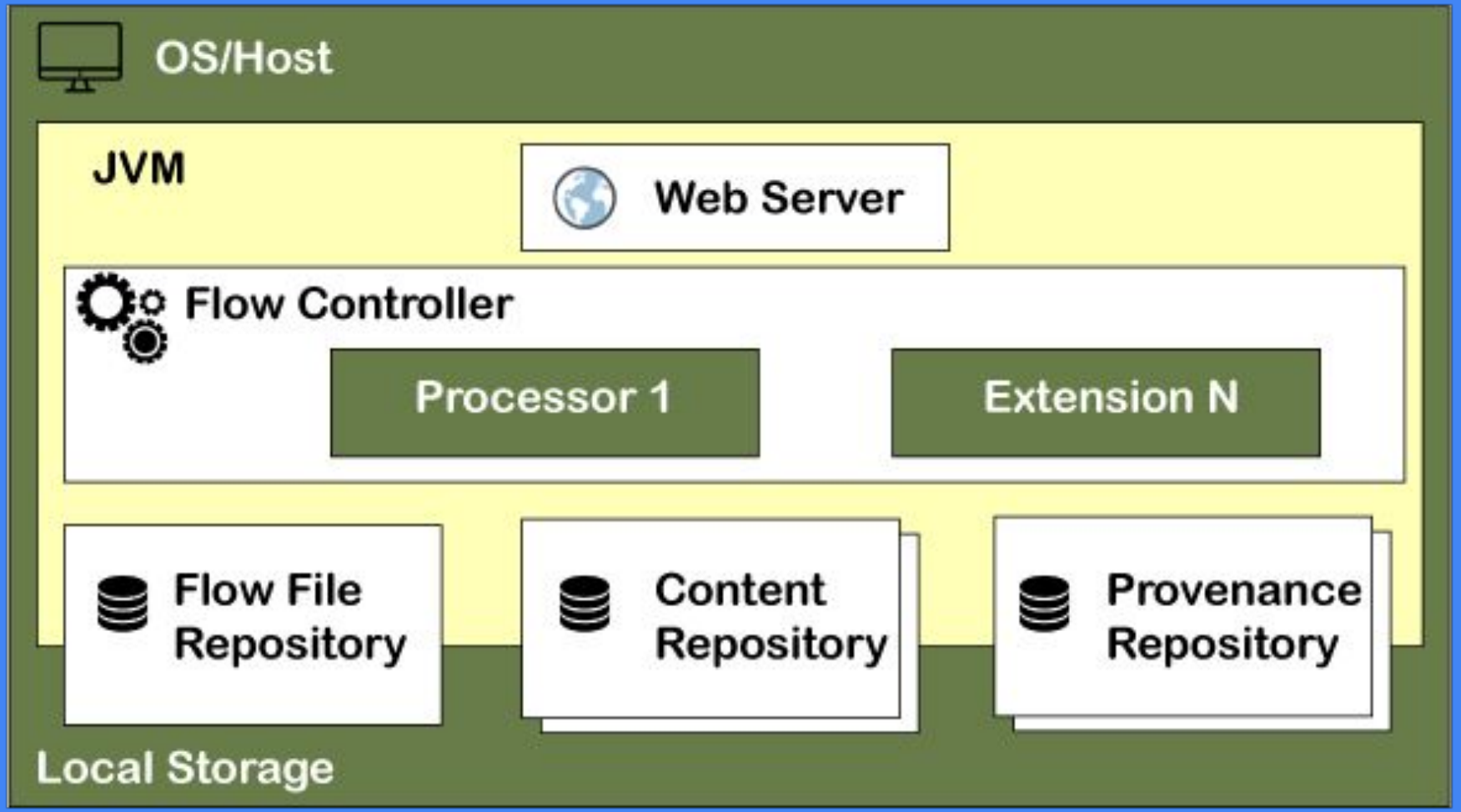
Apache NiFi is an integrated data logistics platform for automating the movement of data between disparate systems.

It provides real-time control that makes it easy to manage the movement of data between any source and any destination.

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.


OS/Host
JVM
Web Server
Flow Controller
Processor 1
Extension N
Flow File Repository
Content Repository
Provenance Repository
Local Storage

Nifi Components

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

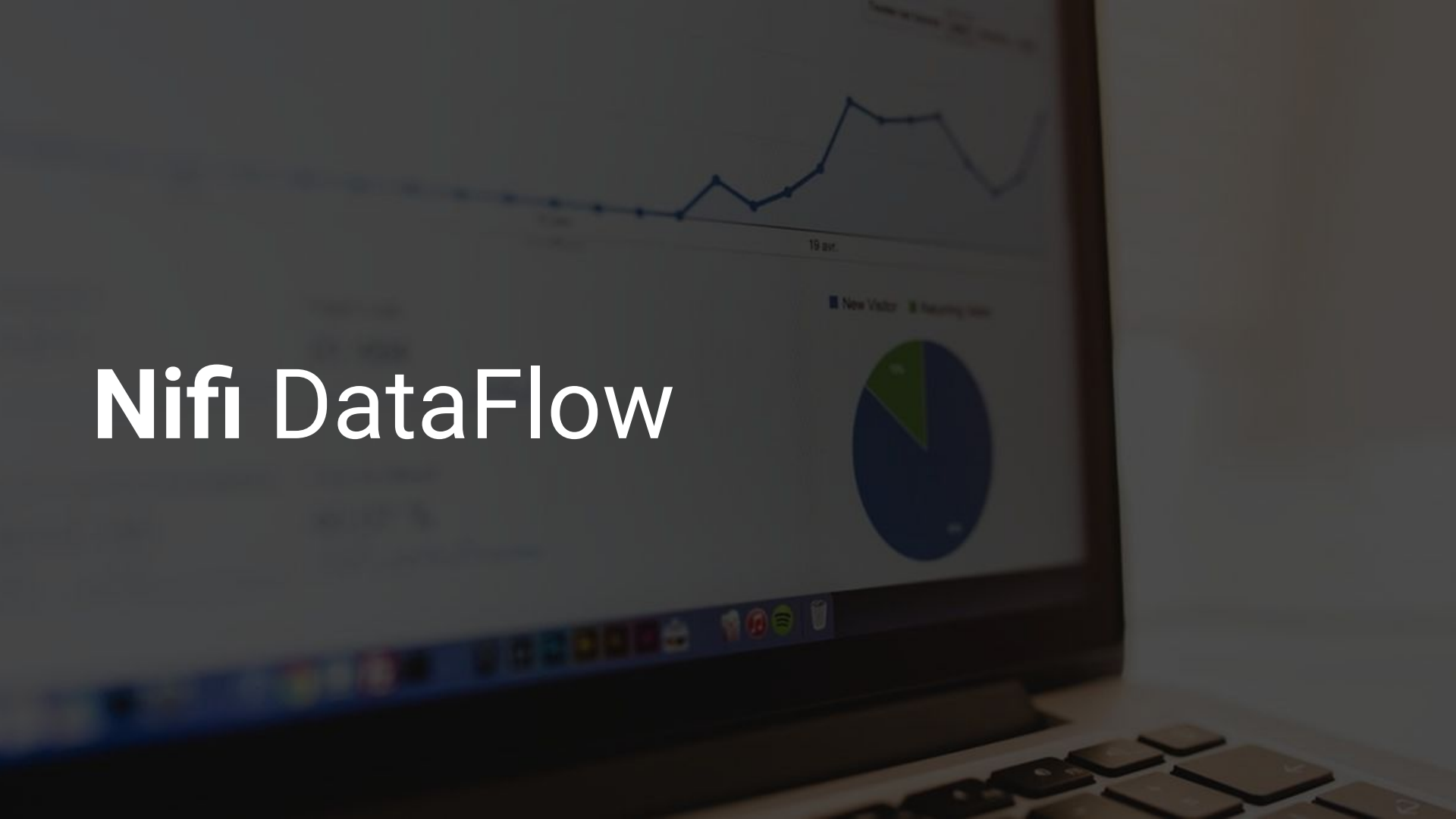
The Content Repository is an area where the actual content bytes of a given FlowFile exist.

Provenance Repository

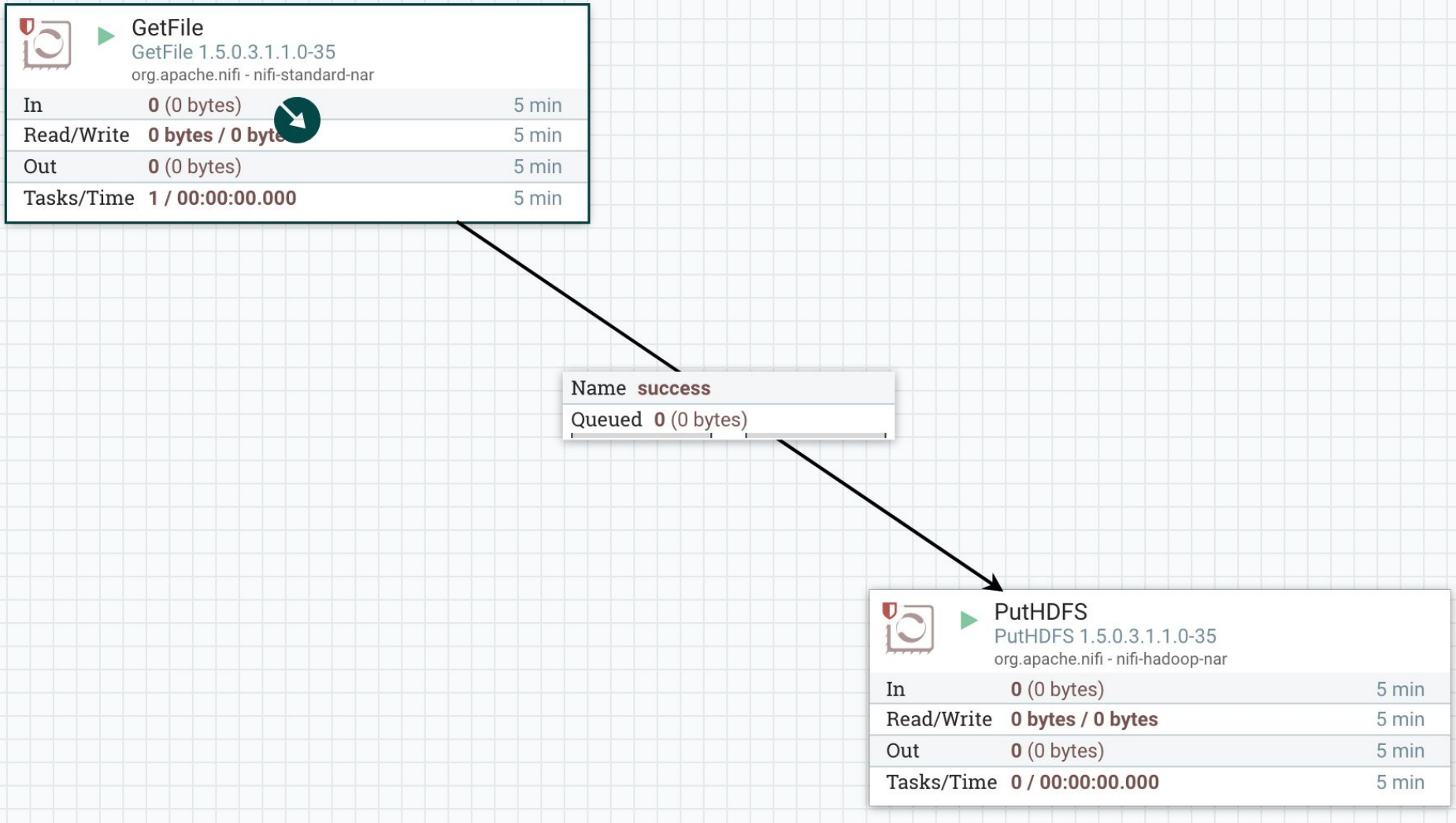
The Provenance Repository is an area where all provenance event data is gathered.

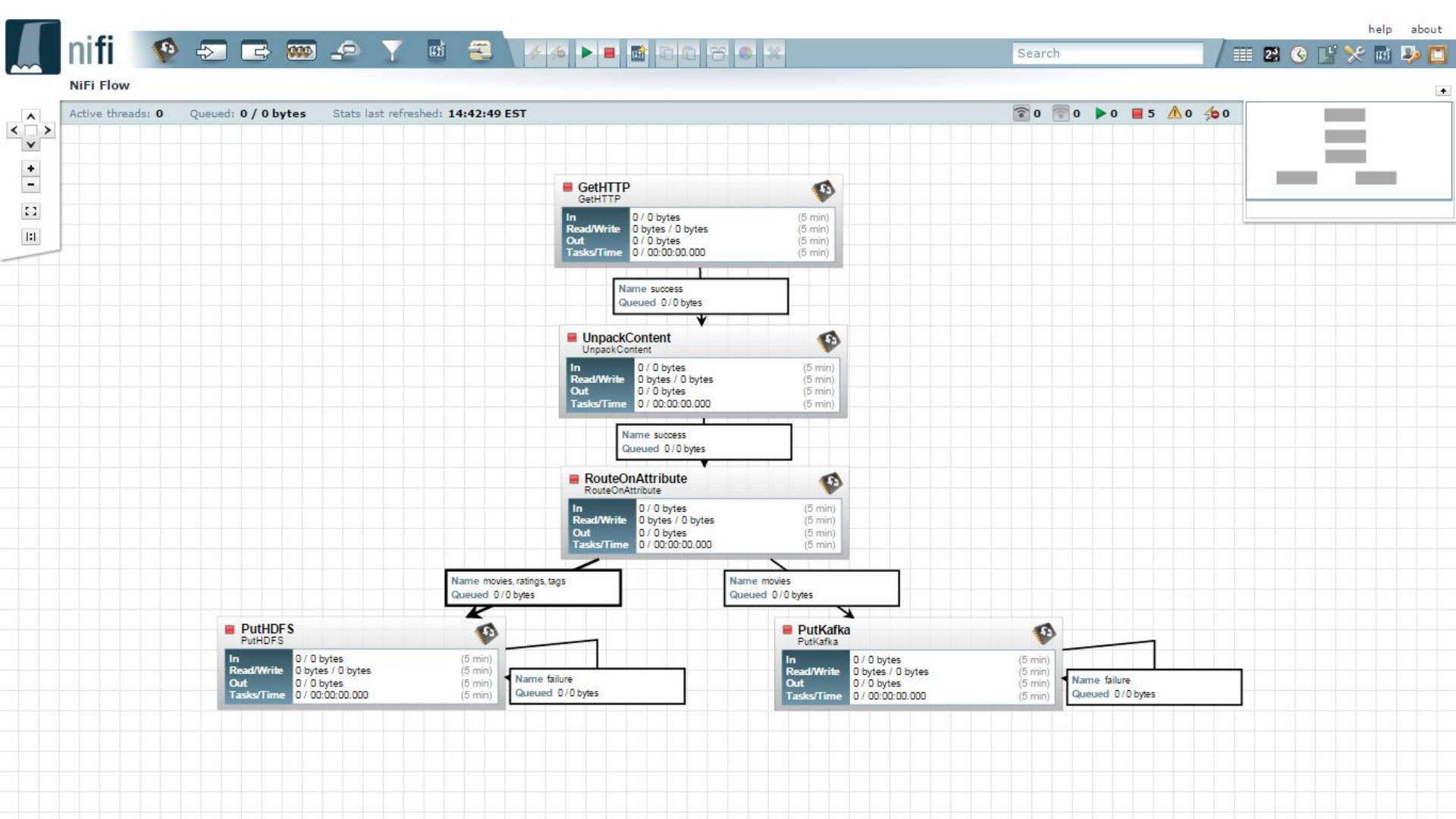
Nifi Features

- 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



Nifi DataFlow


GetFile
GetFile 1.5.0.3.1.1.0-35
org.apache.nifi - nifi-standard-nar
In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 1 / 00:00:00.000 5 min
Name success
Queued 0 (0 bytes)
PutHDFS
PutHDFS 1.5.0.3.1.1.0-35
org.apache.nifi - nifi-hadoop-nar
In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 00:00:00.000 5 min


nifi
help
about
Search
NiFi Flow
Active threads: 0
Queued: 0 / 0 bytes
Stats last refreshed: 14:42:49 EST
0
0
0
5
0
0
GetHTTP
GetHTTP
In 0 / 0 bytes (5 min)
Read/Write 0 bytes / 0 bytes (5 min)
Out 0 / 0 bytes (5 min)
Tasks/Time 0 / 00:00:00.000 (5 min)
Name success
Queued 0 / 0 bytes
UnpackContent
UnpackContent
In 0 / 0 bytes (5 min)
Read/Write 0 bytes / 0 bytes (5 min)
Out 0 / 0 bytes (5 min)
Tasks/Time 0 / 00:00:00.000 (5 min)
Name success
Queued 0 / 0 bytes
RouteOnAttribute
RouteOnAttribute
In 0 / 0 bytes (5 min)
Read/Write 0 bytes / 0 bytes (5 min)
Out 0 / 0 bytes (5 min)
Tasks/Time 0 / 00:00:00.000 (5 min)
Name movies, ratings, tags
Queued 0 / 0 bytes
Name movies
Queued 0 / 0 bytes
PutHDFS
PutHDFS
In 0 / 0 bytes (5 min)
Read/Write 0 bytes / 0 bytes (5 min)
Out 0 / 0 bytes (5 min)
Tasks/Time 0 / 00:00:00.000 (5 min)
Name failure
Queued 0 / 0 bytes
PutKafka
PutKafka
In 0 / 0 bytes (5 min)
Read/Write 0 bytes / 0 bytes (5 min)
Out 0 / 0 bytes (5 min)
Tasks/Time 0 / 00:00:00.000 (5 min)
Name failure
Queued 0 / 0 bytes