

Enabling Real Time Analytics For IoT



Instructor: Ekpe Okorafor

1. Accenture – Big Data Academy
2. Computer Science African University of Science & Technology

Objectives

Objectives:

- Streaming and real-time data
- Stream & Real Time Processing Frameworks
- Streaming and Real Time Data Analytics for IoT

Introduction

- Modern data pipelines receive data at a high ingestion rate
- Volume, variety and velocity are important considerations for real time analytics in the context of Big Data / IoT
- To maximize the benefit of IoT data, we need an integrated platform to leverage the ability to collect, analyze and act upon this streaming data in real-time.
- Stream and real time processing frameworks together with analytics accommodating the variety, velocity and volume of big data generated by the Internet of Things

Streaming and Real-Time Data Processing

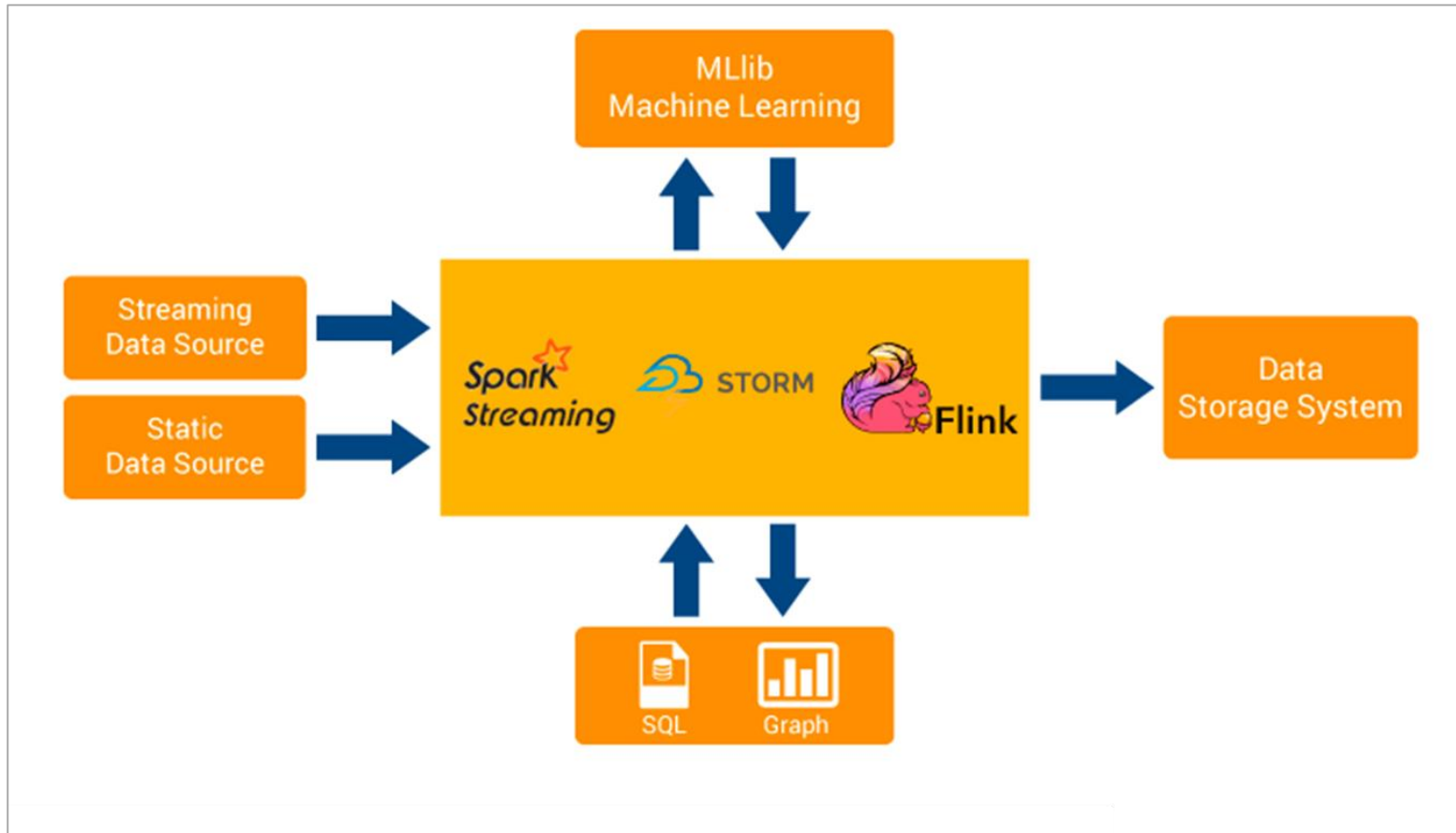
Stream Processing

- Refers to a method of continuous computation that happens as data is flowing through the system. There are no compulsory time limitations in stream processing.

Real-time Processing

- But Real-time data needs to have tight deadlines in the terms of time. So we normally consider that if our platform is able to capture any event within 1 ms, then we call it as real-time data or true streaming.

Stream & Real Time Processing Frameworks



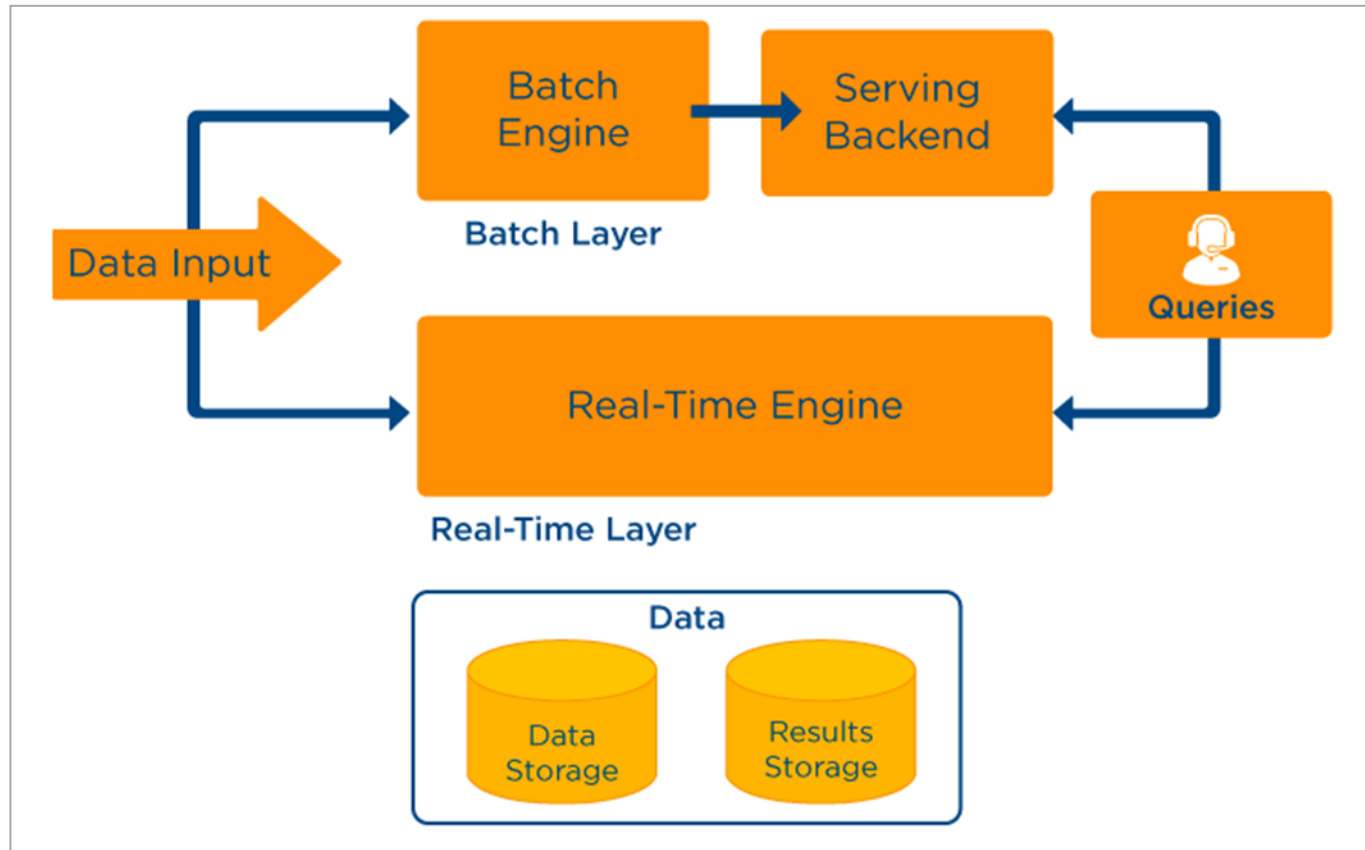
Processing of huge volumes of data is not enough. We need to process them in real-time so that decisions are taken immediately whenever any important event occurs.

Streaming Architectures

There are two types of architectures which are used while building real-time pipelines

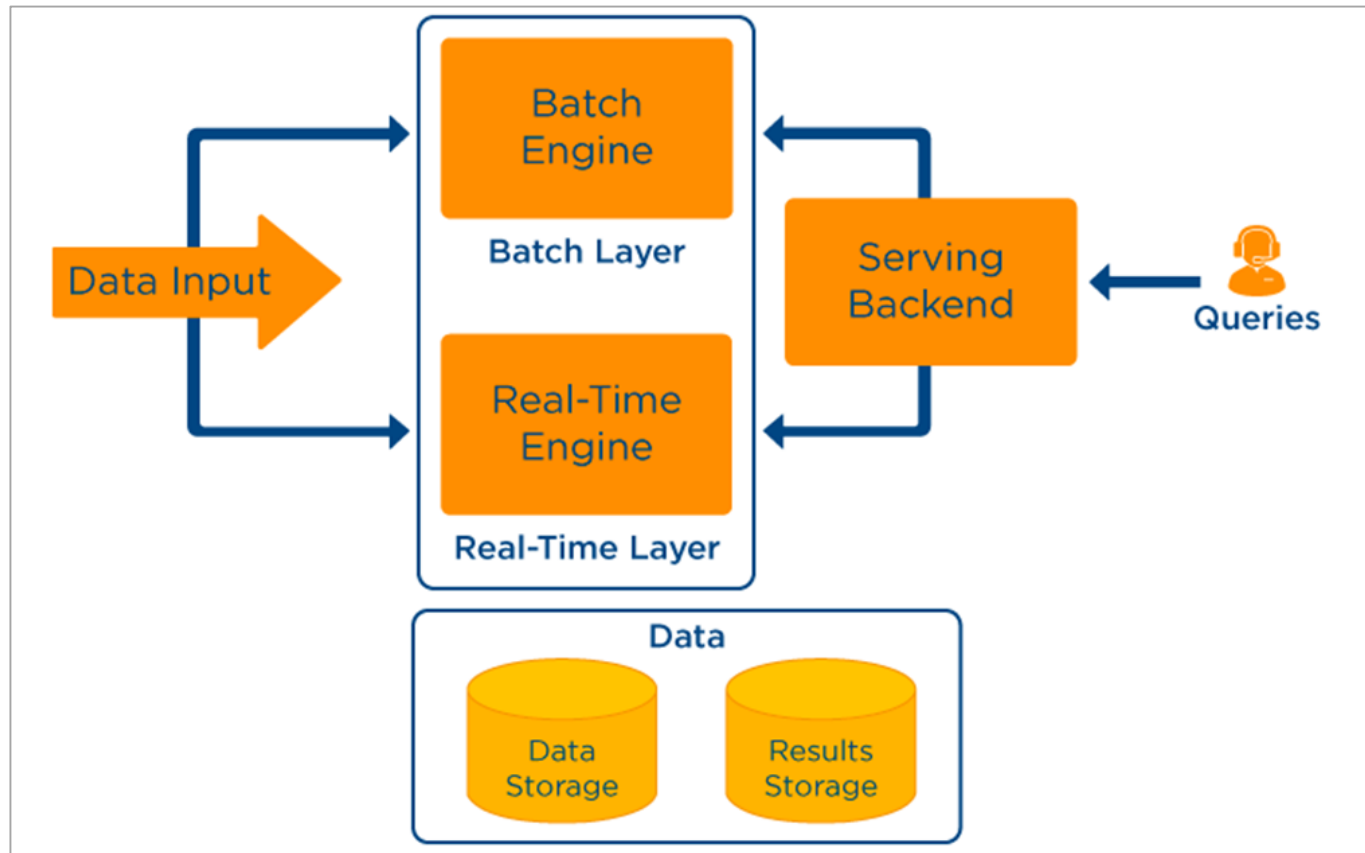
- Lambda Architecture
 - This architecture was introduced by Nathan Marz in which we have three layers to provide real-time streaming and compensate any data error occurs if any. The three layers are Batch Layer, Speed layer, and Serving Layer.
- Kappa Architecture
 - One of the important motivations for inventing the Kappa architecture was to avoid maintaining two separate code bases for the batch and speed layers. The key idea is to handle both real-time data processing and continuous data reprocessing using a single stream processing engine.

Lambda Architecture



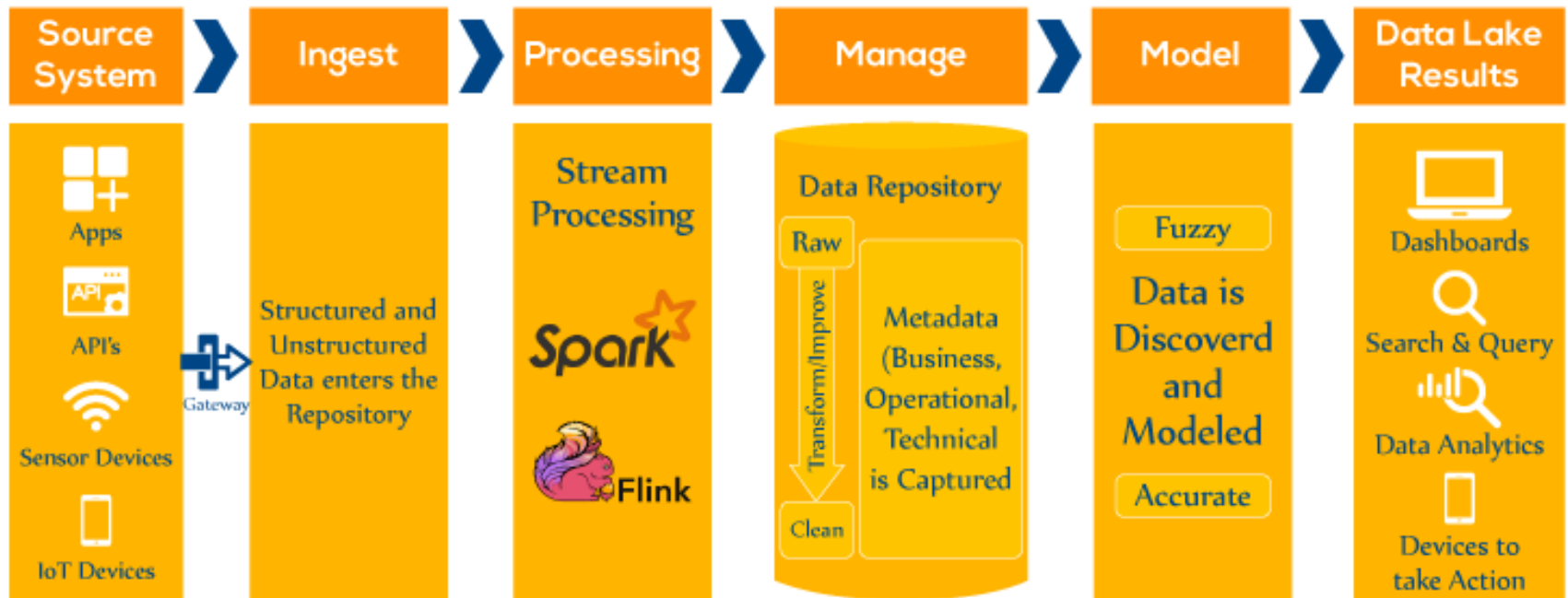
- The batch layer has two major tasks: (a) managing historical data; and (b) recomputing results such as machine learning models.
- The speed layer is used in order to provide results in a low-latency, near real-time fashion.

Kappa Architecture



- The key idea is to handle both real-time data processing and continuous data reprocessing using a single stream processing engine.

Streaming and Real Time Data Analytics for IoT



- To maximize the benefit of IoT data, there is a need for integrated platform to leverage the ability to collect, analyze and act upon this streaming data in real-time.

Summary
