Assignment 9.5

1. Explain the core components of flume

The core components of flume are channel, sink and source

* **Source**– the entity through which data enters into Flume. Sources either actively poll for data or passively wait for data to be delivered to them. A variety of sources allow data to be collected, such as log4j logs and syslogs.
* **Sink**– the entity that delivers the data to the destination. A variety of sinks allow data to be streamed to a range of destinations. One example is the HDFS sink that writes events to HDFS.
* **Channel**– the conduit between the Source and the Sink. Sources ingest events into the channel and the sinks drain the channel.

2. Can flume provide 100% reliability while transferring the data?

* Yes Flume provides 100% reliability because of its transactional approach in data flow.
* Flume can guarantee that all data received by an agent node will eventually make it to the collector at the end of its flow as long as the agent node keeps running.
* Even when using the end-to-end reliability level, a Flume flow can fail to make progress if there aren’t any nodes available to process events.

3. Explain the consolidation in flume.

* It collect data from different sources and from different flume agents.
* Flume source can collect all data flow from different sources and flows through channel and sink and send this data to HDFS or target destination.
* This is called as consolidation in flume.

4. Explain what is an event in flume?

* A unit of data with set of string attributes called flume event
* The external source like web-server sends events to the source.
* Flume has internal built in functions to understand source format.

5. Explain what is an agent in flume.

* An agent is an independent daemon process (JVM) in Flume.
* It receives the data (events) from clients or other agents and forwards it to its next destination (sink or agent).
* Flume may have more than one agent.