**Big Data Ingestion Tools: Flume vs Kafka vs Nifi**

~ Written by: Deward Seneh

**Overview**

Big data as the name goes is a term that describes large amount of data. Nowadays, huge amount of data is being generated by almost every device every second of the day. To overcome the problem of handling these large volumes of data that are being generated, Hadoop was developed. Hadoop is an open source framework that is used to store, process, and analyze data in a distributed fashion. Hadoop provides a set of tools, methodologies, and technologies to capture, store and analyze data in cluster. This process can be referred to as Big Data Pipeline.

There are many stages in the big data pipeline, here are the six important ones: Data Source, Ingest, Process, Story and Analyze. For the objective of this blog, I am going to focus on the Ingestion stage in particular.

**Data Ingestion**

In the pipeline, the first thing to do is have a source of the data, we need to obtain the data from those source or sources we have into our pipeline. This process of obtaining the data is call ingestion. It is also the way of importing data for immediate use or for the purpose of storage. There are many tools used for data ingestion, but for this blog, I will be brefely discussing three of them: *Apache Flume, Apache Kafka, and Apache Nifi*.

**Apache Flume**

Apache Flume is a tool, service, or data ingestion mechanism for collecting aggregating and transporting large amounts of streaming data such as log files, events, etc. form various centralized data sources.

Apache Flume is highly fault-tolerant, which mean it offers reliable mechanism for fail-over and recovery. Flume is also capable of collecting data in both streaming (real-time) and bash models.

**Apache Kafka**

Apache Kafka is an open source system for processing ingests data in real-time. Kafka is a fast, salable, fault-tolerant, publish-subscribe messaging system. One of the best futures of Kafka is, it is highly available and resilient to node failures and supports automatic recovery. This feature makes Kafka ideal for communication and integration between components of large-scale data systems in real-world data systems.

**Apache Nifi**

Apache Nifi provides us a powerful and flexible management tool for data ingestion and data flow. One key feature of Apache Nifi is, it provides us with a nice graphical user interface which allows user to create, monitor and control data flow with drag and drop.

Apache Nifi is a real time data ingestion tool which can transfer and manage data transfer between different sources and destination system.

**All Together: Flume vs Kafka vs Nifi**

Flume, Kafka and Nifi are great tools that offer great performances and can be scaled horizontally.

The function of most of these tools are almost the same, therefore, combining them may appear wasteful, as it seems to introduce some overlap in functionality. To give you an example, both NiFi and Kafka provide brokers to connect producers and consumers. However, they do so differently: in NiFi, the bulk of the data flow logic lays not inside the producer/consumer, but lives in the broker, allowing for centralized control. NiFi was built to do one important thing well: data flow management. With both tools combined, NiFi can take advantage of Kafka’s reliable stream data storage, while taking care of the dataflow challenges that Kafka was not designed to solve. Combining Flume and Kafka on the other hand allows Kafka to avoid custom coding and take advantage of Flume’s battle-tested sources and sinks, while Flume events passing through the Kafka channel are stored and replicated across Kafka brokers for resiliency.

**Conclusion**

When building big data pipelines, we need to think on how to ingest the volume, variety, and velocity of data showing up at the gates of what would typically be a Hadoop ecosystem. Preliminary considerations such as sociability, reliability, adaptability, cost in terms of development time, etc. will all come into play when deciding on which tools to adopt to meet our requirements.