

Introduction & Types of Data

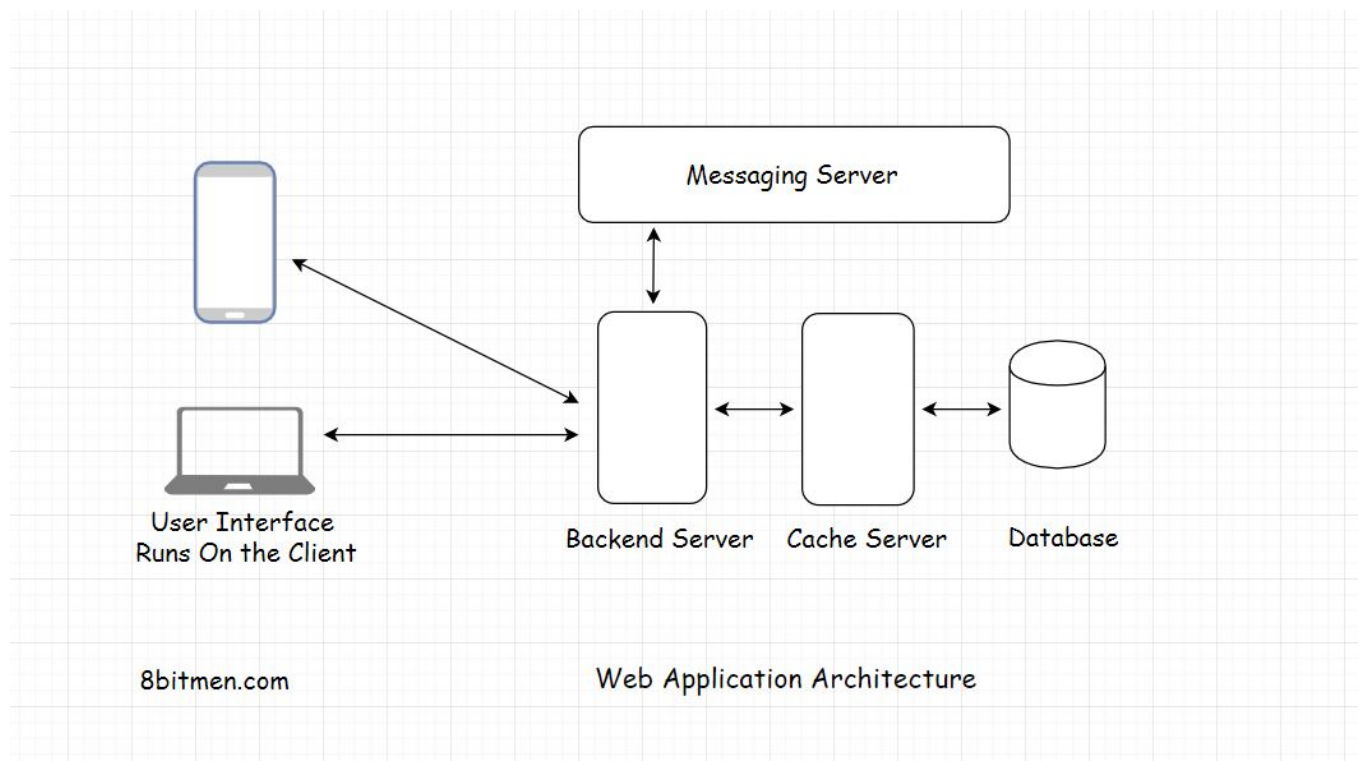
In this lesson, we will have an introduction to databases and the different types of data.

WE'LL COVER THE FOLLOWING ^

- What Is A Database?
- Structured Data
- Unstructured Data
- Semi-structured Data
- User state

What Is A Database?

A database is a component required to persist data. Data can be of many forms: structured, unstructured, semi-structured and user state data.



Let's quickly have an insight into the classification of data before delving into the databases.

Structured Data

Structured data is the type of data which conforms to a certain structure, typically stored in a database in a *normalized* fashion.

There is no need to run any sort of data preparation logic on structured data before processing it. Direct interaction can be done with this kind of data.

An example of structured data would be the personal details of a customer stored in a database row. The customer id would be of *integer* type, the name would be of *String* type with a certain character limit etc.

So, with structured data, we know what we are dealing with. Since the customer name is of String type, without much worry of errors or exceptions, we can run String operations on it.

Structured data is generally managed by a query language such as *SQL* (*Structured query language*).

Unstructured Data

Unstructured data has no definite structure. It is generally the heterogeneous type of data comprising of text, image files, video, multimedia files, pdfs, Blob objects, word documents, machine-generated data etc.

This kind of data is often encountered in data analytics. Here the data streams in from multiple sources such as IoT devices, social networks, web portals, industry sensors etc. into the analytics systems.

We cannot just directly process unstructured data. The initial data is pretty raw, we have to make it flow through a data preparation stage which segregates it based on some business logic & then runs the analytics algorithms on it.

Semi-structured Data

Semi-structured data is a mix of structured & unstructured data. Semi-structured data is often stored in data transport formats such as XML or JSON and is handled as per the business requirements.

User state

User state

The data containing the user state is the information of activity which the user performs on the website.

For instance, when browsing through an e-commerce website, the user would browse through several product categories, change the preferences, add a few products to the reminder list for the price drops.

All this activity is the user state. So next time, when the user logs in, he can continue from where he left off. It would not feel like that one is starting fresh & all the previous activity is lost.

Storing user state improves the browsing experience & the conversion rate for the business.

So, now we are clear on the different types of data. Let's have a look into different types of databases.

There are multiple different types of databases with specific use cases. We'll quickly go through each of them in order to have a bird's eye view of the database realm.