Part1

1. When was the last time you tried/learned something new? What did you do and how?

Last month , Feb-2018, I tried to integrate real-time Cloudera system with Hortonworks Nifi. On Cloudera system, a single node , multiple Kafka broker system having 4 topics, 2 of which stream to Spark streaming and two to Flume is set up. Flume is sinked to hdfs. Spark streaming consume Kafka topics to be merged with pre-computed views from batch layer.

Kafka topics from Cloudera, are tested to connect to NiFi for ingestion and ETL process and found that configuration through GUI is easy and user-friendly. Various integration , even to the Cloud environment is more of a GUI, though not tried. So on-premise data can be interoperate among private, public, colocate services via Cloud services.

The work I did:

* Download Kafka and Zookeeper and configured them
* Tested Kafka and Zookeeper
* Created topics and verified both producer and consumer, thru data and redirection
* Consume Kafka topics via Flume and sink to Hdfs
* Move data through scripts to different layers of raw, decomposed, model and cleansing areas
* Tested Spark streaming windows analytics functions.
* Data loaded to table through Spark SQL, queried