



Data Ingestion

In this lesson, you will gain 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 and 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.

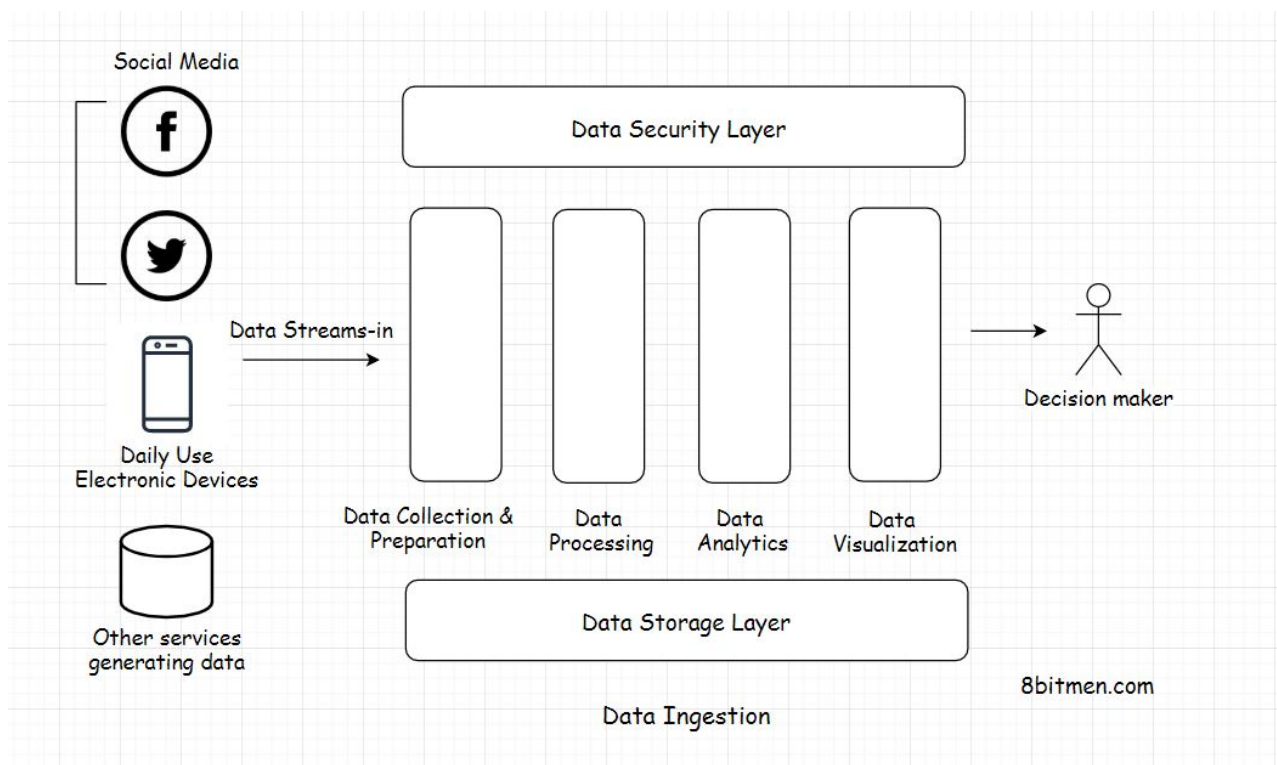
In a data processing system, the data is ingested from the IoT devices and other sources into the system to be analyzed. It is routed to different components/layers through the *data pipelines*, algorithms are run on it, and it is eventually archived.



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, including structured, unstructured, and 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 and sizes from web-based services, social networks, IoT devices, industrial machines, and 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 data standardization process 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 and routed to different destinations.

Data analysis#

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



Data visualization#

Once the analytics are run and 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.

Data storage and 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 the gist of how massive amounts of data are 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, and an in-depth detailed microscopic view of each layer demands a dedicated data analytics course for itself.

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

[< Back](#)[Next >](#)[Introduction](#)[Different Ways of Ingesting Data and ...](#)[Mark as Completed](#)



Report an Issue