

The background is a light cream color with a subtle pattern of small, faint geometric shapes. It is decorated with large, flowing abstract shapes in orange and dark blue. Various geometric elements are scattered throughout: a cluster of small 'x' marks in the top left, a grid of dots with an 'x' in the top center, a stylized 'D' shape in the top right, a circular pattern of dots in the bottom left, and a cluster of 'L' shapes in the bottom right. The title text is centered in a bold, dark blue serif font.

Introduction to Data and Database

Profile



 [linkedin.com/in/romansyasetyo/](https://www.linkedin.com/in/romansyasetyo/)



**Data will talk to you
if you're willing to listen**

- Jim Bergeson

Table of Content

What will We Learn Today?

1. What is Data?
2. Data Types
3. About Database
4. Data Architecture
5. The World of Big Data

Introducing Google Colab

Introducing DBeaver





What is Data?





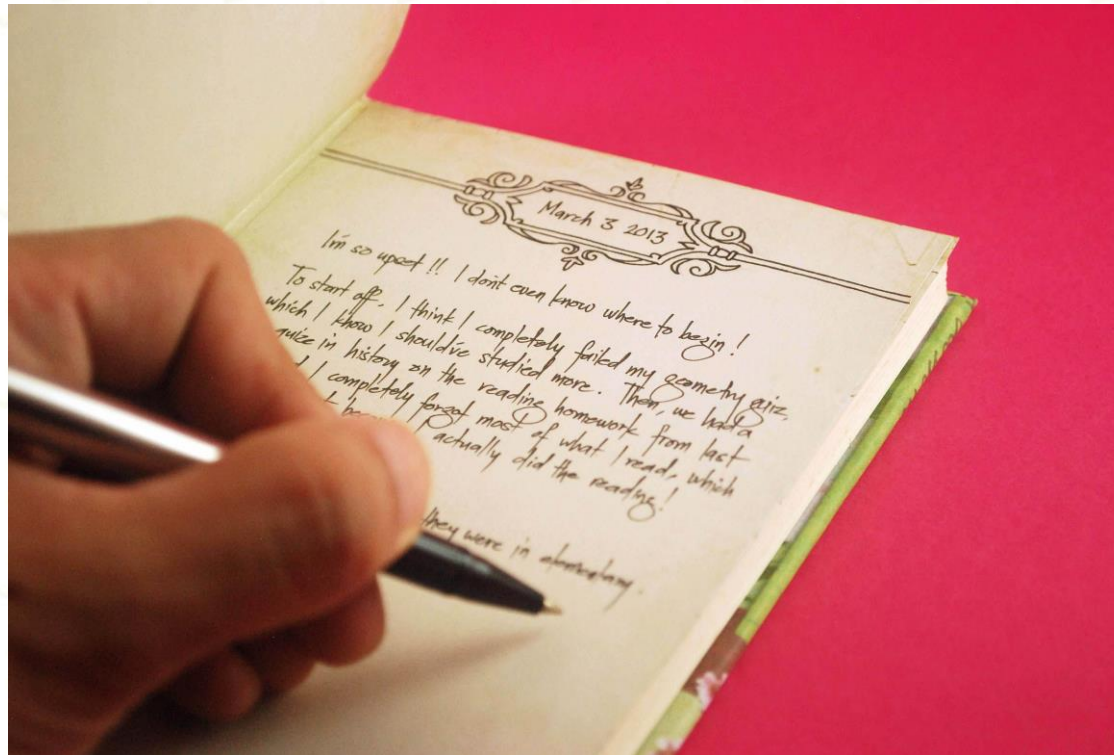
Data / Datum



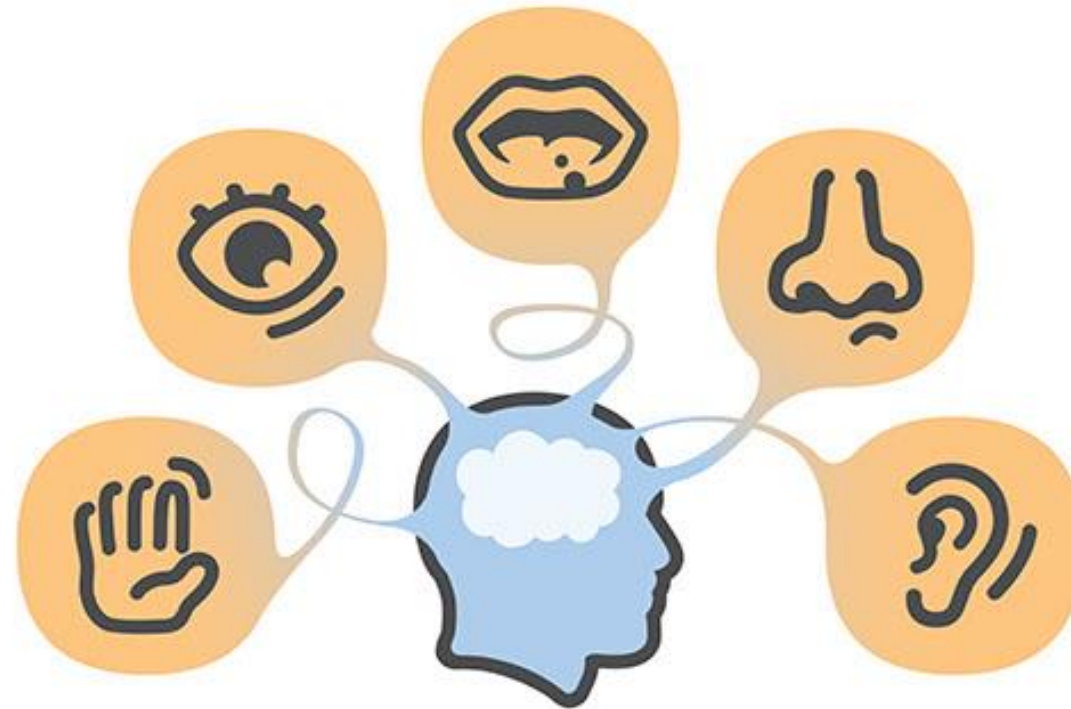
Factual information

Measurement result of
something that is understood by
both the recorder and the reader











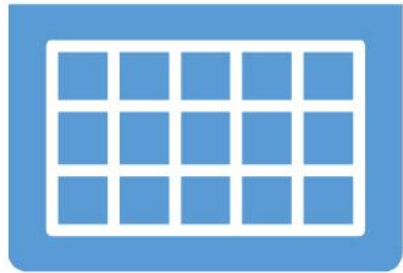
Any recorded facts
are
DATA



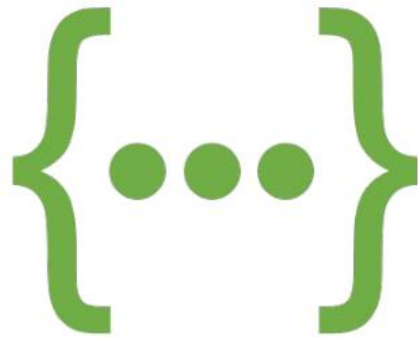


Data Types





Structured

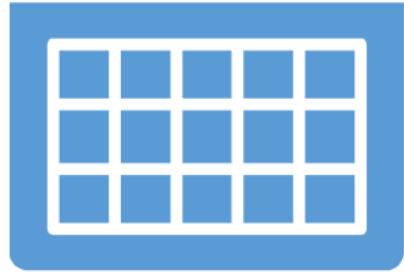


Semi-structured



Unstructured





Structured Data

- Displayed in rows, columns
- Stored in relational databases
- It has strict structure and data type (hard to be changed)
- Easier to manage, protect, transform, and analyze
- Example: Excel, CSV, MySQL, PostgreSQL, Data Warehouse





Semi-structured Data

- Displayed key-paired values
- Stored in semi-structured database (like MongoDB)
- It is flexible and can be nested
- Common transactional data type (for API)
- Example: JSON, XML, Parquet, Avro





Unstructured Data

- Can be found in almost any file
- Stored in folder
- It represents specific object
- It is hard to manage, secure, and process
- Example: text, image, video, music





About Database





Database



A database is an **organized** collection of data, generally stored and accessed electronically from a computer system



Type of Database



OLTP

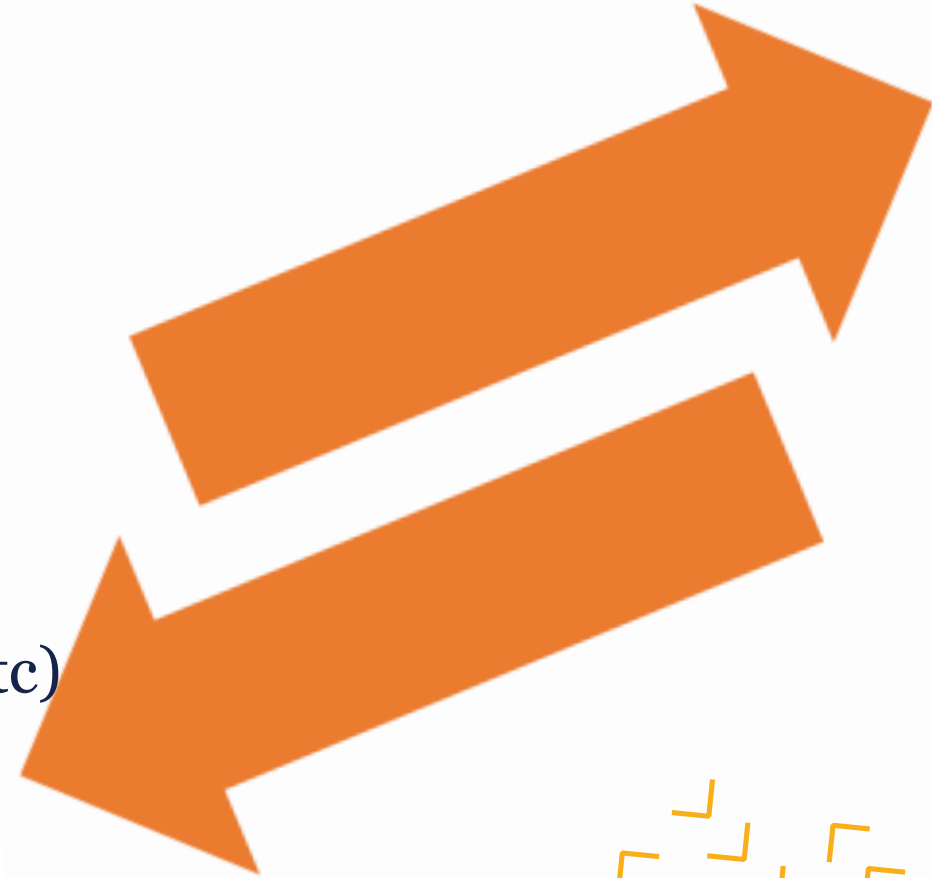


OLAP



OLTP


- On Line Transaction Processing
- It is high volume transaction database
- Faster to WRITE
- Row-based record
- Used for external needs (users, partners, API, etc)
- Normalized Data (Data separated to different tables)
- Example: MySQL, MariaDB, MongoDB, PostgreSQL




OLAP


- On Line Analytical Processing
- It is heavy processing database
- Faster to READ
- Column-based record
- Limited to internal usage (Dashboard, Report, etc)
- Denormalized Data (Data merged into analytic ready table)
- Example: Data Warehouse, RedShift (AWS), BigQuery (GCP), Teradata






	OLTP	OLAP
Characteristics	Handles a large number of small transactions	Handles large volumes of data with complex queries
Query types	Simple standardized queries	Complex queries
Operations	Based on INSERT, UPDATE, DELETE commands	Based on SELECT commands to aggregate data for reporting
Response time	Milliseconds	Seconds, minutes, or hours depending on the amount of data to process





	OLTP	OLAP
Design	Industry-specific, such as retail, manufacturing, or banking	Subject-specific, such as sales, inventory, or marketing
Source	Transactions	Aggregated data from transactions
Purpose	Control and run essential business operations in real time	Plan, solve problems, support decisions, discover hidden insights
Data updates	Short, fast updates initiated by user	Data periodically refreshed with scheduled, long-running batch jobs

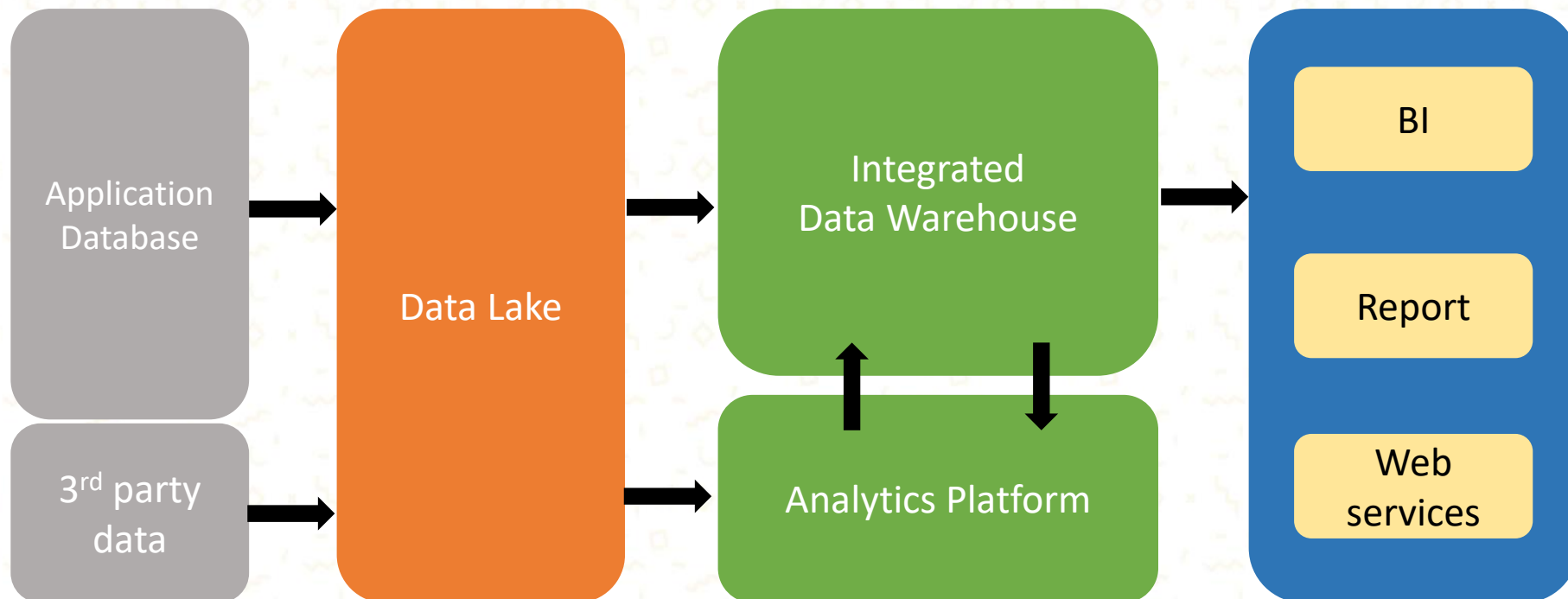




Data Architecture



Ideal Architecture





The World of Big Data



BIG DATA





ABC

1 letter = 1 Byte





One page has +1000 letters so it is **1 KB**
A book is consists of hundreds pages so
it is **1 MB**

Total Data Produced

2005 – 130 ExaBytes

2010 – 1,200 ExaBytes

2015 – 7,900 ExaBytes

2020 – 40,900 ExaBytes

IDC's Digital Universe Study 2012

*1 Exabyte = 1.000.000.000 GB



5V of Big Data

Volume

Refers to the vast amounts of data generated every seconds

Variety

Refers to the different types of data we can now use

Velocity

Refers to the speed at which new data is generated and the speed at which data data movers around

Veracity

Refers to the messiness or trushworthieness of the data

Value

Refers to having access to big data is no good unless we can turn it into value



Tools





Google Colab

<https://colab.research.google.com/>





Welcome To Colaboratory

File Edit View Insert Runtime Tools Help

Share



Sign in

Table of contents

- Getting started
- Data science
- Machine learning
- More Resources
- Machine Learning Examples
- + Section

+ Code + Text Copy to Drive

Connect

Editing

What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

Getting started






DBEaver

<https://dbeaver.io/download/>







DBEaver Community
Free Universal Database Tool

Star 19,808 Follow @dbeaver_news

search here ... Go

HomeAboutDownloadSourcesDocumentationNewsSupportEnterprise EditionCloudBeaver

Download

<h3>Community Edition 21.0.1</h3> <p>Released on March 22, 2021 (Milestones). It is free and open source (license). Also you can get it from the GitHub mirror.</p>	<h3>Enterprise Edition 21.0</h3> <p>Released on March 8, 2020 EE version web site: dbeaver.com Trial version is available.</p>
<div> Windows</div> <ul style="list-style-type: none">• Windows 64 bit (installer)• Windows 64 bit (zip)• Install from Microsoft Store• Chocolatey (choco install dbeaver) <div> Mac OS X</div> <ul style="list-style-type: none">• Mac OS X (dmg)• Mac OS X (zip)	<div>Enterprise Edition features:</div> <ul style="list-style-type: none">• Support of NoSQL databases:<ul style="list-style-type: none">• MongoDB• Cassandra• InfluxDB• Redis• Amazon DynamoDB• Amazon DocumentDB• Amazon Keyspaces• Google Bigtable• Couchbase• CouchDB

Thank You

