**ASSIGNMENT-9.3**

**1.What is meant by FlumeNG ?**

* Flume can be used to produce streams of data that can be aggregated, stored, and analyzed using Hadoop.
* Flume NG is work related to new major revision of Flume .
* At a high-level, Flume NG uses a single-hop message delivery guarantee semantics to provide end-to-end reliability for the system.
* Flume-ng is a command that is used to start the flume agent given as,

flume-ng agent \ --conf-file \ --name \ --conf $FLUME\_HOME/conf \

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

* Yes, flume provides 100 % reliability to the data flow.
* By default,flume uses a transactional approach in the data flow which means the source and sink encapsulate in a transactional repository provided by the channels.
* This channels responsible to pass reliably from end to end flow which strengthens the reliability to data flow.

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

* Yes,flume can distribute data to multiple destinations as it supports multiplexing flow.
* The event flows from one source to multiple channel and multiple destinations.
* This can be acheived by defining a flow multiplexer.

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

The different channel types available in Flume are,

* **MEMORY Channel -** Events are read from the source into memory and passed to the sink.
* **JDBC Channel** - JDBC Channel stores the events in an embedded Derby database.
* **FILE Channel** - File Channel writes the contents to a file on the file system after reading the event from a source. The file is deleted only  after the contents are successfully delivered to the sink.
* **Kafka channel** - can be used to write to Hadoop directly from Kafka without using a source,to write to Kafka directly from Flume sources without additional buffering,as a reliable and highly available channel for any source/sink combination.

**MEMORY Channel** is the fastest channel among the three however has the risk of data loss. The channel that you choose completely depends on the nature of the big data application and the value of each event.