NiFi can be downloaded from the NiFi Downloads Page.

https://mirrors.gigenet.com/apache/nifi/1.12.1/nifi-1.12.1-bin.zip

Decompress and untar into the desired installation directory.

Navigate to the NiFi installation directory.

Then run the command:

$ sudo bin/nifi.sh start

By following the command you can track Apache NiFi is started or not:

$ sudo bin/nifi.sh status

sudo vi /opt/nifi/conf/bootstrap.conf

# The defaults.

# JVM memory settings

java.arg.2=-Xms512m

java.arg.3=-Xmx512m

# If you wanted to change it to 1 GB each:

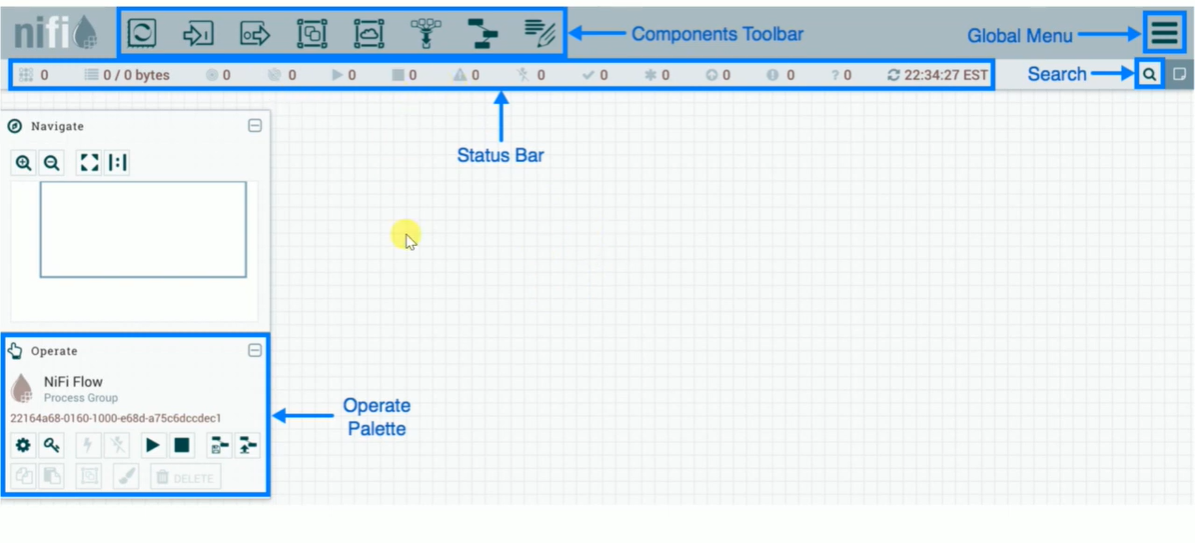
# JVM memory settings

java.arg.2=-Xms1g

java.arg.3=-Xmx1g

sudo ./nifi.sh start

http://localhost:8080/nifi/



NiFi Components

Internally, NiFi pipeline consists of below components.

FlowFile

FlowFile represents the real abstraction that NiFi provides i.e., the structured or unstructured data that is processed.

Semi Structured data such as JSON or XML message and unstructured data such as images, videos, audios. FlowFile contains two parts – content and attribute.

Content keeps the actual information of the data flow which can be read by using GetFile, GetHTTP etc.

while the attribute is in the key-value pair form and contains all the basic information about the content.

Processor

Processor acts as a building block of NiFi data flow.

It performs various tasks such as create FlowFiles, read FlowFile contents, write FlowFile contents, route data, extract data, modify data and many more.

As of today we have 280+ in built processors in NiFi. Do remember we can also build custom processors in NiFi as per our requirement.

Reporting Task

Reporting task is able to analyse and monitor the internal information of NiFi and then sends this information to the external resources.

Processor Group

It is a set of various processors and their connections that can be connected through its ports.

Queue

Queue as the name suggests it holds processed data from a processor after it’s processed.

FlowFile Prioritizer

It gives the facility to prioritize the data that means the data needed urgently is sent first by the user and remaining data is in the queue.

Flow Controller

Flow Controller acts as the brain of operations. It keeps the track of flow of data that means initialization of flow, creation of components in the flow,

coordination between the components.

It is responsible for managing the threads and allocations that all the processes use. Flow controller has two major components- Processors and Extensions.

Ex:1

Generate Flow File(File Size:1KB, Sec:5secs) -> Putfile(Directory:/tmp, Settings: Check-in- Failure & Success)

Ex:2

GetFile(Properties:location(out)) -> Putfile(Properties:location(in))

https://www.batchiq.com/s3-ingest-with-nifi.html

https://www.batchiq.com/database-ingest-with-nifi.html

https://github.com/jfrazee/awesome-nifi

[https://github.com/hortonworks-gallery/nifi-templates/tree/master/templates](https://github.com/jfrazee/awesome-nifihttps://github.com/hortonworks-gallery/nifi-templates/tree/master/templates)

https://thirdeyedata.io/etl-pipeline-demonstration-using-apache-nifi/

<https://medium.com/@suci/hello-world-nifi-dcafcba0fdb0>

Nifi task  
----------  
1. FTP a file from remote server to your local server.  
2. Move it to HDFS  
3. Move this data from HDFS to mysql/postgres/mssql/hive external database