Assignment-12.3:

***What is meant by Flume NG?***

 Flume NG - refactoring of Flume.

It is mainly used to solve certain issues as well limitations

And there is a need for certain types of refactoring certain type of core class.

And these are the following issues:

* Complexity in code.
* Source or sink can be started or stopped at any time in a core component lifecycle as well control-code.
* It will help to simplify the data paths which are common.
* Packages which are available can be renamed as org.apache.flume

***Can Flume provides 100 % reliability to the data flow?***

Reliability of flow is possible.

Transactional repository which provides repository to the channels encapsulates the sources and sink.

In flume the events are transferred in an order so there are no chances for the data to get lost.

Whenever a data is received an acknowledgement will be sent.

Yes, it provide end-to-end reliability of the flow. Sources and sinks encapsulated in a transactional repository provides by the channels.

***Can Flume can distributes data to multiple destinations?***

It is possible for flume to distribute the data to multiple destinations.

And even multiplexing flow is supportable. The flow of event will be from the source to various channels and from there to multiplexers.

And there is a possibility for declaration of more than one source or channel or sink.

So it could be possible from numerous number of sources and buffer from numerous number of channels which will be forwarded to multiple number of sinks.

For this process the sink should follow the respective configuration as well sinks should be properly allocated.

For Example:

For Setting up of sink channels:

agent\_foo.sinks.hdfs-ca-sink.channel = mem-channel-ca

agent\_foo.sinks.hdfs-ny-sink.channel = mem-channel-ny

***Explain about the different channel types in Flume. And which channel type is faster?***

Channels will be present between source and sink.

And it will operate at different rates.

And they act as key which guarantee that there will not be any data loss.

So that the data can be read by one or more number of sinks.

The different channel available in sink are:

1. **Memory Channel**

It is a memory channel which stores the events onto heap.

A high throughput will be supported which helps to store a huge data in memory.

Memory channel acts as in-memory queue.

It acts as a source to read the tail of sink of its head.

All the events will be stored up to maximum size and the flow will be from higher throughput.

And the data will be lost if there is any agent failure.

* To the maximum of 100 channels can be stored in a event.
* To the maximum of 100-transaction can happen between source to sink.
* It will take 3 seconds to add or to remove a data.

1. **File Channel:**

It acts as persistent channel for flume.

All the details which all need to be written to the disc will be present and the data will not be lost while processing or even during the time of machine shut down or crash.

The work of the channel to ensure all the data which all entered are removed only after they are taken up by the sink and after the process of transaction.

When there is a chance for the machine to get crashed then the process will restart.

It was specially designed to take more tasks (i.e0several number of sinks and events.

* 100-transaction is the maximum size
* 30000 milliseconds is the time taken between check points.
* 1000000 is the maximum channel capacity.
* 3 seconds is the time taken for put command to take through.