ASSIGNMENT 9.3

1. WHAT IS MEANT BY FLUMENG?

* Apache Flume is a distributed, reliable, available service for efficiently collecting, aggregating, and moving large amounts of streaming data into the HDFS
* 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 \

1. CAN FLUME PROVIDES 100 % RELIABILITY TO THE DATA FLOW?

* Yes it provide end to end reliability of the flow.
* By default flume uses a transactional approach in the data flow which means sources and sinks encapsulated in a transactional repository provides by the channels.
* This channels responsible to pass reliably from end to end flow which strengthens the reliability to data flow

1. CAN FLUME DISTRIBUTES DATA TO MULTIPLE DESTINATIONS?

* Yes Flume can distribute data to multiple destinations as it supports multiplexing flow
* Here the event flows from single source through multiple channels to multiple destinations
* Flow multiplexer is defined to achieve this multiplexing flow

1. 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 among all the other channels. But disadvantage of this channel is it suffers from data loss