Lesson 9 – Apache Sqoop and Flume

Knowledge is gained from inside and outside

WHOLENESS OF THE LESSON

Sometimes there is a need to ingest data from outer sources into HDFS. Sqoop and Flume are two data integration frameworks that are heavily used for ingesting data into HDFS. This HDFS data can then be analyzed to get a better insight. Science & Technology of Consciousness: Nature of life is to grow. Similarly, we want more data into HDFS to gain better insights and grow in our understanding of the data.

MAIN POINTS

1. Apache Sqoop is an open source, batch-oriented program developed by Cloudera that allows users to import data from a database into HDFS or export data from HDFS back to a database. To use Sqoop, you specify the tool you want to use and the arguments that control the tool. Sqoop uses MapReduce to import and export the data, which provides parallel operation as well as fault tolerance. Science & Technology of Consciousness: TM is a tool which imports all the Nature support in our lives and exports the creative potential within into the atmosphere, thereby creating a blissful surrounding.

2. Flume is a service that lets you ingest high-volume streaming data into HDFS. To use Flume, we need to run a Flume agent, which is a long-lived Java process that runs sources and sinks, connected by transactional channels through which events flow from an external source to the next destination. Science & Technology of Consciousness: Maharishi's Science of Consciousness locates three components to any kind of knowledge: the knower, the object of knowledge, and the process of knowing. These can be found in the structure of a Flume Agent: the data that is coming into the Source is the "object of knowledge." The "knower" aspect of the agent is the Sink which is the destination of the data and the Channels are like the "process of knowing".

UNITY CHART

CONNECTING THE PARTS OF KNOWLEDGE WITH THE WHOLENESS OF KNOWLEDGE:

Ingesting Data into HDFS

1. Sqoop is based upon a connector architecture which supports plugins to provide connectivity to new external systems. An example use case of Sqoop is an enterprise that runs a nightly Sqoop import to load the day's data from a production transactional RDBMS into a Hive data warehouse for further analysis.

2. Flume has a flexible design based upon streaming data flows. It is fault tolerant and robust with multiple failovers and recovery mechanisms.

3. <u>Transcendental consciousness</u>: is the state where the mind has become an unlimited and silent ocean of pure consciousness without a single wave of thought.

4. <u>Impulses within the Transcendental Field</u>: Nature accomplishes what it needs by having its impulses in the transcendental field be as efficient as possible.

5. <u>Wholeness moving within itself</u>: In Unity Consciousness, all things are experienced in terms of one's own unbounded nature, the Self.