

# Data Ingestion

In this lesson, we will have an insight into the process of data ingestion.

## We'll cover the following

- What Is Data Ingestion?
- Layers Of Data Processing Setup
- Data Standardization
- Data Processing
- Data Analysis
- Data Visualization
- Data Storage & Security

## What Is Data Ingestion?

*Data Ingestion* is a collective term for the process of collecting data streaming-in from several different sources and making it ready to be processed by the system.

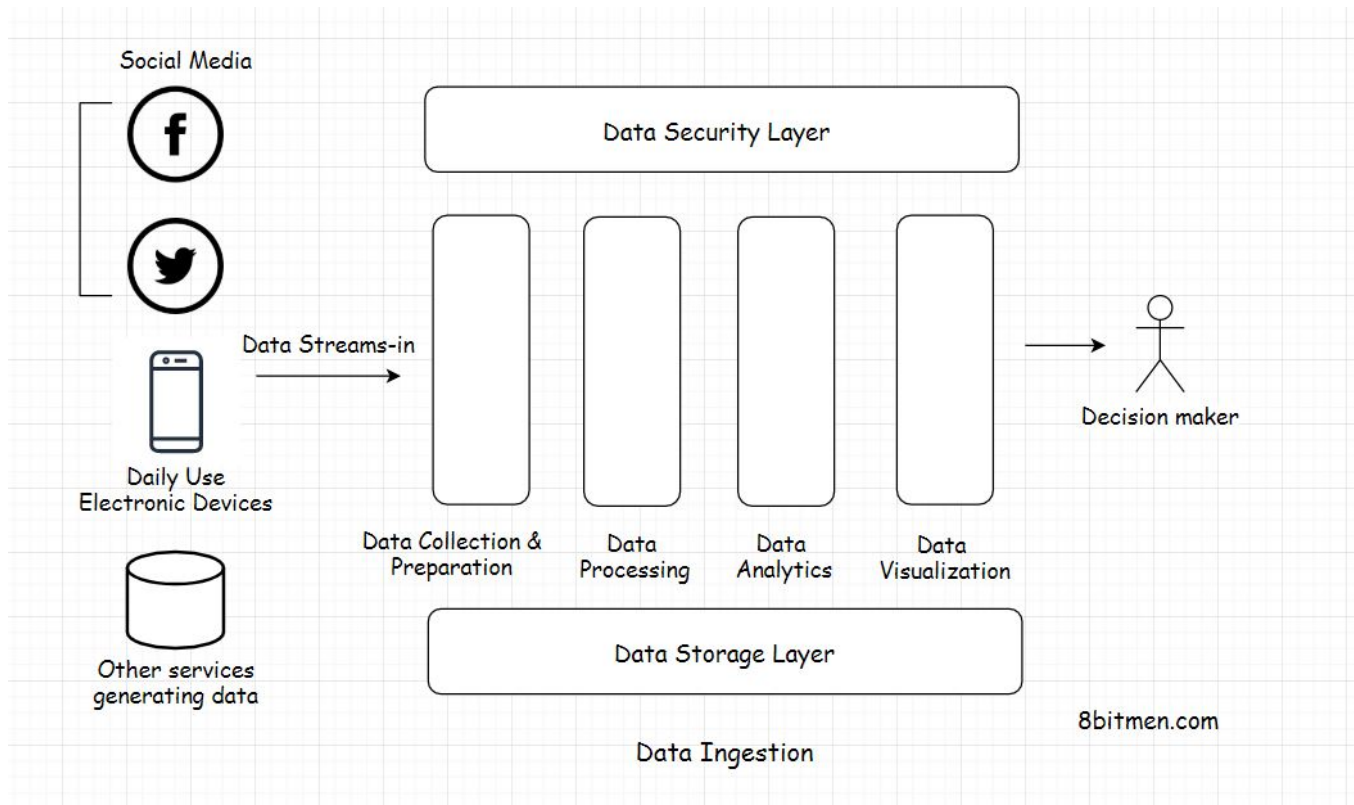
In a data processing system, the data is ingested from the IoT devices & other sources, into the system to be analysed. It is routed to different components/layers through the *data pipelines*, algorithms are run on it and is eventually archived.

## Layers Of Data Processing Setup

There are several stages/layers to this whole data processing setup such as the:

- Data collection layer
- Data query layer
- Data processing layer

- Data visualization layer
- Data storage layer
- Data security layer



As you can see in the diagram all the data processing layers are pretty self-explanatory.

## Data Standardization #

The data which streams in from several different sources is not in a homogeneous structured format. We have already gone through different types of data, structured, unstructured, semi-structured in the database lesson. So, you have an idea of what unstructured heterogeneous data is.

Data streams-in into the system at different speeds & sizes, from the web-based services, social networks, IoT devices, industrial machines & whatnot. Every stream of data has different semantics.

So, in order to make the data uniform and fit for processing, it has to be first collected and converted into a standardized format to avoid any future processing issues. This process of data standardization occurs in the *Data collection and preparation layer*.

## Data Processing #

Once the data is transformed into a standard format it is routed to the *Data processing layer* where it is further processed based on the business requirements. It is generally classified into different flows, routed to different destinations.

## Data Analysis #

After being routed, analytics is run on the data which includes execution of different analytics models such as predictive modelling, statistical analytics, text analytics etc. All the analytical events occur in the *Data Analytics layer*.

## Data Visualization #

Once the analytics are run & we have valuable intel from it. All the information is routed to the *Data visualization layer* to be presented before the stakeholders, generally in a web-based dashboard.

*Kibana* is one good example of a data visualization tool, pretty popular in the industry.

## Data Storage & Security #

Moving data is highly vulnerable to security breaches. *The Data security layer* ensures the secure movement of data all along. Speaking of the *Data Storage layer*, as the name implies, is instrumental in persisting the data.

So, this is a gist of how massive amounts of data is processed and analyzed for business use cases. This is just a bird's eye view of things. The field of data analytics is pretty deep, an in-depth detailed microscopic view of each layer demands a dedicated data analytics course for itself.

Alright, now let's have a look at the different ways in which the data can be ingested.