Visualization of Streaming Data

Data Streaming

- What is meant by data streaming?

Streaming data is **data that is generated continuously by thousands of data sources, which typically send in the data records simultaneously, and in small sizes (order of Kilobytes)**.

Data streaming can also be explained as a technology used to deliver content to devices over the internet, and it allows users to access the content immediately, rather than having to wait for it to be downloaded.[

Streaming

- **What is Streaming?**
- The term "streaming" is used to describe continuous, never-ending data streams with no beginning or end, that provide a constant feed of data that can be utilized/acted upon without needing to be downloaded first.
- Similarly, data streams are generated by all types of sources, in various formats and volumes.

- From applications, networking devices, and server log files, to website activity, banking transactions, and location data, they can all be aggregated to seamlessly gather real-time information and analytics from a single source of truth.

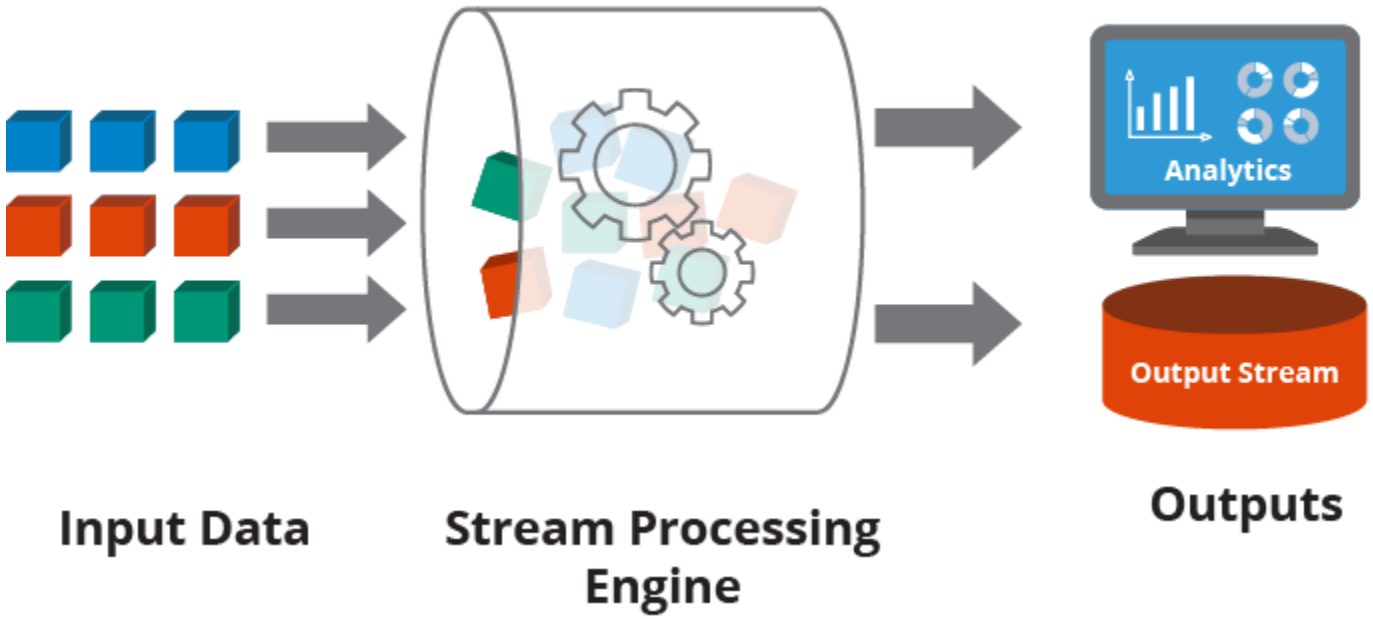
Applications

- Finance
- Real-estate
- Gaming
- Ecommerce
- Healthcare
- Transport
- Video Industry
- Music Industry

Characteristics of Data Streams

- Large volumes of continuous data, possibly infinite.
- Steady changing and requires a fast, real-time response.
- Data stream captures nicely our data processing needs of today.
- Random access is expensive and a single scan algorithm
- Store only the summary of the data seen so far.
- Maximum stream data are at a pretty low level or multidimensional in creation, needs multilevel and multidimensional treatment.

Stream Processing



Streaming Data Architecture

- A streaming data architecture is a framework of software components built to ingest and process large volumes of streaming data from multiple sources.

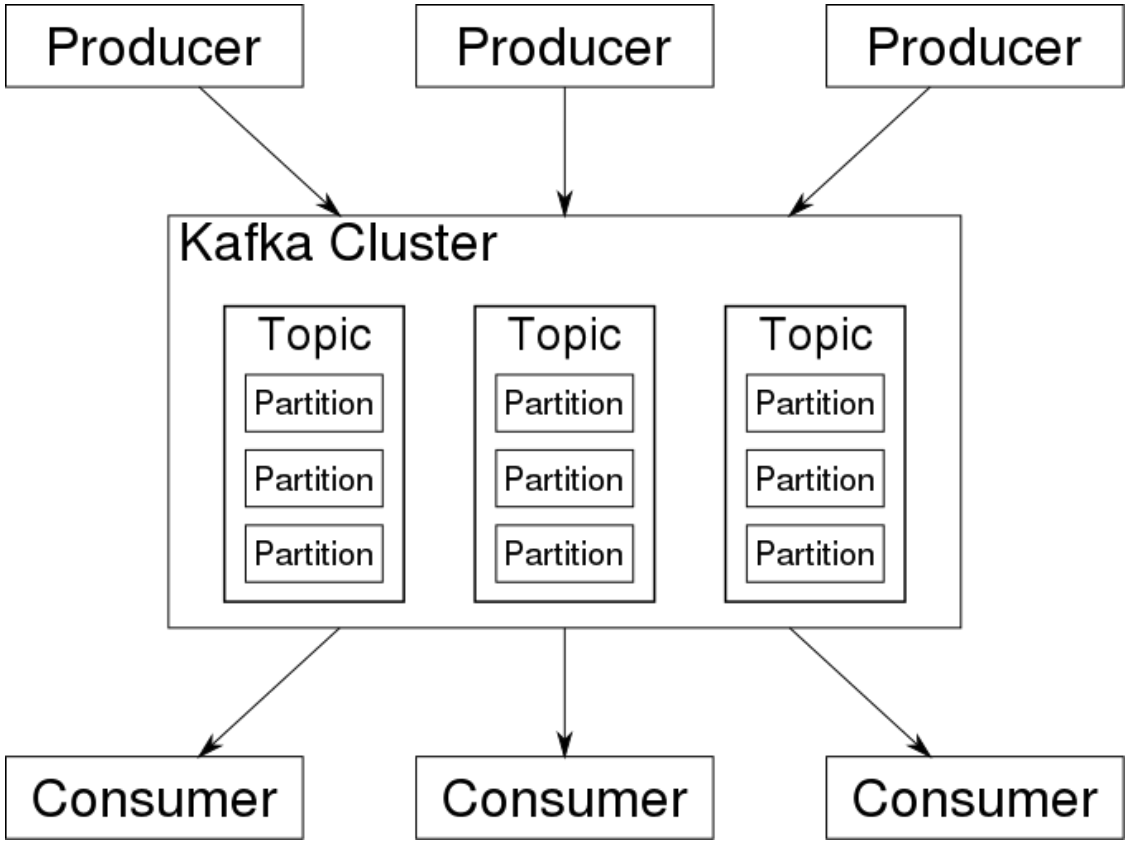
- While traditional data solutions focused on writing and reading data in batches, a streaming data architecture consumes data immediately as it is generated, persists it to storage, and may include various additional components per use case – such as tools for real-time processing, data manipulation, and analytics.

Benefits of Stream Processing

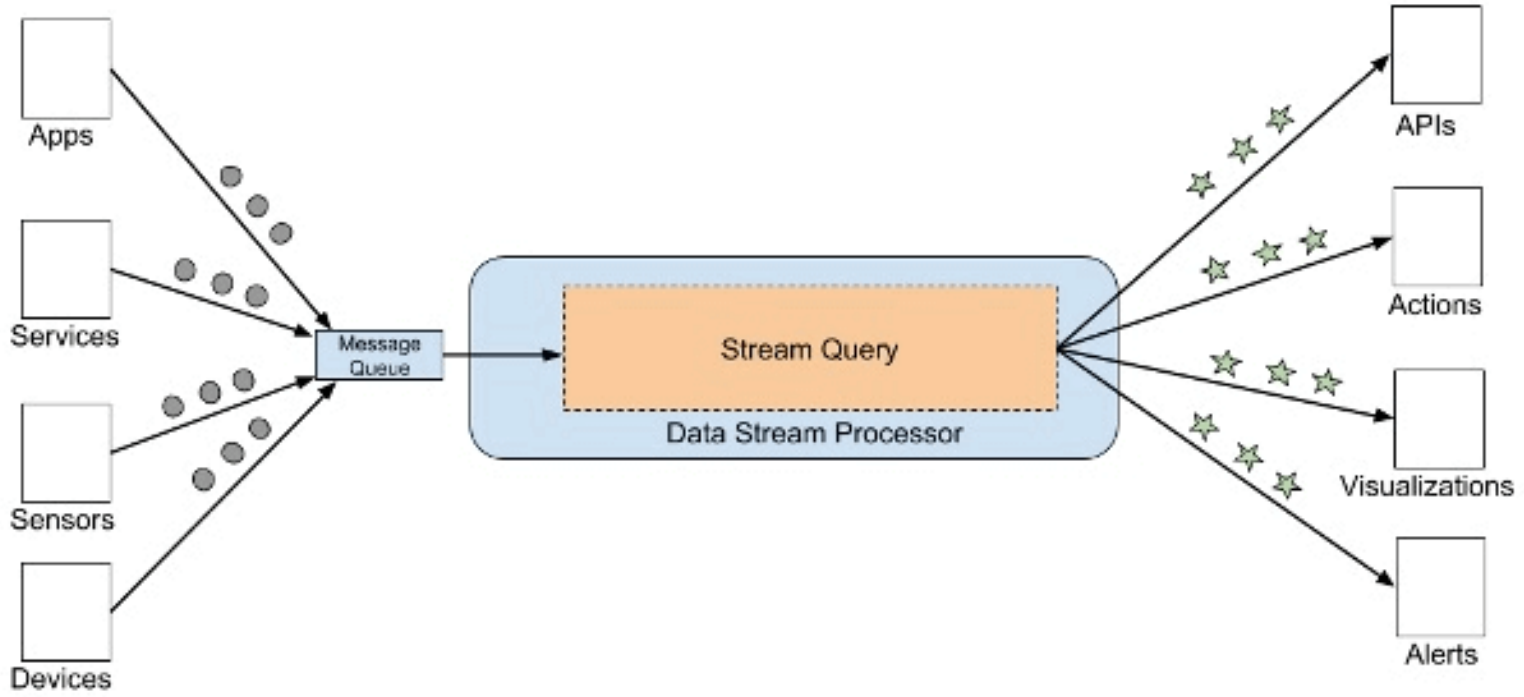
- **Able to deal with never-ending streams of events**
- **Real-time or near-real-time processing**
- **Detecting patterns in time-series data**
- **Easy data scalability**

The Components of a Streaming Architecture

- **1. The Message Broker / Stream Processor**



- **2. Batch and Real-time ETL Tools**



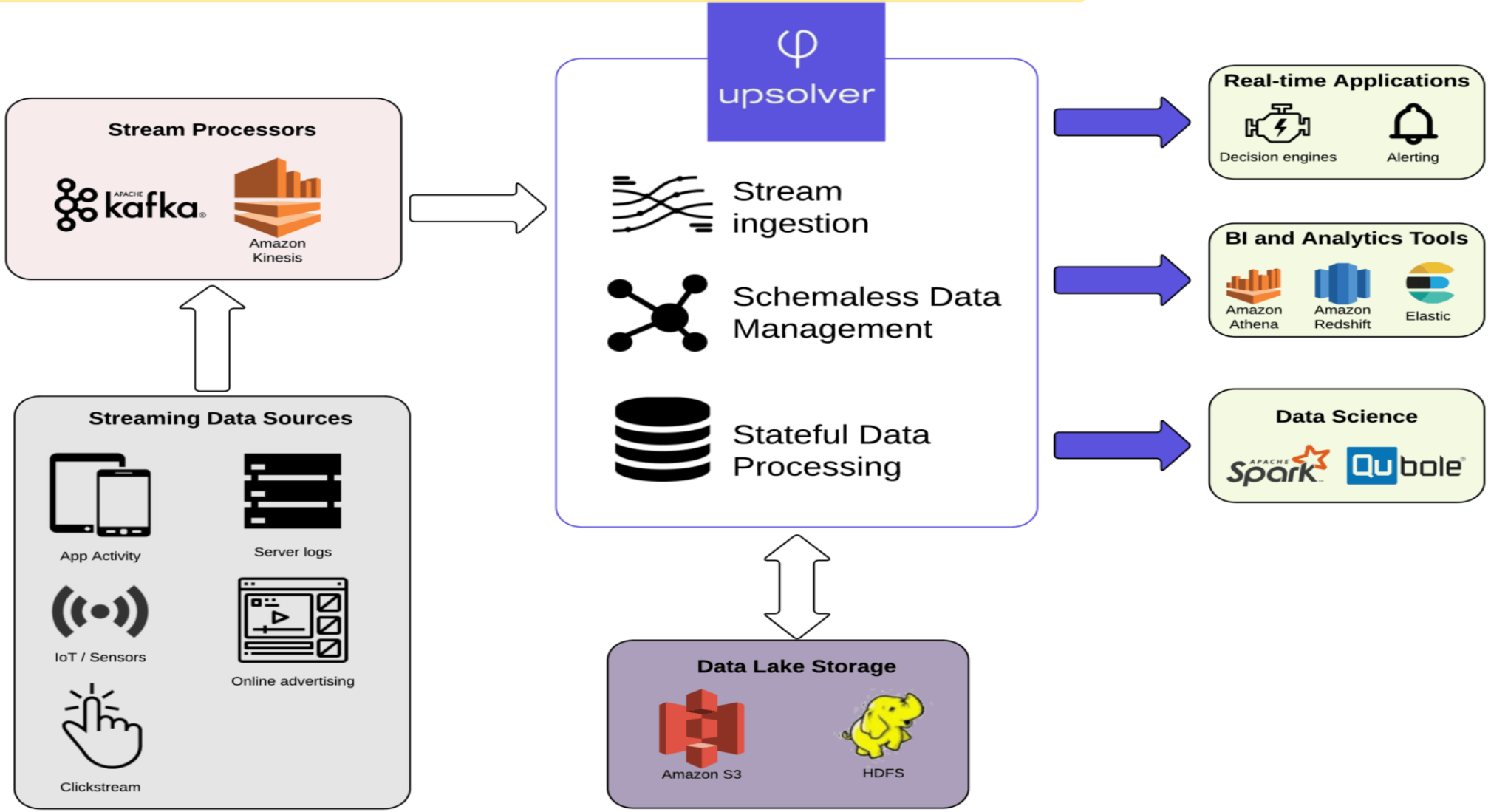
- **3. Data Analytics / Serverless Query Engine**
- After streaming data is prepared for consumption by the stream processor, it must be analyzed to provide value. There are many different approaches to streaming data analytics. Here are some of the tools most commonly used for streaming data analytics.

- Amazon Athena
- Amazon Redshift
- ElasticSearch
- Cassandra

Streaming Data Storage

- Database/Data Warehouse
- Message Broker
- Data lake – Amazon S3

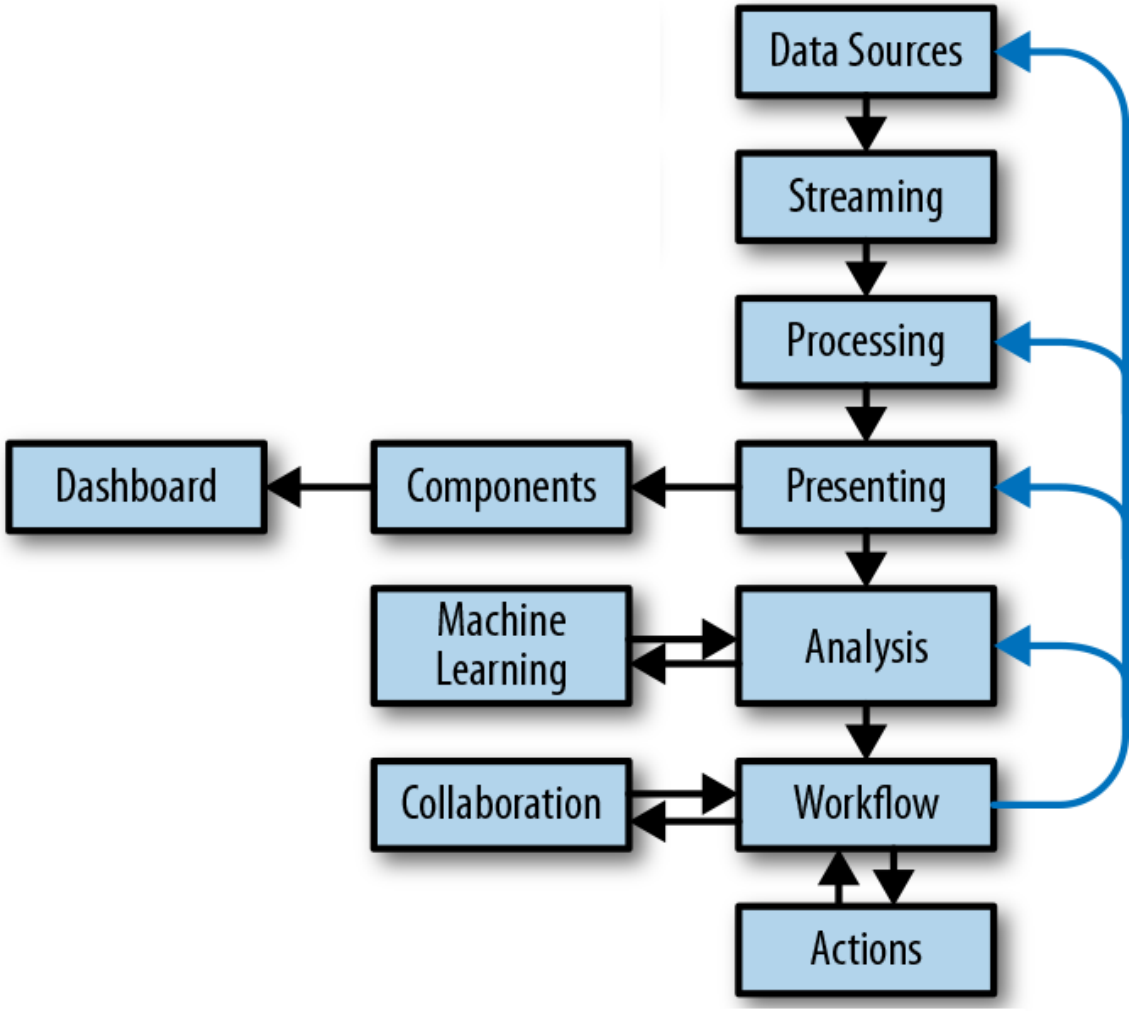
Modern Streaming Architecture



Benefits of a modern streaming architecture:

- Can eliminate the need for large data engineering projects
- Performance, high availability, and fault tolerance built in
- Newer platforms are cloud-based and can be deployed very quickly with no upfront investment
- Flexibility and support for multiple use cases

Streaming Data Visualization



Streaming Analytics

- Streaming analytics is the processing and analysis of fast-moving live data from a variety of sources, including IoT devices, to raise automated, real-time actions or alerts. It's essential for enterprises that want to extract immediate insights from fast and ever-growing volumes of data