# Data Ingestion Use Cases

In this lesson, we will discuss some common data ingestion use cases in the industry.

**We'll cover the following**



This is the part where I talk about some of the data streaming use cases commonly required in the industry.

## Moving Big Data Into Hadoop #

This is the most popular use case of data ingestion. As discussed before, Big Data from IoT devices, social apps & other sources, streams through data pipelines, moves into the most popular distributed data processing framework Hadoop for analysis & stuff.

## Streaming Data from Databases to Elasticsearch Server #

*Elastic search* is an open-source framework for implementing search in web applications. It is a defacto search framework used in the industry simply because of its advanced features, & it being open-source. These features enable businesses to write their own custom solutions when they need them.

In the past, with a few of my friends, I wrote a product search software as a service using *Java, Spring Boot* & *Elastic Search*. Speaking of its design, we would stream & index quite a large amount of product data from the legacy storage solutions to the Elastic search server in order to make the products come up in the search results.

All the data intended to show up in the search was replicated from the main storage to the Elastic search storage. Also, as the new data was persisted in the main storage it was asynchronously rivered to the Elastic server in real-time for indexing.

indexing.

## Log Processing #

If your project isn't a hobby project, chances are it's running on a cluster. When we talk about running a large-scale service, monolithic systems are a thing of the past. With so many microservices running concurrently. There is a massive number of logs which is generated over a period of time. And logs are the only way to move back in time, track errors & study the behaviour of the system.

So, to study the behaviour of the system holistically, we have to stream all the logs to a central place. Ingest logs to a central server to run analytics on it with the help of solutions like ELK Elastic LogStash Kibana stack etc.

## Stream Processing Engines for Real-Time Events #

Real-time streaming & data processing is the core component in systems handling LIVE information such as sports. It's imperative that the architectural setup in place is efficient enough to ingest data, analyse it, figure out the behaviour in real-time & quickly push the updated information to the fans. After all, the whole business depends on it.

Message queues like *Kafka*, Stream computation frameworks like *Apache Storm, Apache Nifi, Apache Spark, Samza, Kinesis* etc are used to implement the real-time large-scale data processing features in online applications.

This is a good read on the topic:

An Insight into Netflix's real-time streaming platform

Alright!! time to have a look into data pipelines in the lesson up-next.