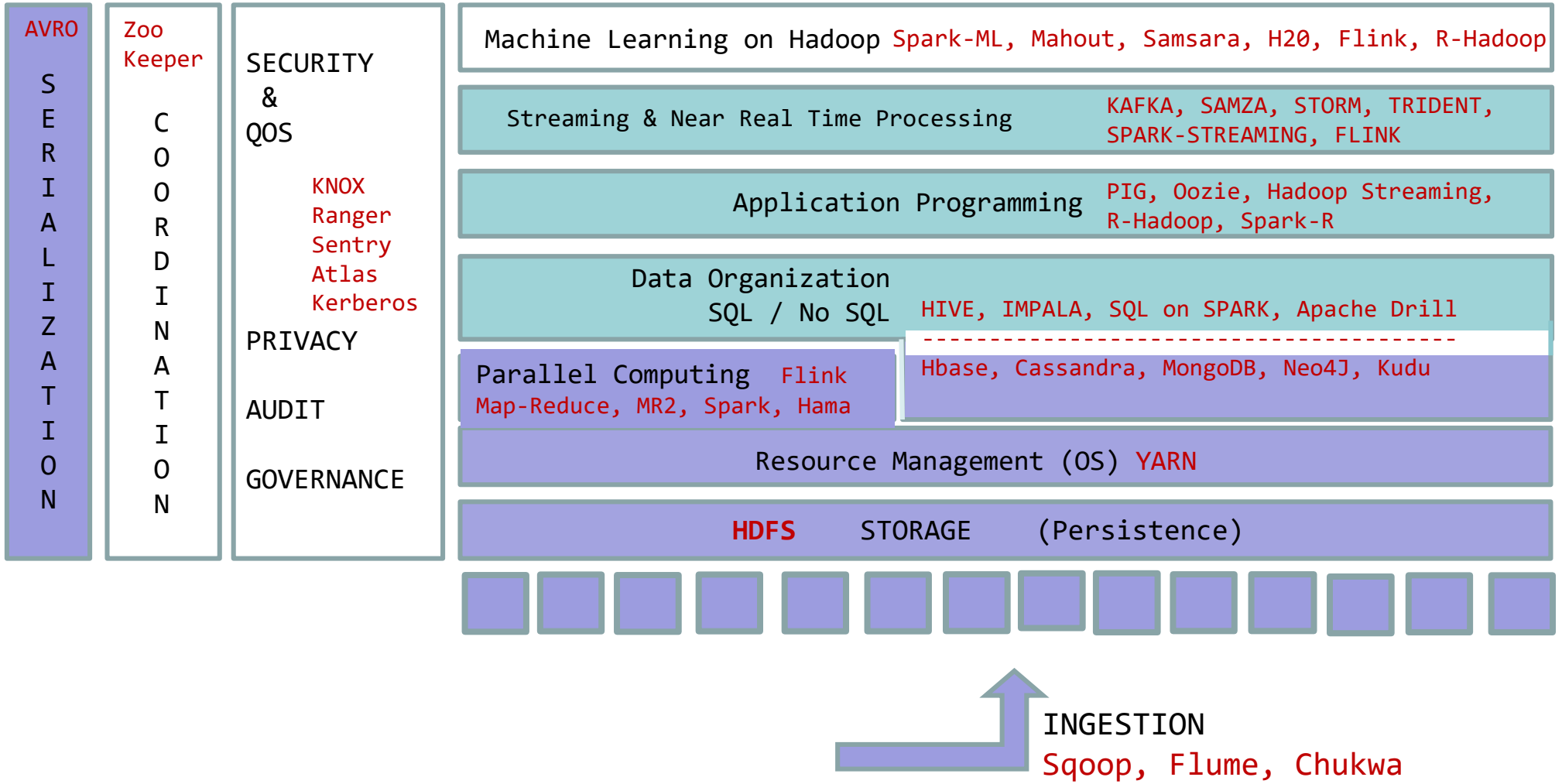
Inspire…Educate…Transform.

Application 3 – Uber App

Flume-Spark-Kafka-SparkStreaming-SparkML

Big Data Team Insofe

Our Focus: The open source big data ("Hadoop + Spark") ecosystem



Uber Use Case

Publicly Available Data

Big Data Processing + Model Building

Analysis

Help Decision Making

Uber Use Case

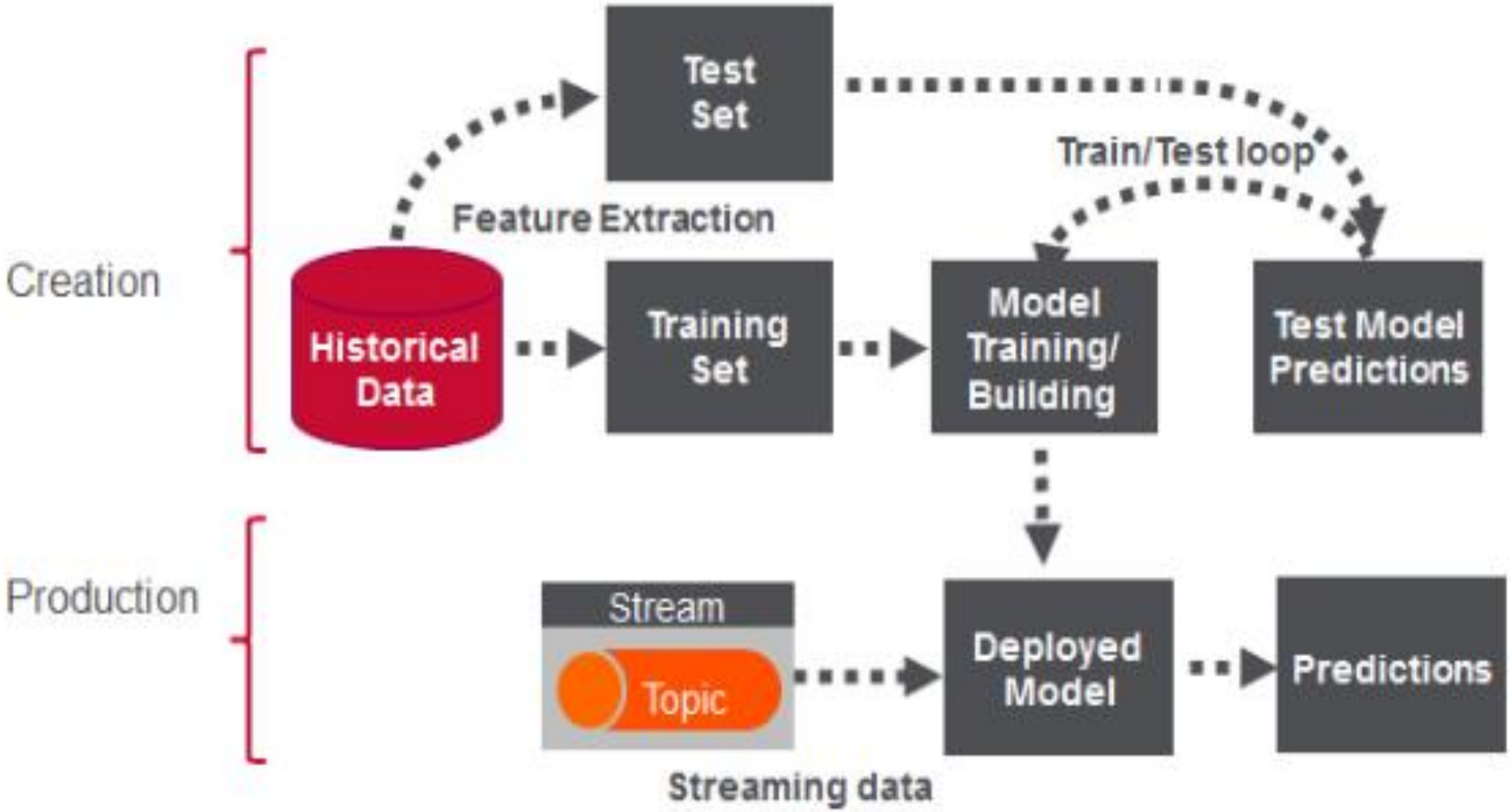
Uber is using big data to perfect its processes, from calculating Uber's pricing, to finding the optimal positioning of cars to maximize profits. In this use-case, we are going to use public Uber trip data to build real-time example for analysis and monitoring of car GPS data.

Other Use cases of data :

1. Observe/Record driving patterns of user using a telematics device
   a. Provide rebates in the sub-sequent purchases/renewals based on the driving history
   b. Get rid of hig-risk customers
   c. Progressive insurance as implemented by many other insurance companies.

2. Track and monitor car's rea-time location – Validate if the car is travelling in the right path or not – Passenger's Safety.

Uber Use Case



Historical Data Structure :

1. Date and Time : Date and Time of Uber pick-up
2. Longitude : Longitude of the uber pick-up
3. Latitude : Latitude of the uber pick-up
4. Base Company : Company affiliated with uber pick-up

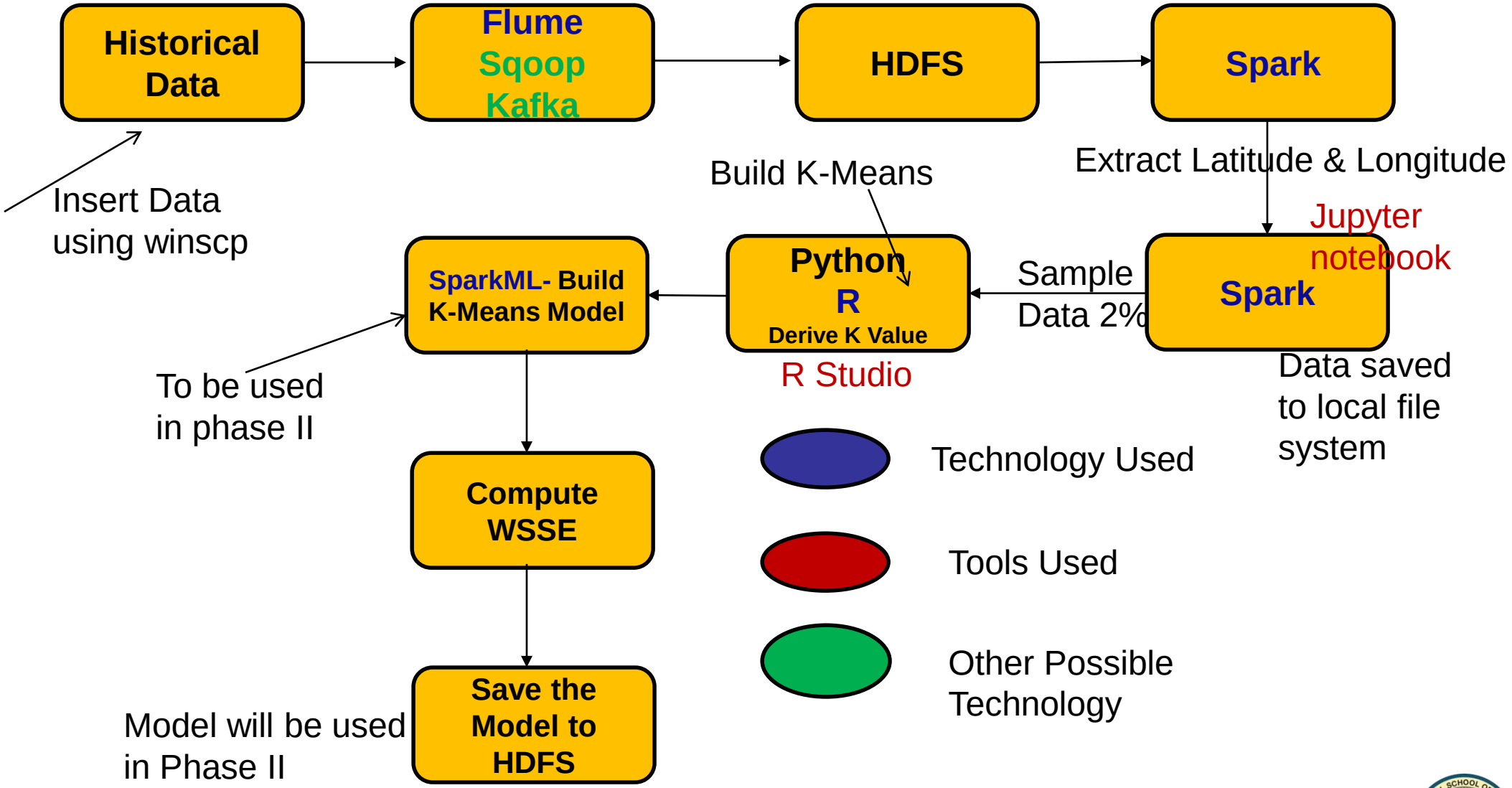
Sample: "11/1/2016 0:00:00",40.7293,-73.992,"B02512"

Streaming Data Structure :

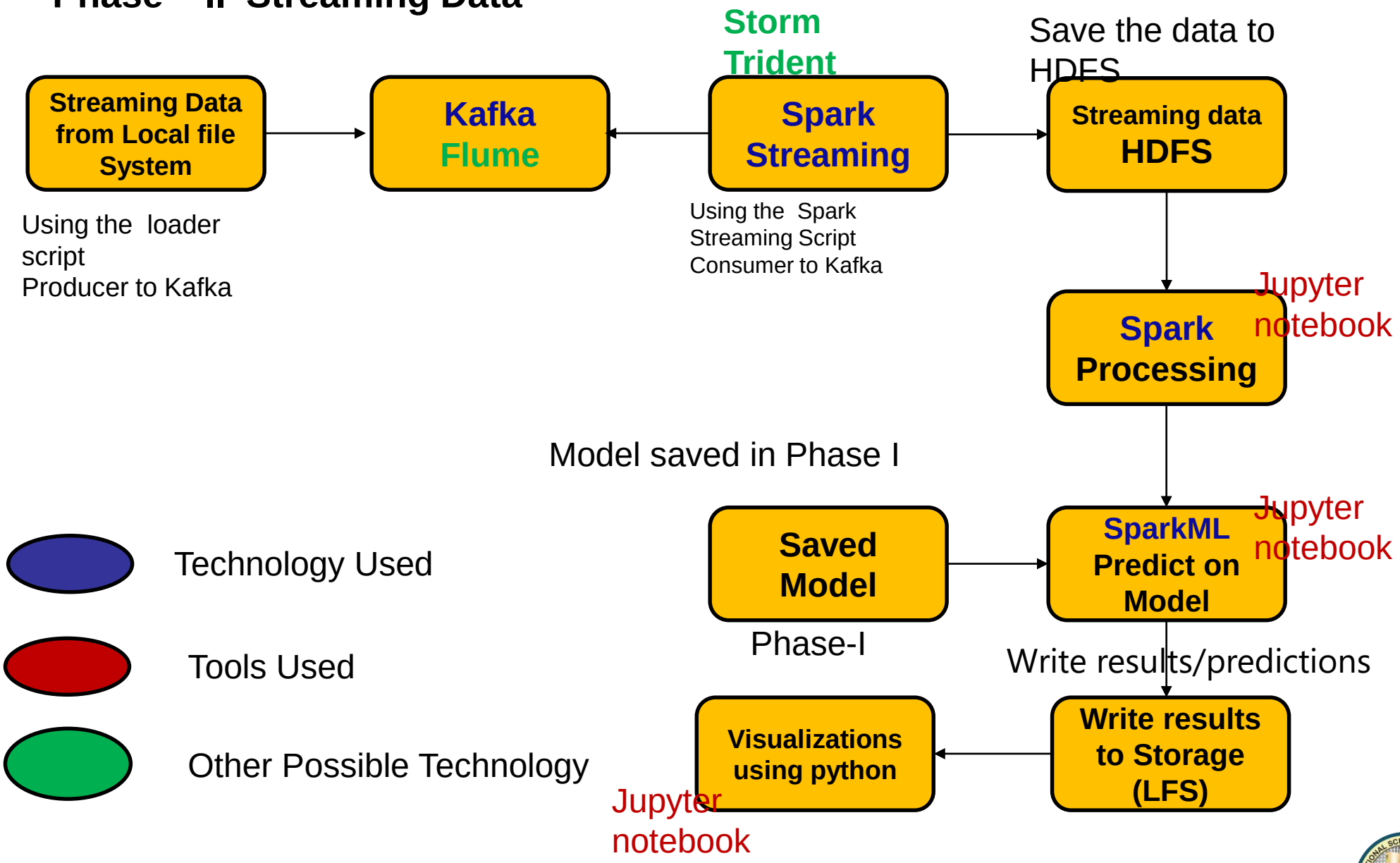
1. Date and Time : Date and Time of Uber pick-up
2. Longitude : Longitude of the uber pick-up
3. Latitude : Latitude of the uber pick-up
4. Base Company : Company affiliated with uber pick-up
5. PickUpStatus : The pick-up status, whether an Active or Cancelled.

Sample: "2/28/2017 0:00:00",40.8039,-73.9678,"B02764","C"

Phase – I Historical Data

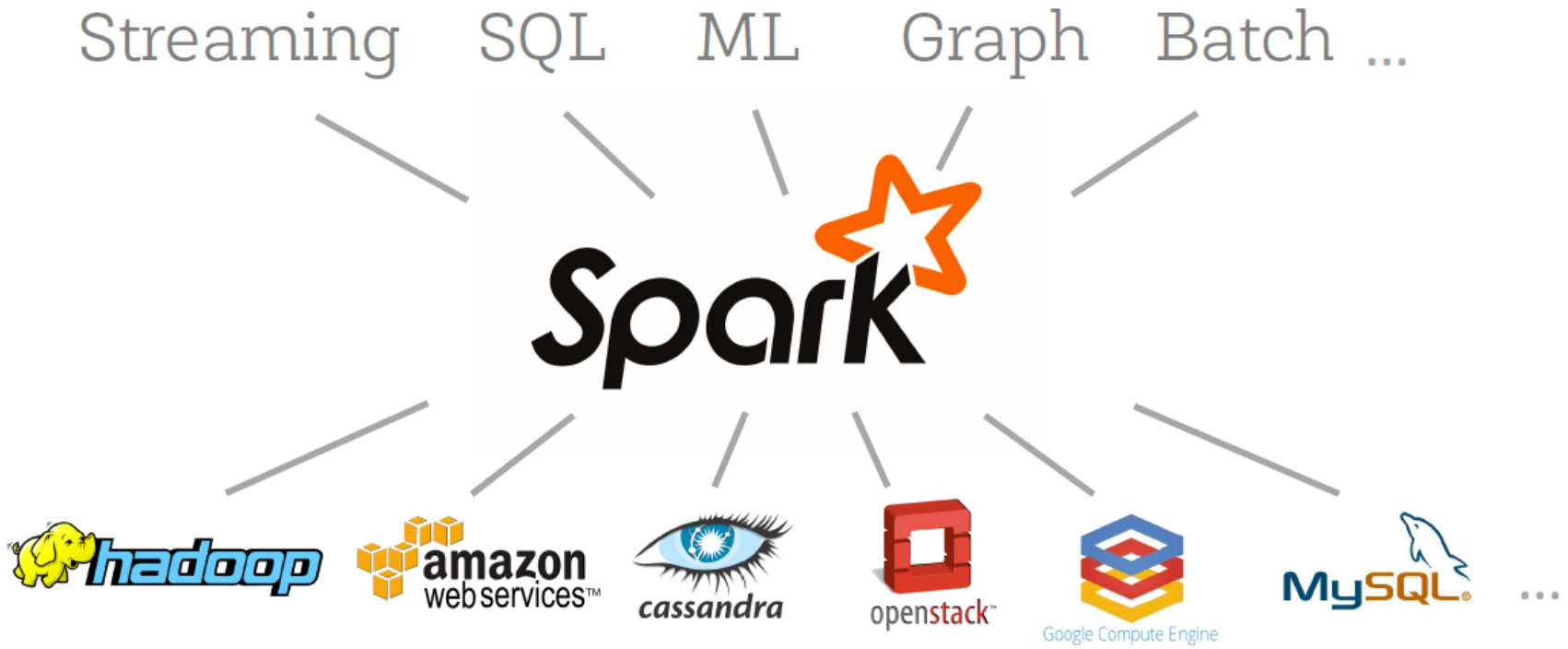


Phase – II Streaming Data

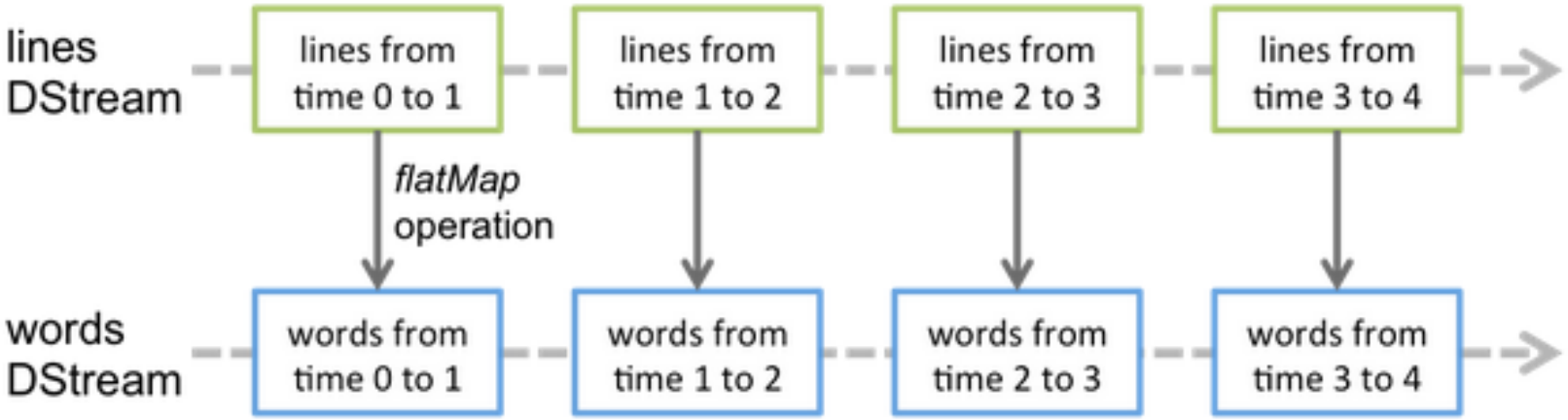


What is Apache Spark?

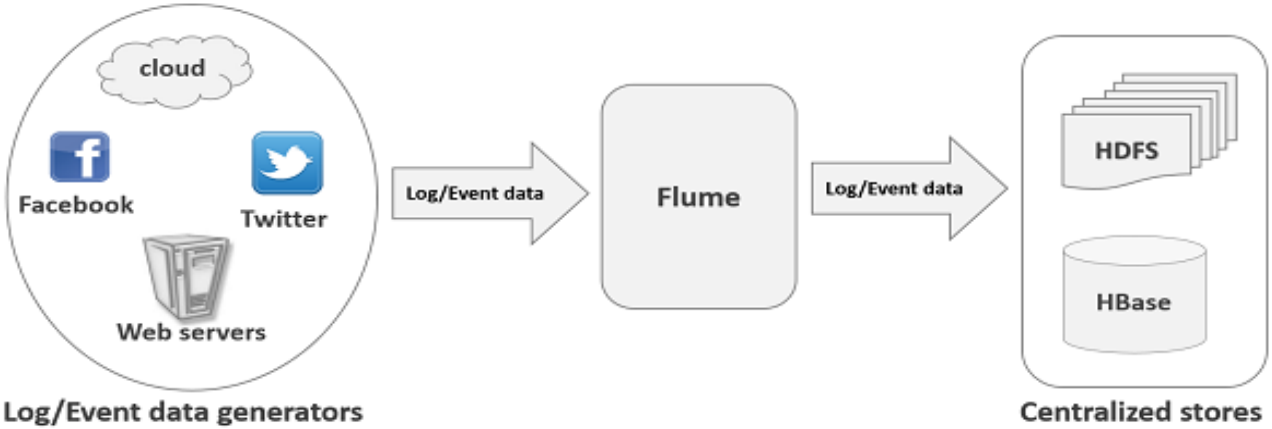
Unified engine across data workloads and platforms

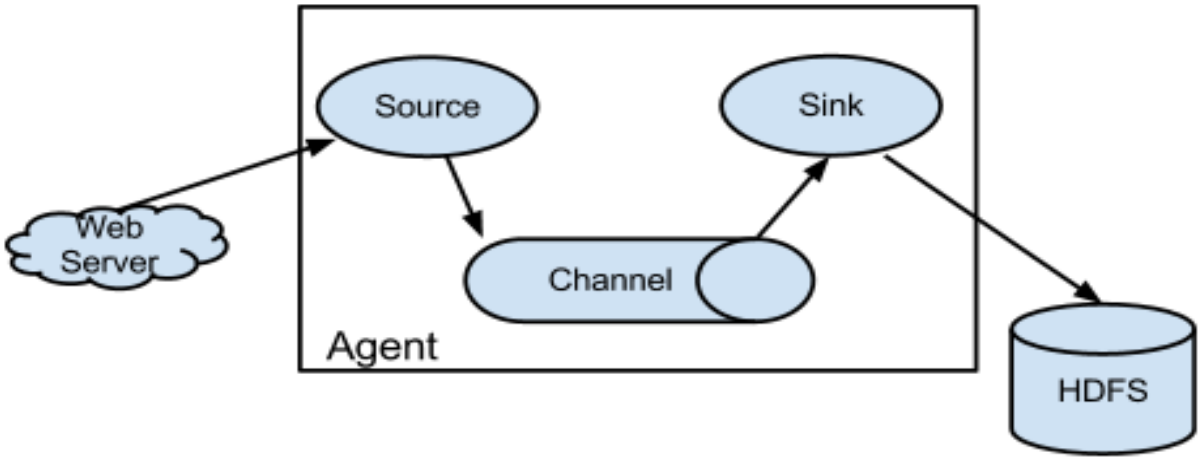


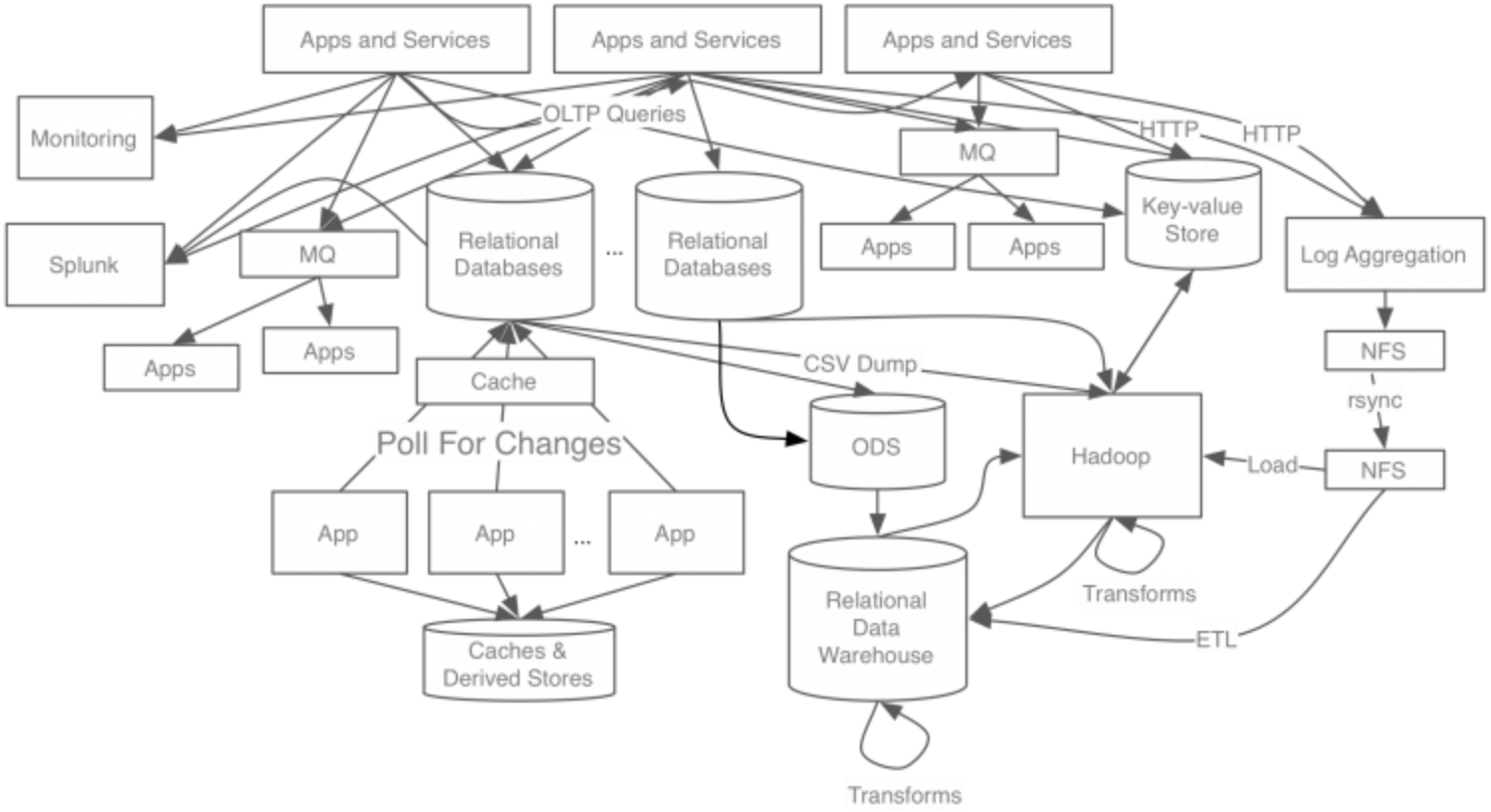
Spark Streaming

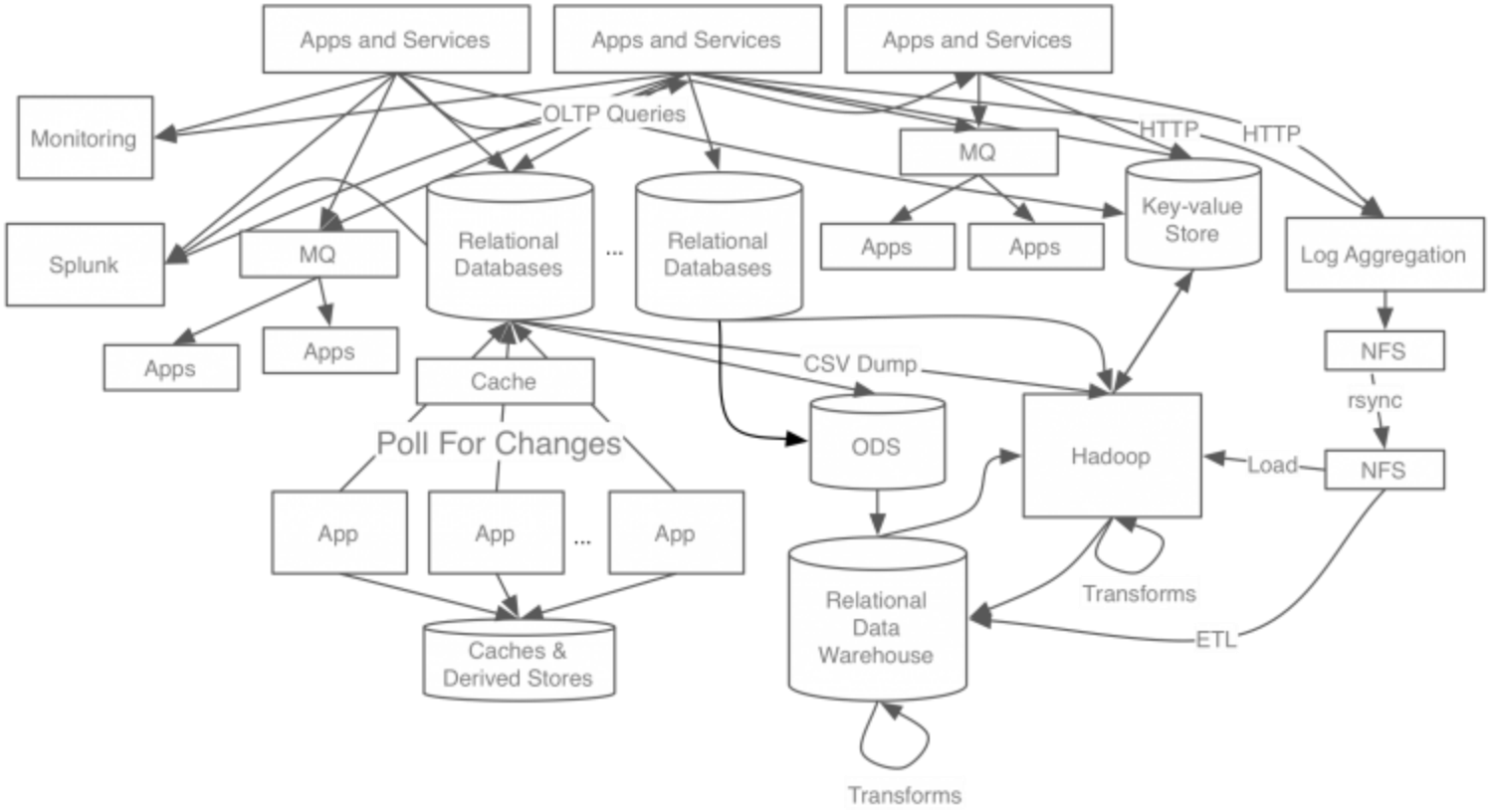


Flume

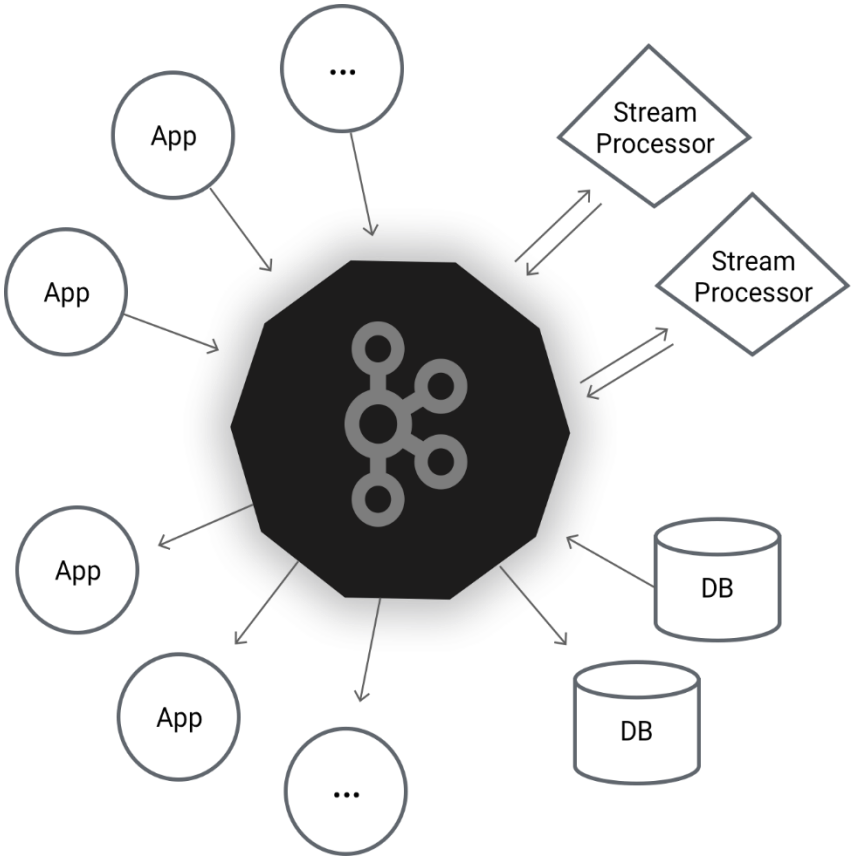

cloud
Facebook
Twitter
Web servers
Log/Event data generators
Log/Event data
Flume
Log/Event data
HDFS
HBase
Centralized stores

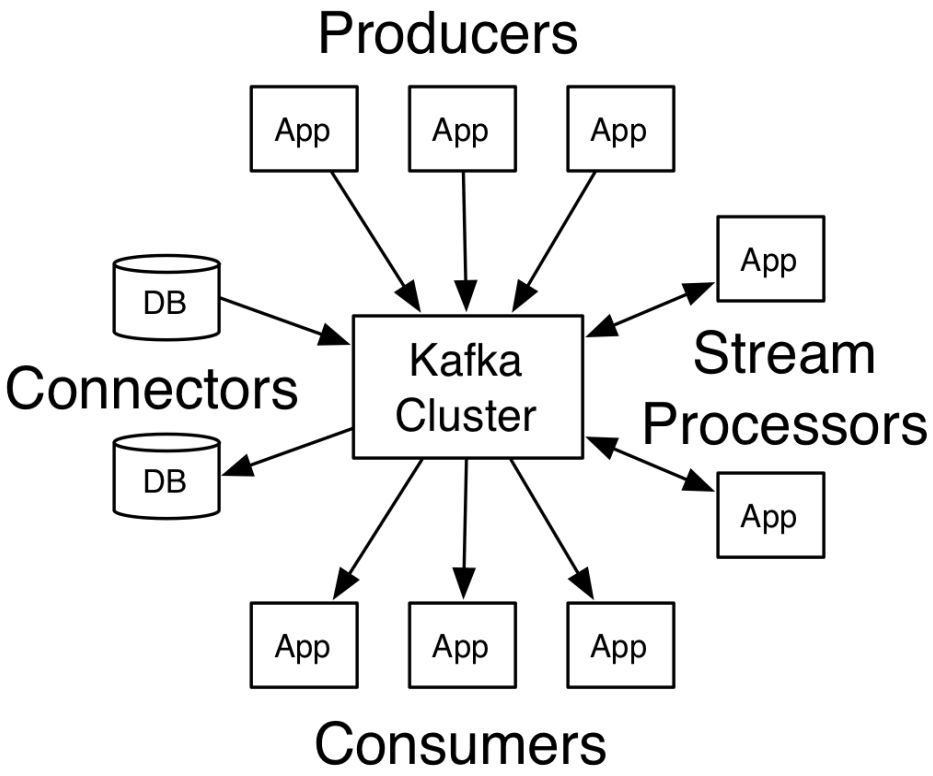

Source
Sink
Web Server
Channel
Agent
HDFS


Apps and Services
Apps and Services
Apps and Services
Monitoring
OLTP Queries
HTTP
HTTP
MQ
Key-value Store
Log Aggregation
Splunk
MQ
Relational Databases
...
Relational Databases
Apps
Apps
Apps
Apps
NFS
rsync
CSV Dump
Cache
ODS
Hadoop
Load
NFS
Poll For Changes
App
App
...
App
Relational Data Warehouse
Transforms
ETL
Caches & Derived Stores
Transforms


Apps and Services
Apps and Services
Apps and Services
Monitoring
OLTP Queries
HTTP
HTTP
MQ
Key-value Store
Log Aggregation
Splunk
MQ
Relational Databases
...
Relational Databases
Apps
Apps
Apps
Apps
NFS
rsync
CSV Dump
Cache
ODS
Hadoop
Load
NFS
Poll For Changes
App
App
...
App
Relational Data Warehouse
Transforms
ETL
Caches & Derived Stores
Transforms

Kafka


...
App
Stream Processor
App
Stream Processor
App
DB
DB
App
...

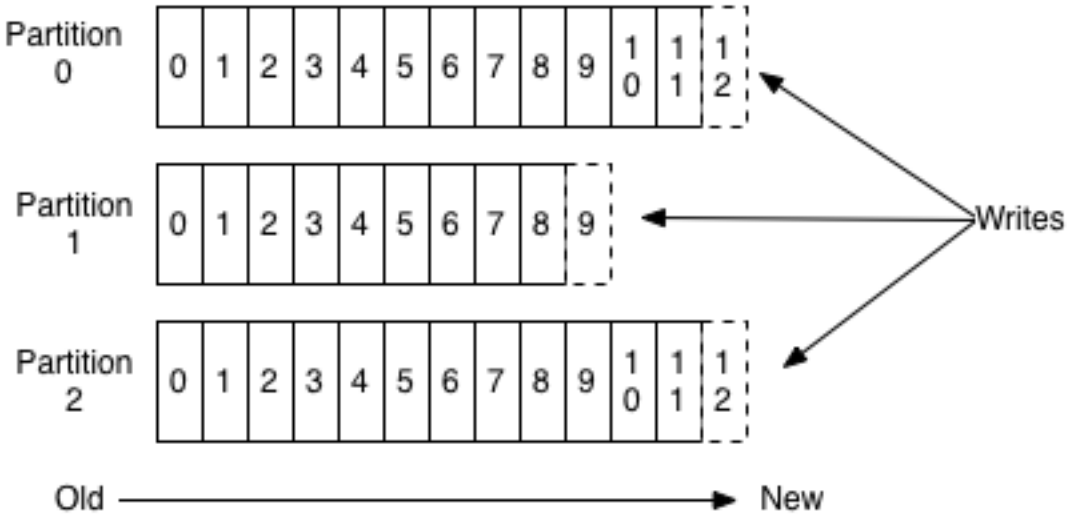

Producers
App
App
App
App
DB
Kafka Cluster
Stream Processors
Connectors
DB
App
App
App
App
Consumers

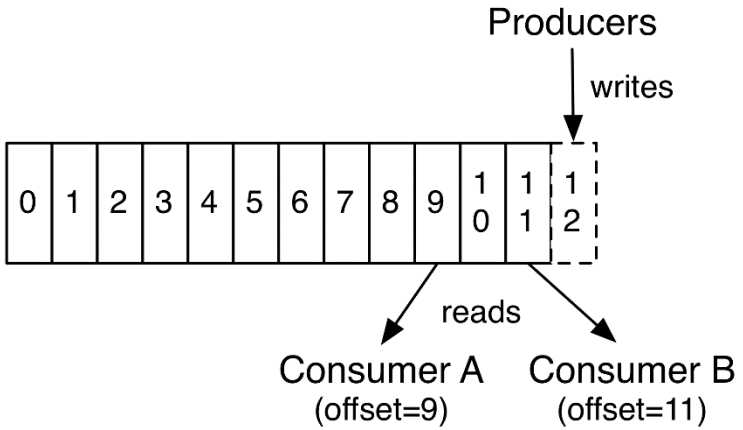
Components in Kafka

Components:

1. Producer
2. Consumer
3. Broker
4. Topic
5. Message
6. Partitions

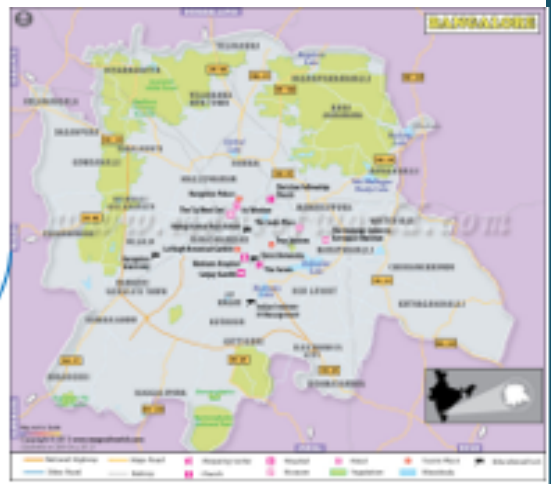
Anatomy of a Topic





INSOFE

Inspire...Educate...Transform.



HYDERABAD

Office and Classrooms

Plot 63/A, Floors 1&2, Road # 13, Film Nagar,
Jubilee Hills, Hyderabad - 500 033
+91-9701685511 (Individuals)
+91-9618483483 (Corporates)

BENGALURU

Office

Incubex, #728, Grace Platina, 4th Floor, CMH Road,
Indira Nagar, 1st Stage, Bengaluru – 560038
+91-9502334561 (Individuals)
+91-9502799088 (Corporates)

Classroom

KnowledgeHut Solutions Pvt. Ltd., Reliable Plaza,
Jakkasandra Main Road, Teacher's Colony, 14th Main
Road, Sector – 5, HSR Layout, Bengaluru - 560102

Social Media

Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering

This presentation may contain references to findings of various reports available in the public domain. INSOFE makes no representation as to their accuracy or that the organization subscribes to those findings.