**ASS-12.3**

**Explain in detail.**

**What is meant by FlumeNG ?**

At a high-level, Flume NG uses a single-hop message delivery guarantee semantics to provide end-to-end reliability for the system.

The purpose of Flume is to provide a distributed, reliable, and available system for efficiently collecting, aggregating and moving large amounts of log data from many different sources to a centralized data store. The architecture of Flume NG is based on a few concepts that together help achieve this objective.

It achieves reliability by the following mechanism

This mechanism also forms the basis for failure handling in Flume NG. When a flow that passes through many different agents encounters a communication failure on any leg of the flow, the affected events start getting buffered at the last unaffected agent in the flow. If the failure is not resolved on time, this may lead to the failure of the last unaffected agent, which then would force the agent before it to start buffering the events. Eventually if the failure occurs when the client transmits the event to its first-hop destination, the failure will be reported back to the client which can then allow the application generating the events to take appropriate action.

On the other hand, if the failure is resolved before the first-hop agent fails, the buffered events in various agents downstream will start draining towards their destination. Eventually the flow will be restored to its original characteristic throughput levels. Figure 4 below illustrates a scenario where a flow comprising of two intermediary agents between the client and the central store go through a transient failure. The failure occurs between agent 2 and the central store, resulting in the events getting buffered at the agent 2 itself.

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

Yes it provides 100% reliability since it provides transactional data flow by default where the source and sink are encapsulated by the repositories provided by the channels. Thus the channels maintain high reliability in data flow.

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

Yes,it supports distributing data to multiple destination or multiple sinks by defining flow multiplexer

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

The 3 different built in 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.  
>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.