

Topics you will find in this deck

- What is Data?
- Data Non Technical prospective
- Data Technical Types
- Data Categorization
- Data Characteristics - Format, Velocity, Volume, Methods of Extraction
- Data Characteristics - Residency, Access, Source Systems, Temperature
- What are Dimension and Measures ?

You are free to

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material for any purpose, even commercially.

Under the following terms*

Attribution — You must give appropriate credit, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the Author endorses you or your use

Please refer for the full terms: <https://creativecommons.org/licenses/by/4.0/>

*The author does not provide any guarantees, warranties for the content and accuracy. The content may or may not be updated. By using this material you indemnify the author against any liabilities including copyrights and/or damages of any sort.

What is data?



Month's Expenses

Mortgage

Car

Electricity

Phone

Mobile Phone
Insurance

Gas

Water

Broadband

Cable

Membership

950.60

372.51

129.29

39.15

101.74

97.31

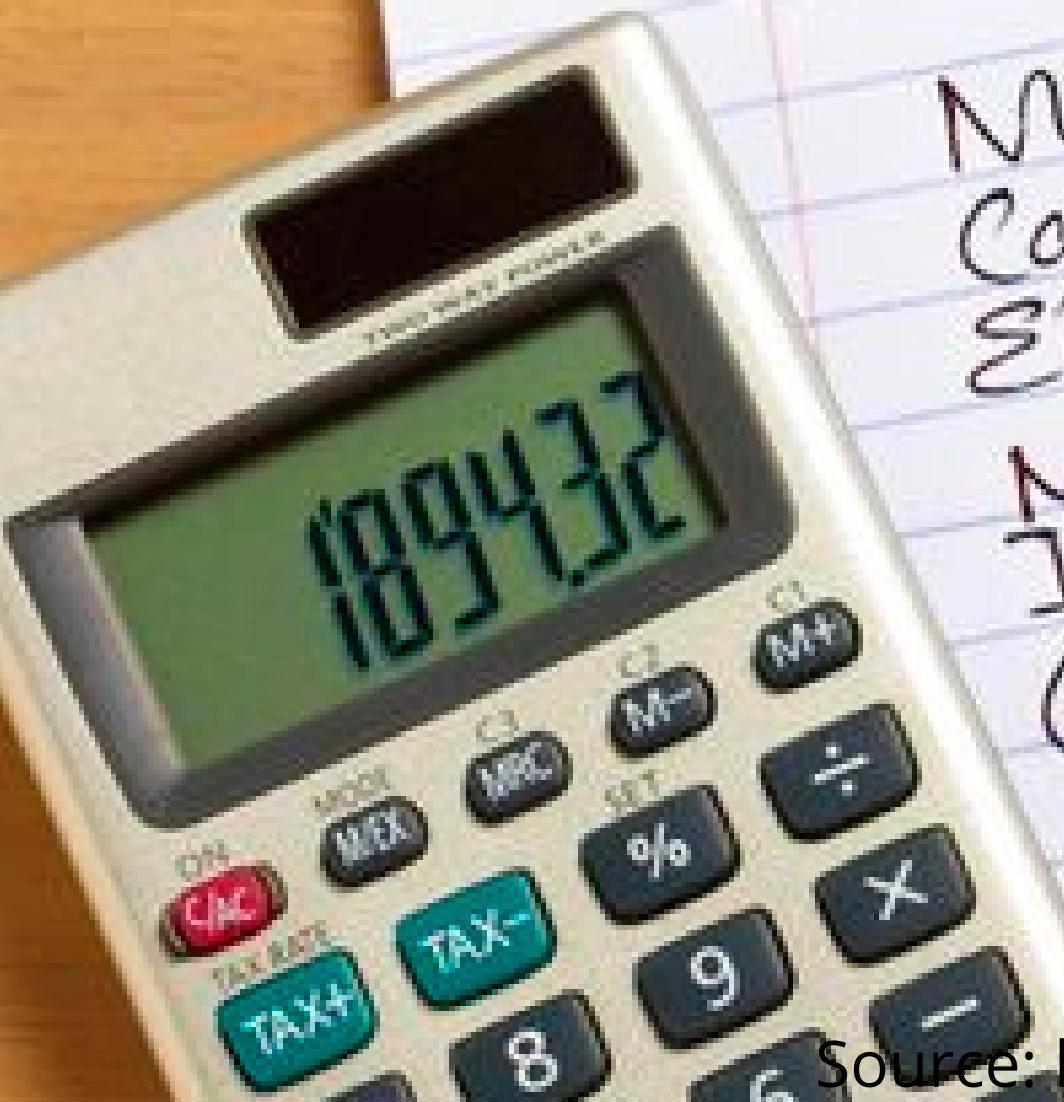
98.77

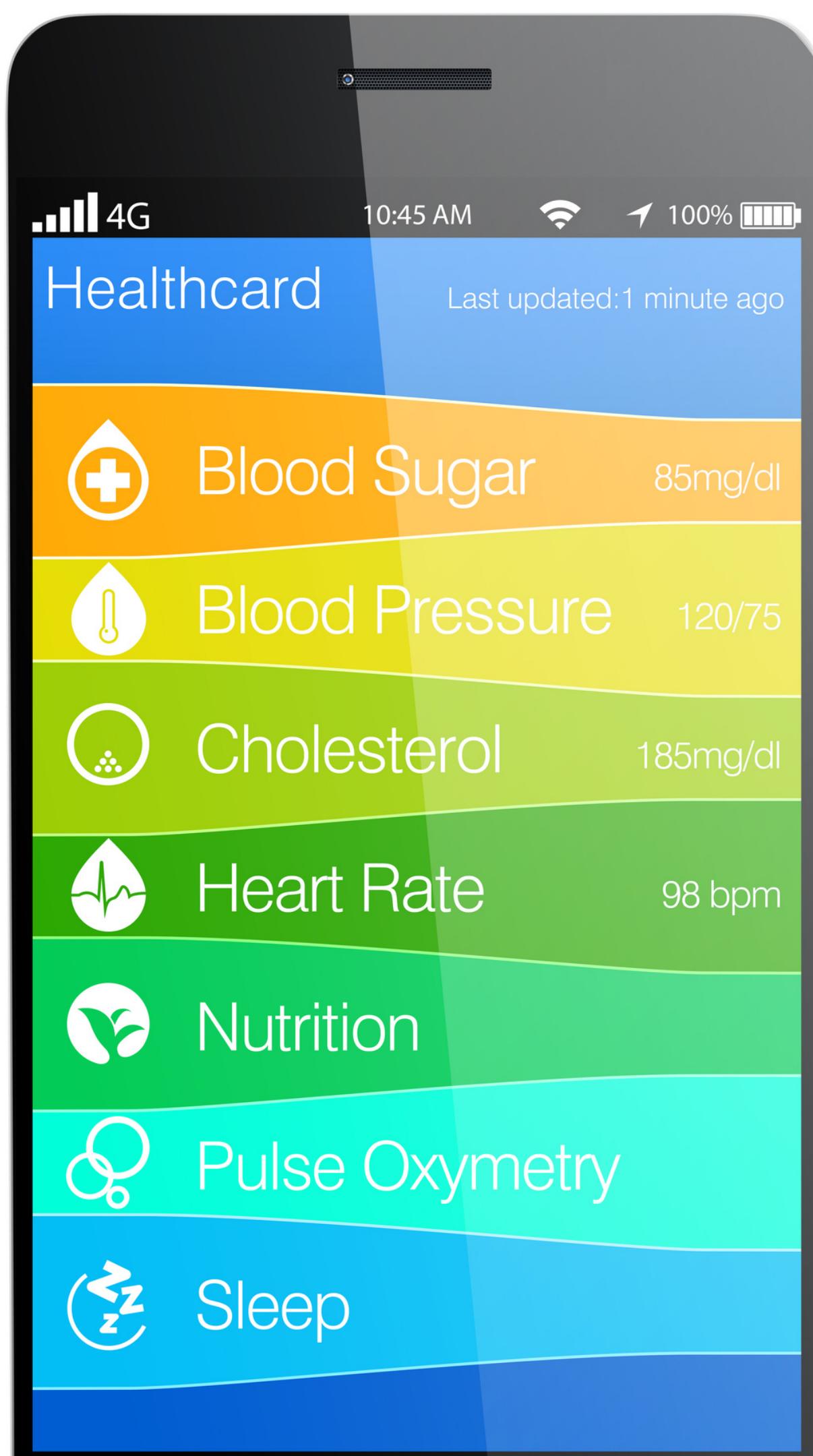
19.95

40.00

45.00

1894.33







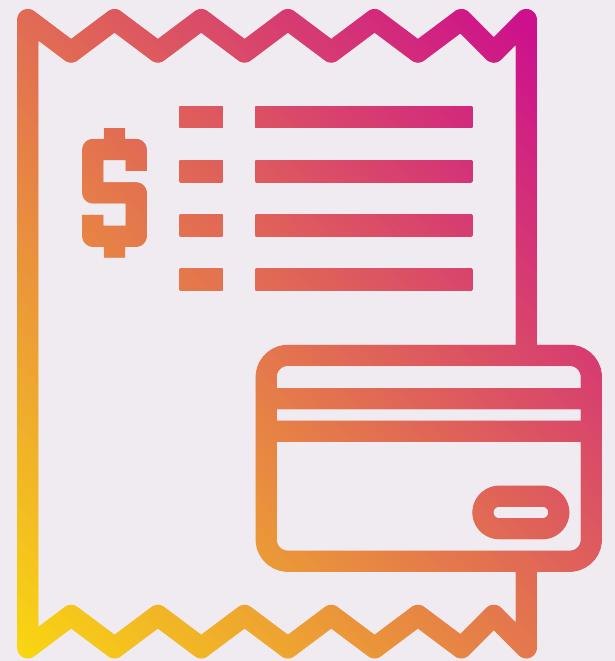
It is a digital record of interactions between humans via technology or between humans and technology systems or in between technology Systems

– Junaid Ahmed

1010
0101
1001



Excel



CC Bills



Notes



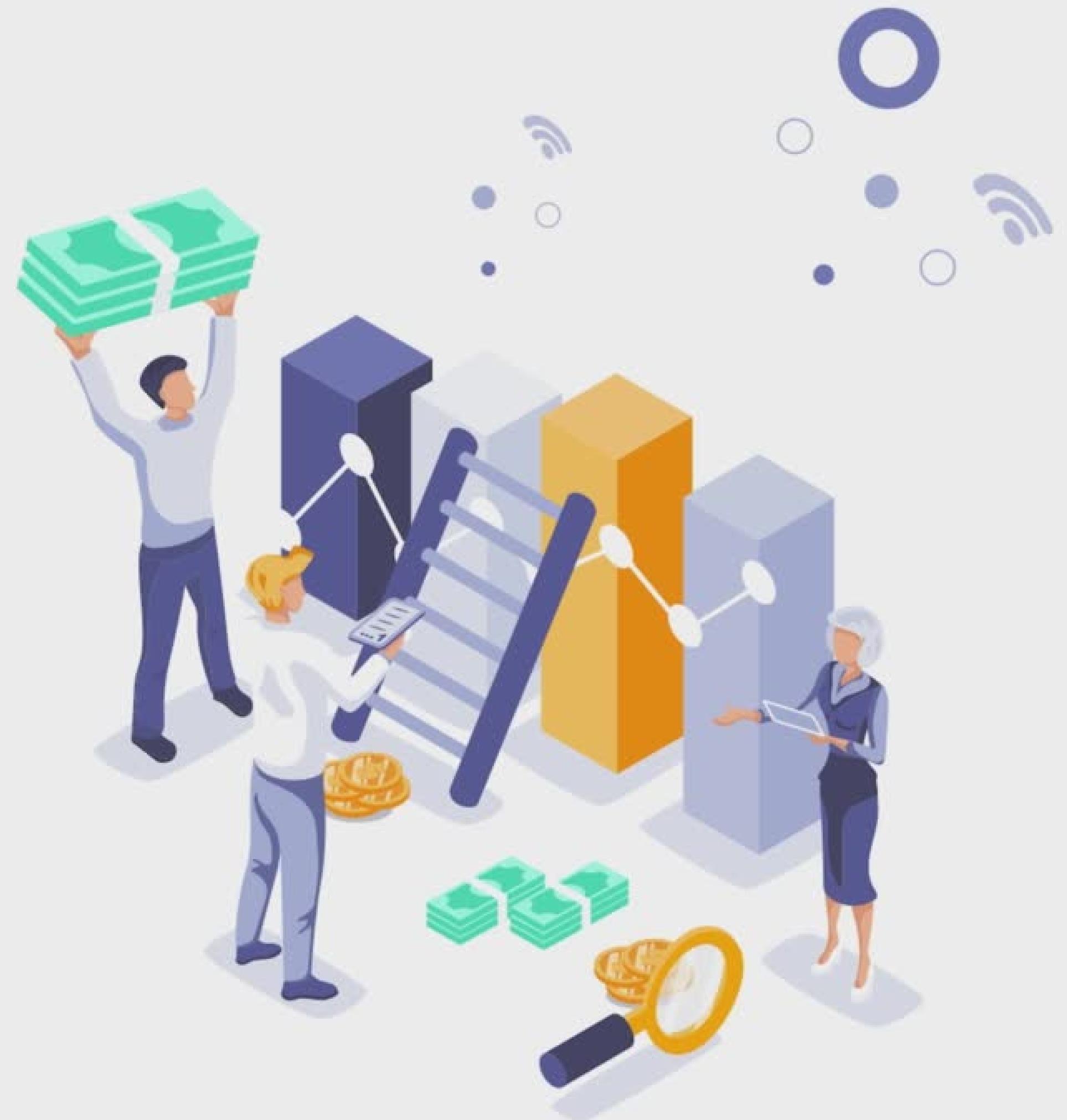
Social Media

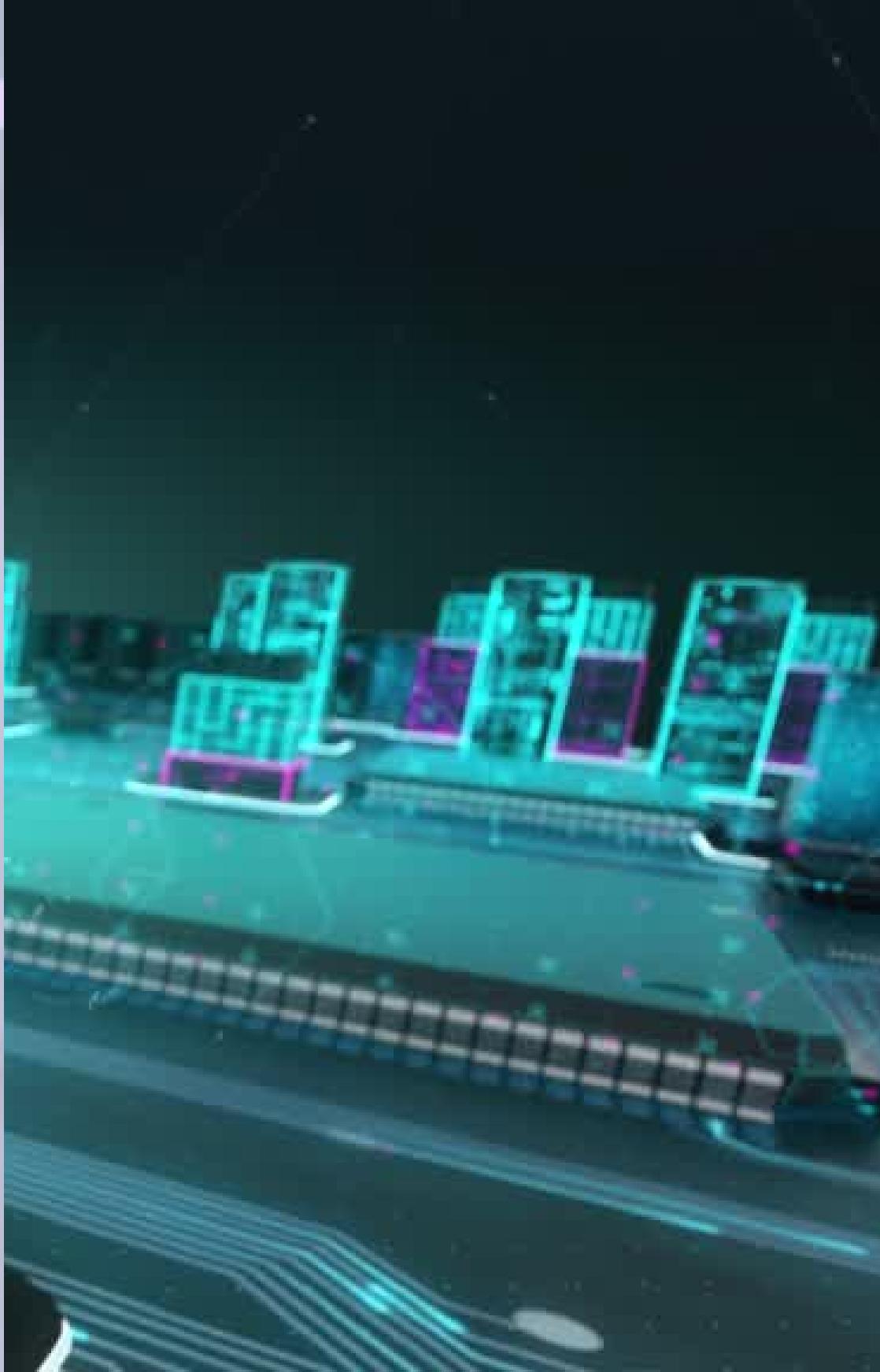


Multi media



Email



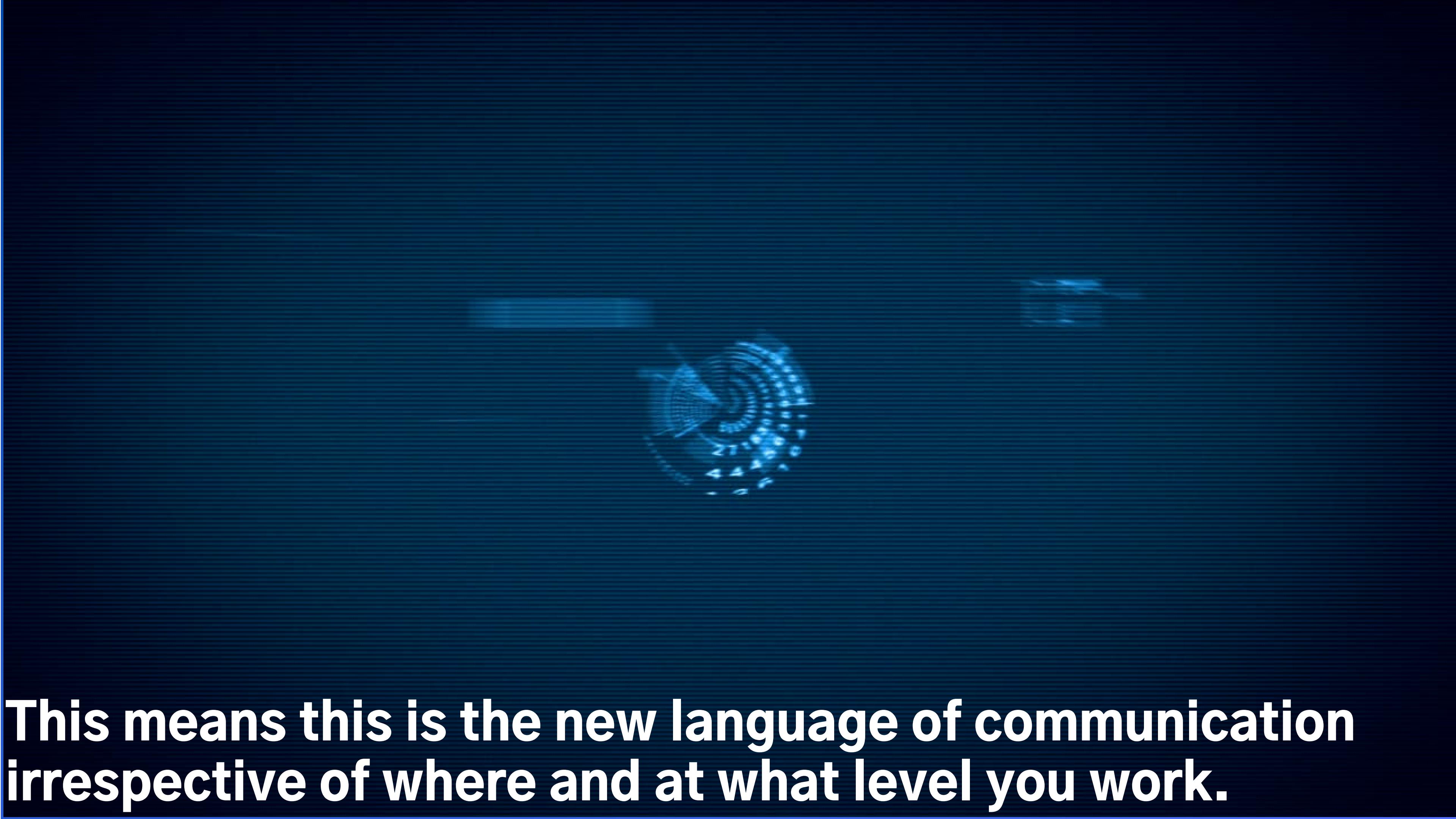


The education system, Monetary systems, financial systems, government system, entertainment, sports, research, technology, citizen systems, security

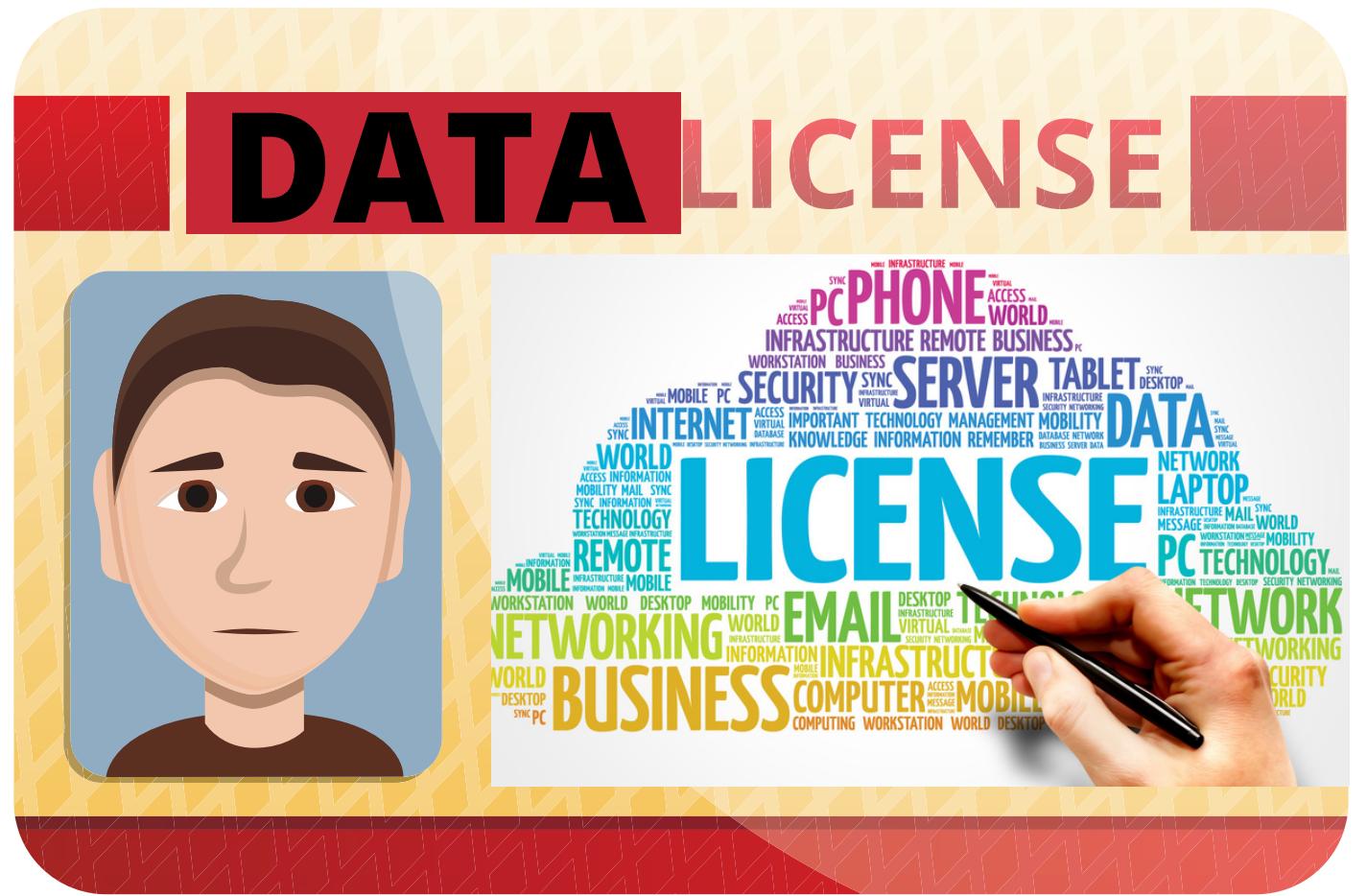
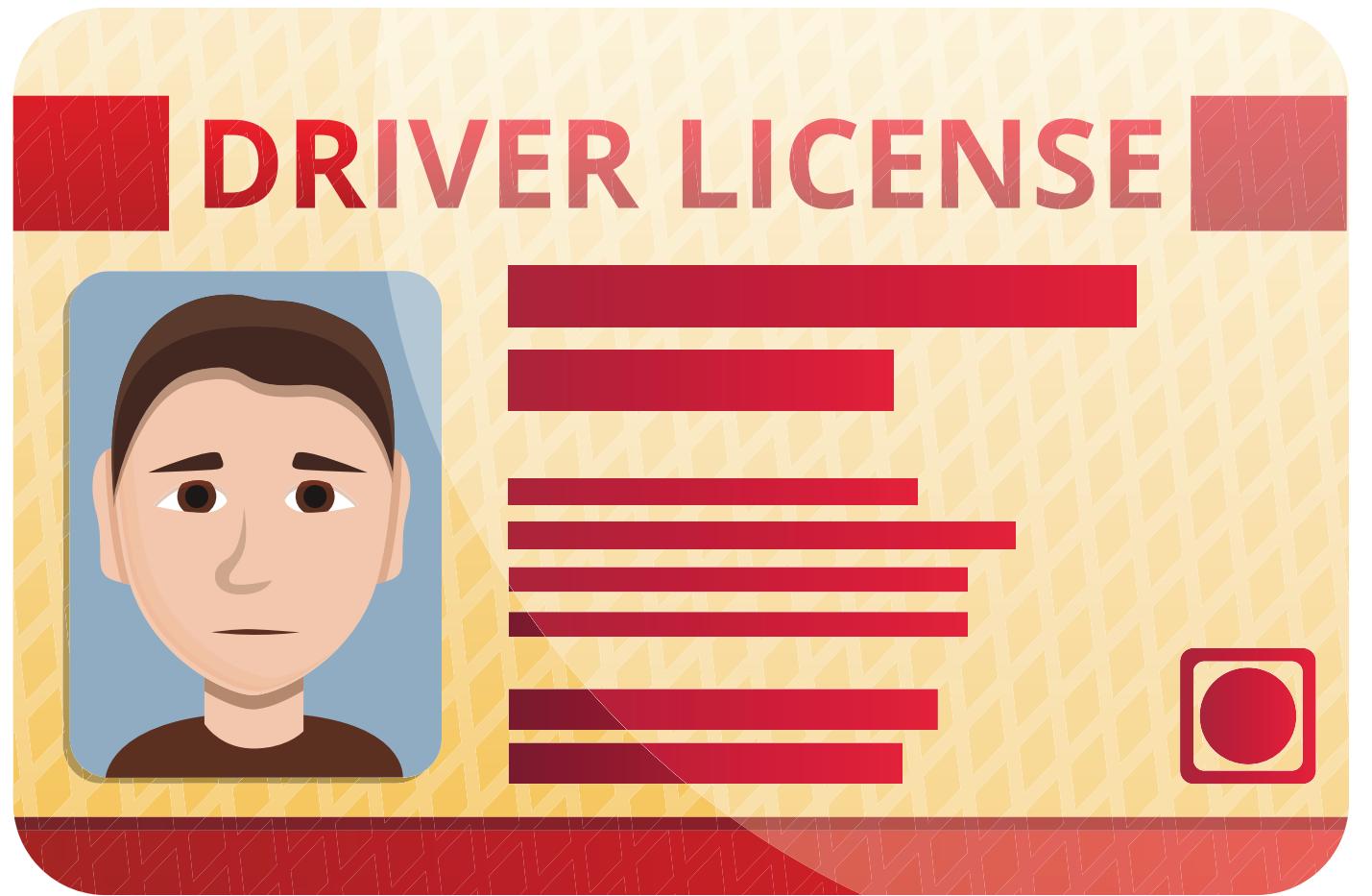
All driven by data.



is data is the fundamental medium of working and progress?



This means this is the new language of communication irrespective of where and at what level you work.



**So many questions must come
to your mind?**

**First might be a feeling of fear or overwhelm
this is so much how can I learn?**

Are there different types of data?

Is data different for different domains and industries?

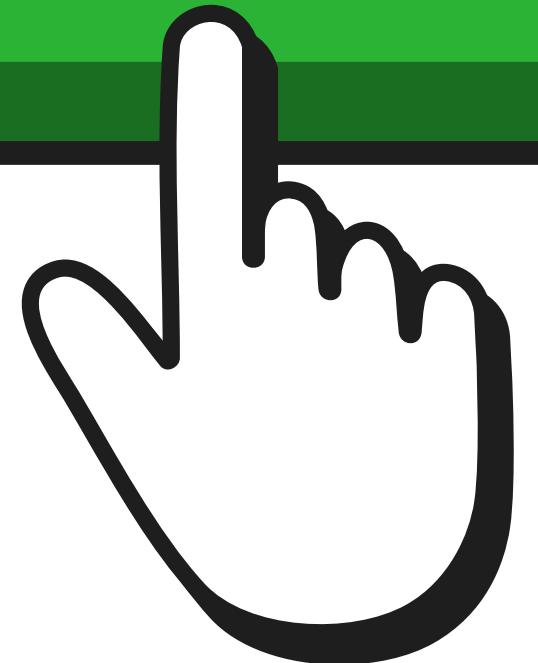
Do I have to know coding?

Will coding help?

How long will it take to learn?

As you must have noticed there are different kinds of data like text, numbers, images, files , videos etc !

MAKE IT EASY



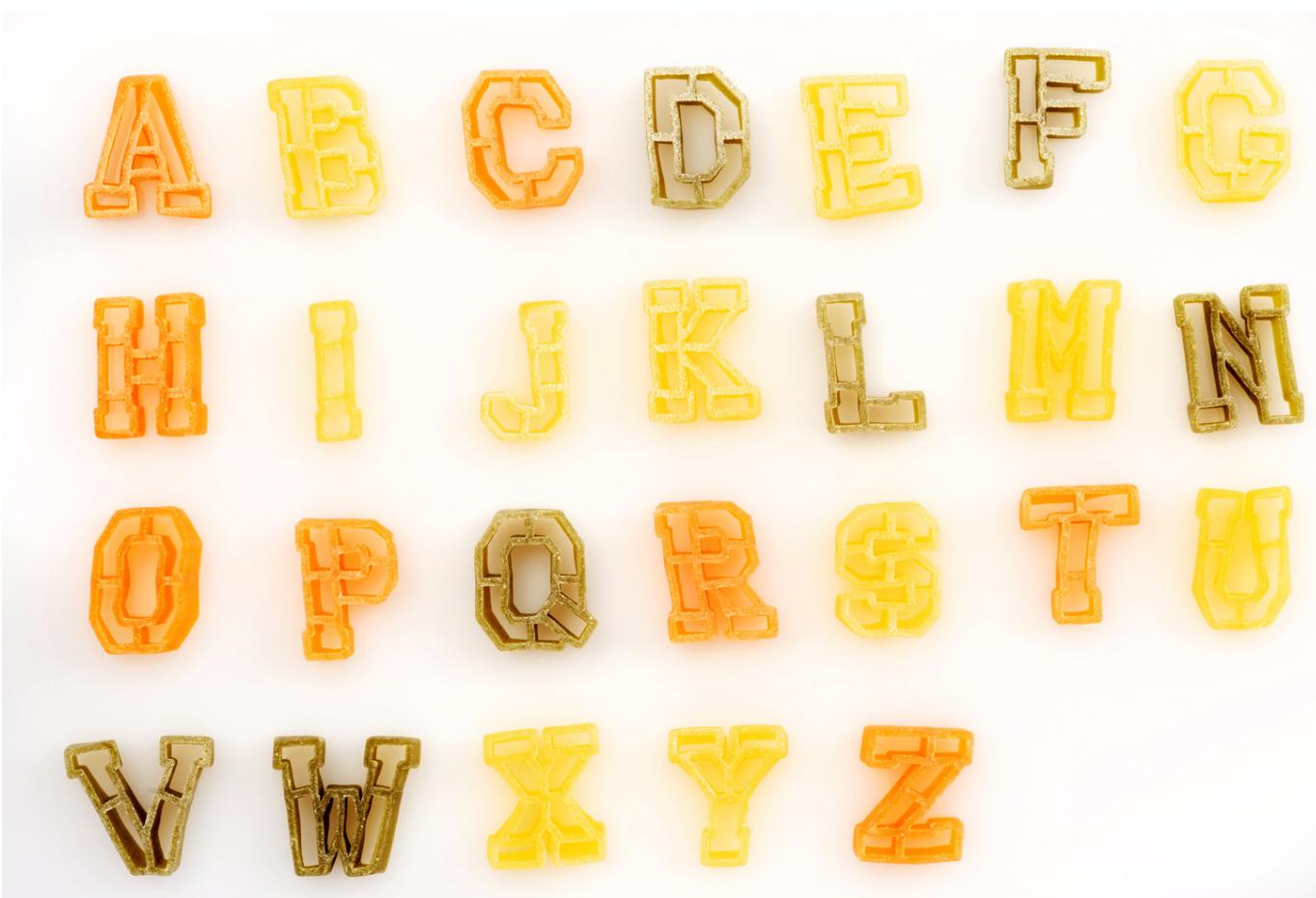
2

Different Types of data

A non technical example to get started to think in the right direction



Pre-defined structure



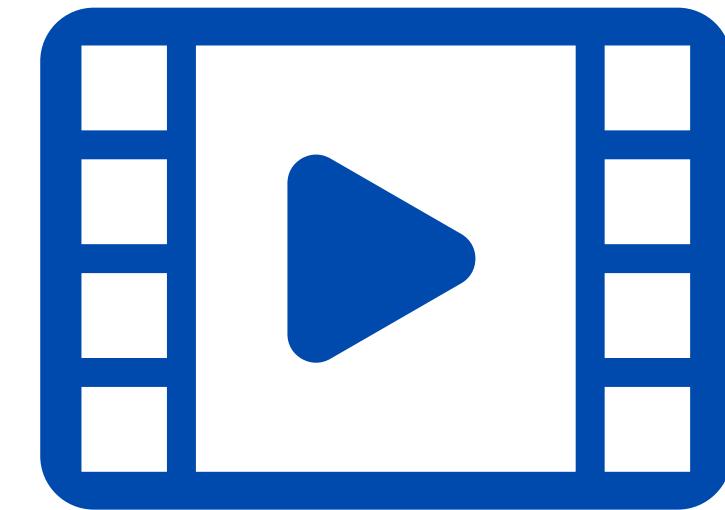
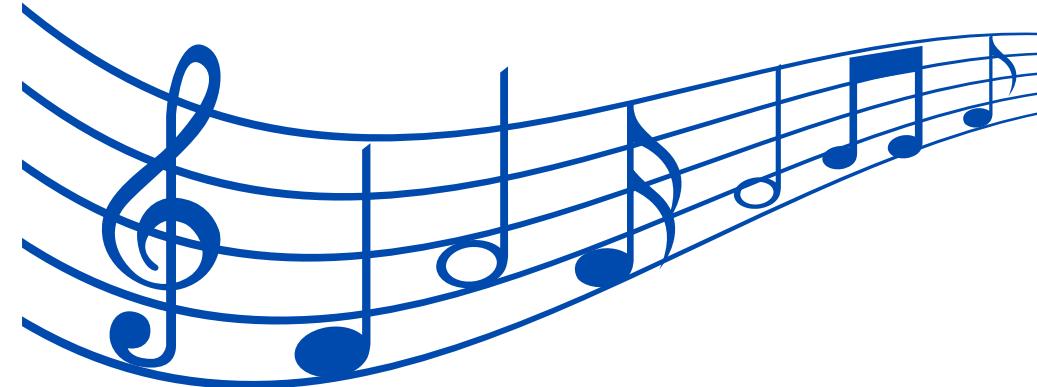
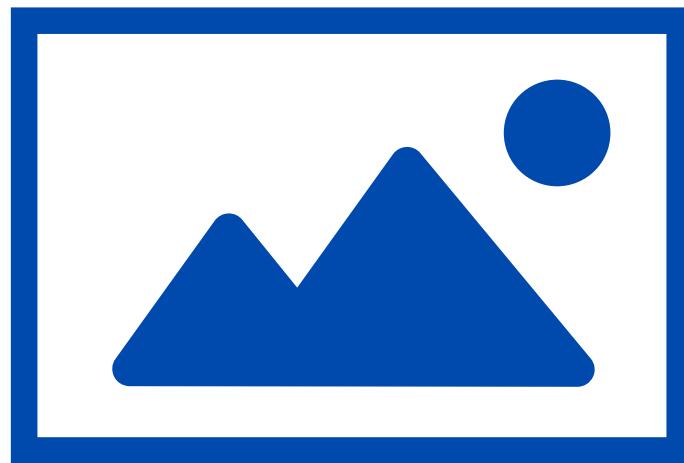
So, for example where I write: “data” – this is understandable and can be comprehended by anybody who has a grasp of the English language.

When I write “11242” – Again this is understandable and can be comprehended by anybody who has a grasp of the number system.

When I write “AAA11123” - A combination of letter and number system both defined

Since English and the number system are well defined (27 letters, 0-9) anything expressed in these terms is understandable – can also be said that these are “structured” because they have the defined (27 letters, 0-9) system.

Now if we were to draw something, or sing something or capture a video or an image?



**These do not have a well-defined structure or a system hence the unlimited possibilities of creativity and creation itself.
These can be categorized as Unstructured.**

A practical angle to structured data

For structured data let's imagine you are making compartments in the kitchen or in the garage.

Design

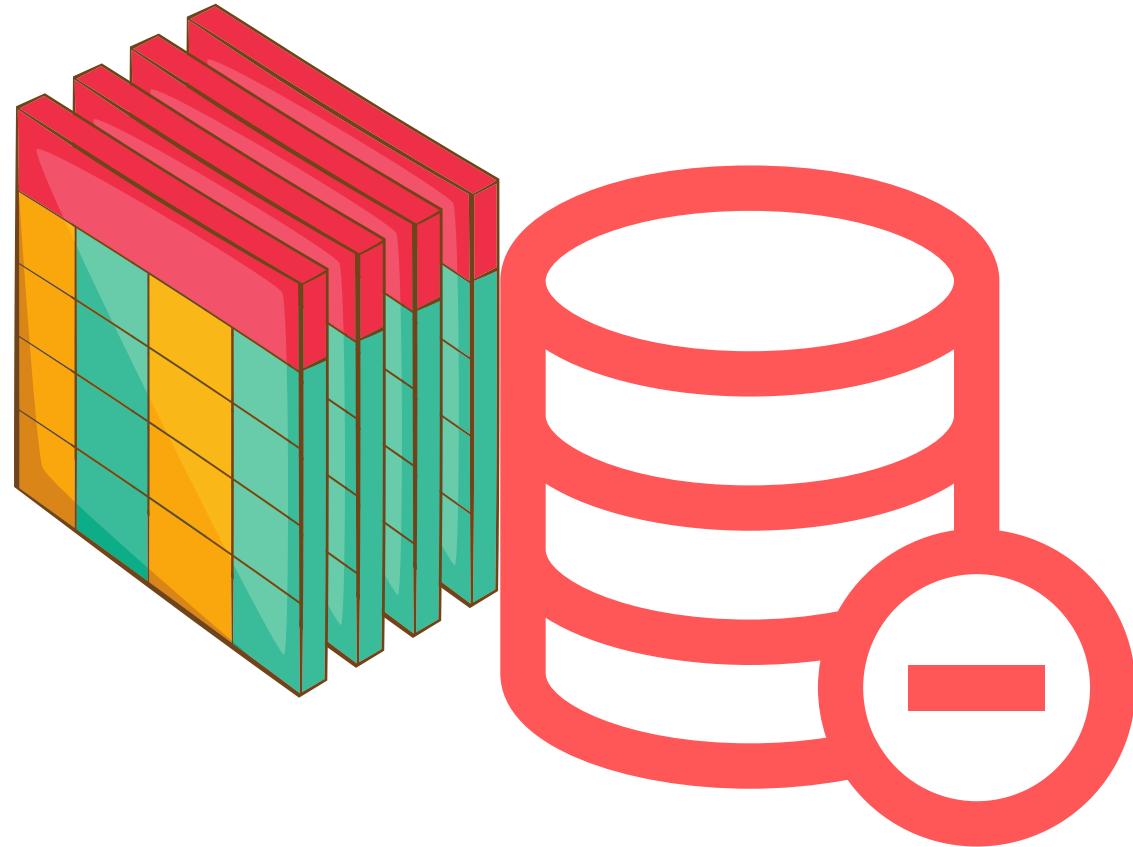
Compartment	Size/Type
Shelf	<= 4 pounds/solids
Jar	2 pounds/liquids
Rack	10 pounds/utensils

Actual

Compartment	Actual	What will happen
Shelf	Onions 3 pounds	No problem
Jar	Frying pan	May fit
Rack	Fill it with 2 pounds of juice	Will spill all over

Compartment	Actual	What will happen
Shelf	Onions 3 pounds	No problem
Jar	Fill it with 2 pounds of juice	No problem
Rack	Frying pan	No problem

Similarly, database or data store is designed in a certain way for a certain purpose with certain specifications in other words limitations.



Unstructured Data example

Imagine a storage facility and you hire a big space for yourself and lets say for the sake of argument it can handle liquids.



Structured (Not exhaustive)

- Numbers, decimals, time
- Letters within the defined length
- Combinations within the defined length

Unstructured (Not exhaustive)

- Images
- Files like PDF
- Videos
- Body of text unlimited - Like books, blogs, tweets etc

Technical Examples: Structured

- Person/ company information (Name, Address, ID, sales, discount, rating, etc.) in any industry healthcare, transport, tech, finance, etc.)
- Financial information (Price, discount, profit, loss, variance, ratios, calculations, formulas, etc.) in any industry healthcare, transport, tech, finance, etc.)
- Social security information
- Market research (price comparisons, market value, financial projections)
- Asset Information (Type, location, value, age, last serviced, etc.)
- IoT device data – Temperature and humidity monitor
- And many more as long as the information has a defined standard structure and has a limitation of length or number of characters.

Technical Examples: Unstructured

User and machine generated

- Facebook/ Instagram posts and videos.
- Medical imaging
- Digital Photos and images
- Lab, research reports
- Invoices
- Books
- Audio files like MP4
- Text and chat messages
- Different file formats
- Even email

Let's Go

Stock Price

Structured

Unstructured



Let's Go

Stock Price

Structured



Unstructured

Let's Go

Stock Price

Structured



Unstructured

Let's Go

CT Scan

Structured

Unstructured



Let's Go

CT Scan

Structured

Unstructured



Let's Go

CT Scan

Structured

Unstructured



To be able to comprehensively describe an entity or object across different industries and domains, you will have to use both structured and unstructured data.

- Student information: Along with vital structured information and Image will also be included
- Financial Information: Along with financial structured data, charts and simulations might be included.
- Medical information: Along with vital structured information, medical scans, reports and patient image might be included

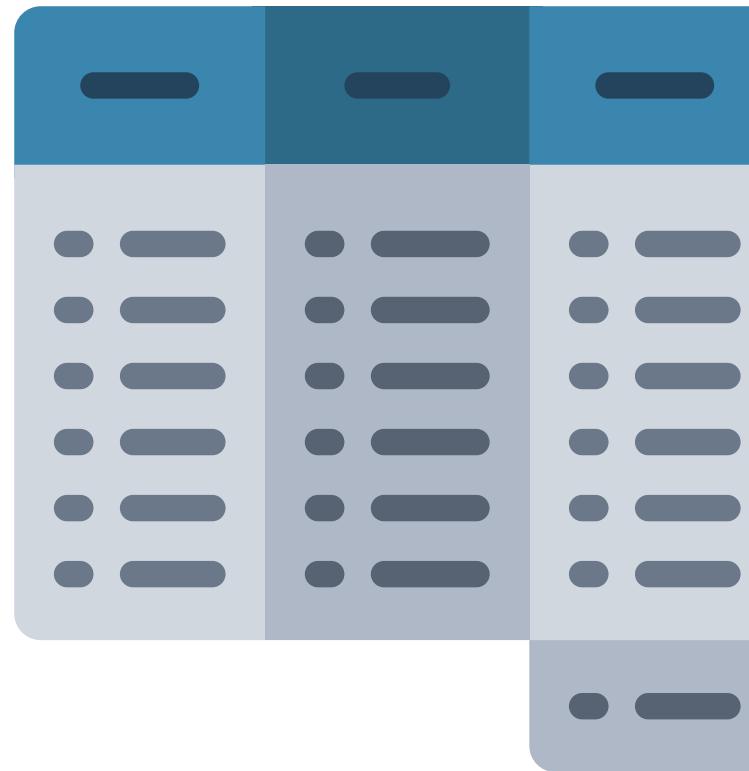
What is data?

3

Now coming to actual data

To understand this lets first build a pictorial table and start to uncover the information and the data types.

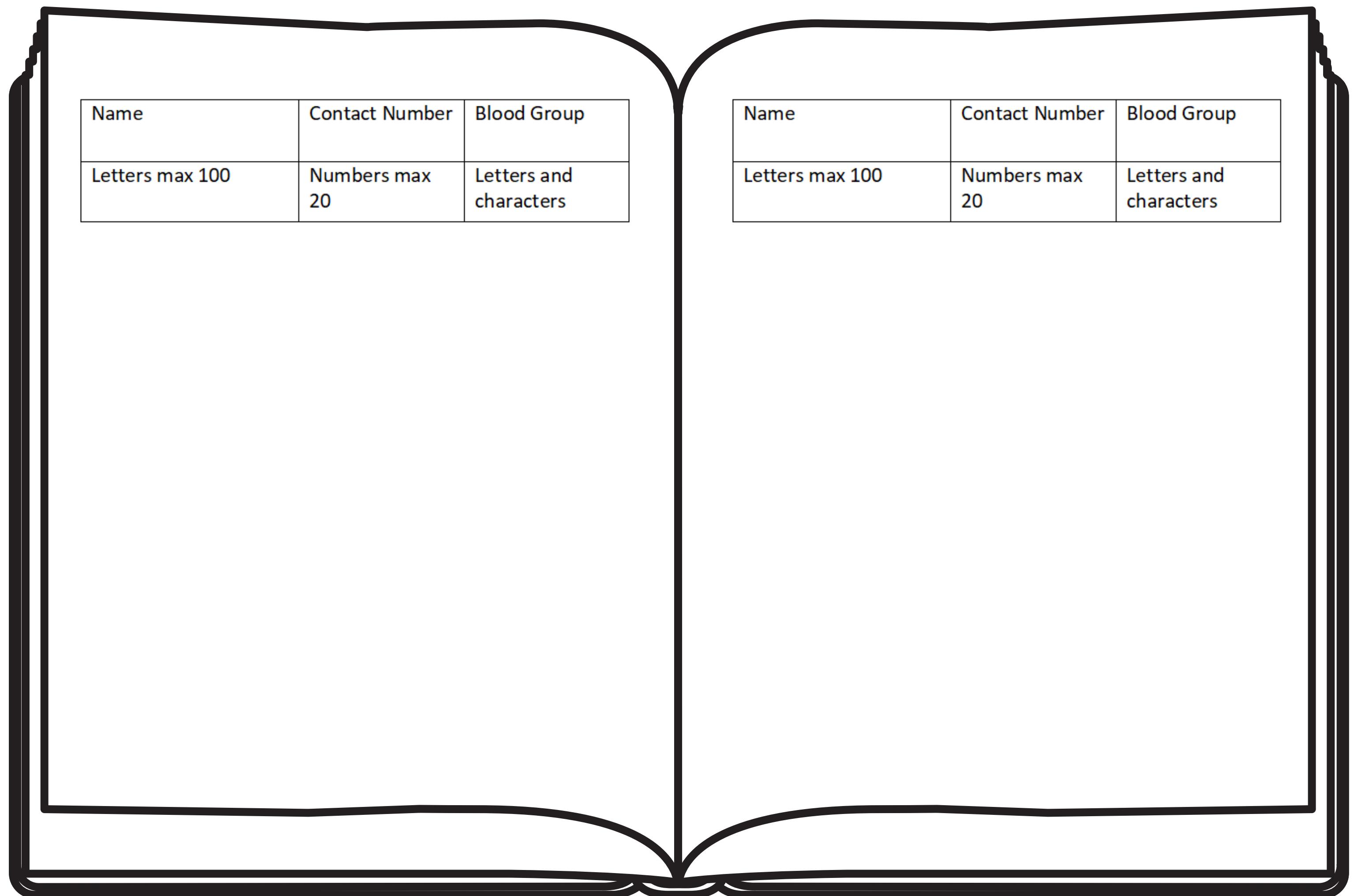
Let's build a database table on paper. Today because of low and no code platforms it is as simple as just naming columns and telling the database what kind of information will be stored in it and how



Let's say we have students in a class. A total of 26

So, imagine a register which captures this information.

So, information about the student's name, Parent's name(s), address, contact number, blood group, height, weight, allergies, emergency contact etc.



Let's take a subset of this information

Name	Contact Number	Blood Group
Juan	217-084-0655	B+
Mark	523-845-9745	B+
Lorie	845-985-3698	A
Juan	217-986-2497	A-

As you can notice we have 2 Juans?

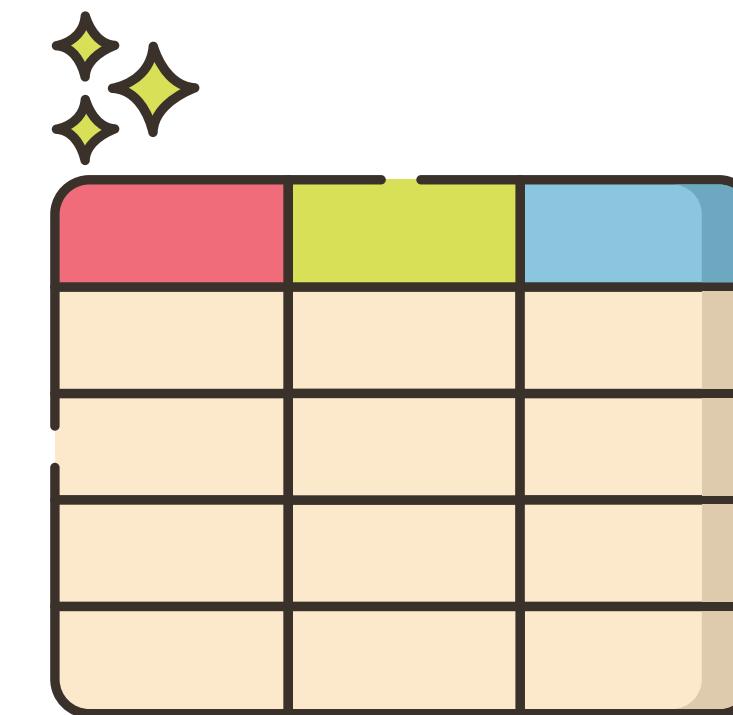
One main property of a table is you need to have a primary identifier (key) to be able to use it effectively.

Also, in general we always give each record a number.

Roll Number	Name	Contact Number	Blood Group
1	Juan	217-084-0655	B+
2	Mark	523-845-9745	B+
3	Lorie	845-985-3698	A
4	Juan	217-986-2497	A-

You can identify each Juan with a unique roll number.
The roll number here is a Unique identifier and hence a Primary key.

Remember registers have a limited space in each column as you can see in this picture.



Similarly, when you define a database column you need to define the type and length of the column

Let's say for this specific class

- Roll Number – Numbers (should be able to capture up to 35 students)
- Name – Letters (should be able to capture up to 100 characters, just to be safe)
- Contact number – This is a standard format so 10 digits or numbers
- Blood Group – Should be able to capture letter and a special character like (-/+)

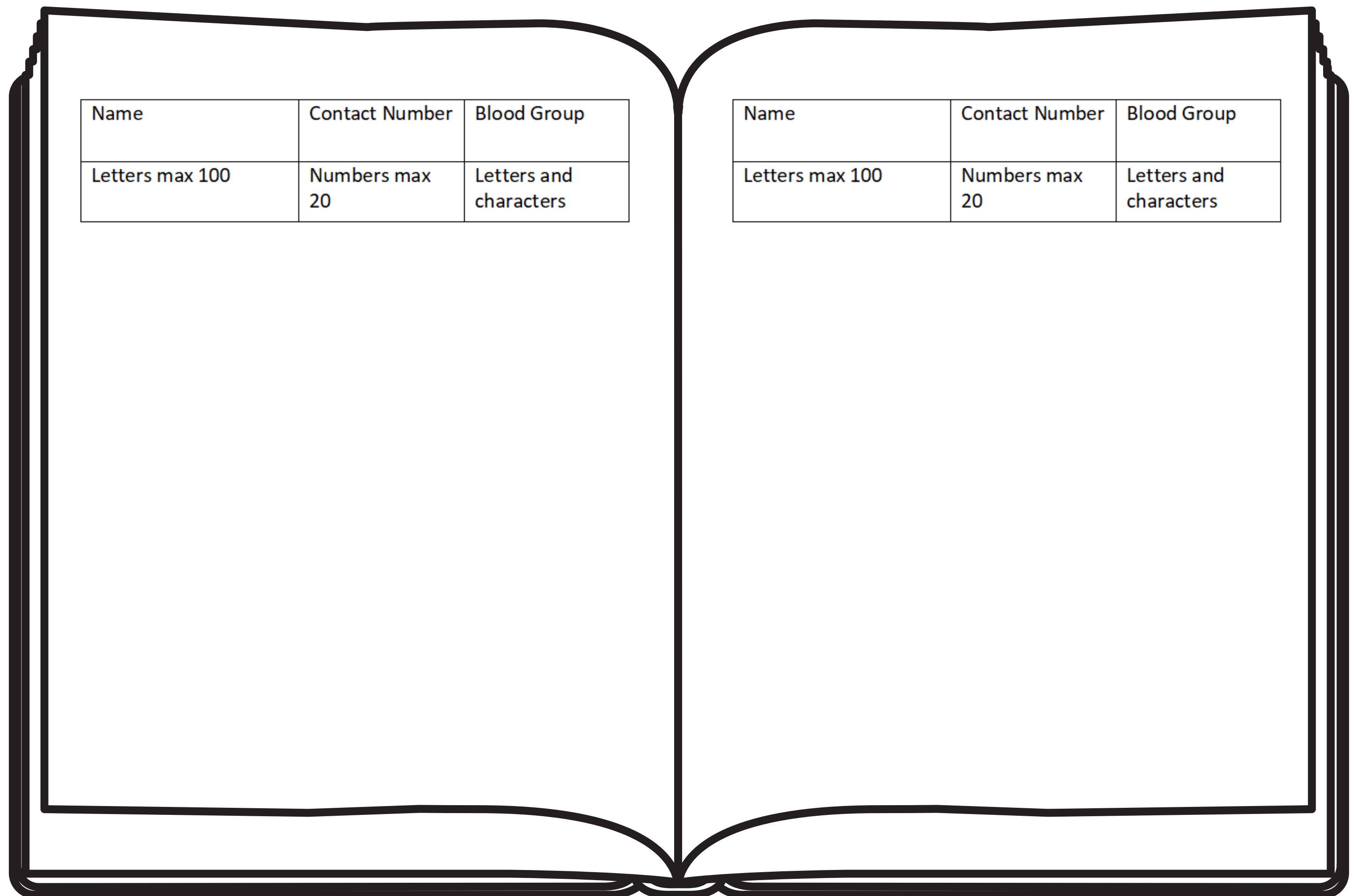
In the database world these are the predefined data type families related to our requirements.

Roll Number – Numbers (should be able to capture up to 35 students)	INT
Name – Letters (should be able to capture up to 100 characters, just to be safe)	VARCHAR
Contact number – This is a standard format so 10 digits or numbers	INT
Blood Group – Should be able to capture letter and a special character like (-/+)	VARCHAR

In the database world these are the predefined data type families related to our requirements.

VARCHAR String (0 – 255)

INT Integer (-2147483648 to 2147483647)



Let's add some more vital information like date of admission and current GPA

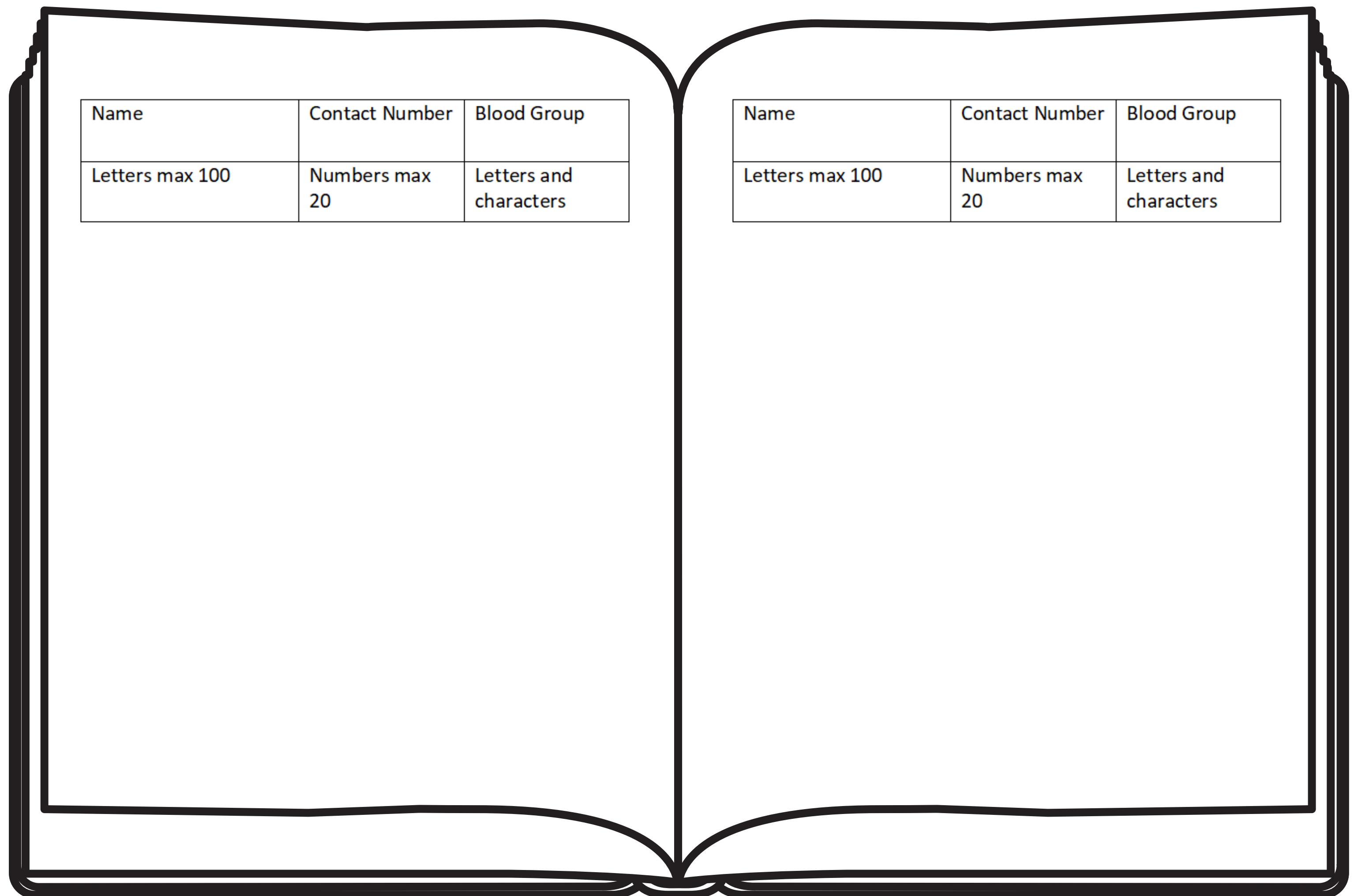
Date of admission	Should be able to hold date values e.g 08/12/2019	DATE
Current GPA	Should be able to hold decimal values 3.53	Decimal

These data types are for structured and each data type has many sub types to make the storing and retrieval efficient.

INT
VARCHAR
TIME
DATE
DECIMAL

Unstructured Data





Unstructured data can be handled by some RDBMS as well like SAP HANA with BLOB data type.

There are databases specifically designed for unstructured data like Hadoop, Cassandra etc.

The data is organized under labels much like folders

Let's Go

Stock Price

Structured

Unstructured



Let's Go

Stock Price

Structured



Unstructured

Let's Go

Stock Price

ABC Inc | Data Type

125.75 | Decimal

Structured



Unstructured

Let's Go

CT Scan

Structured

Unstructured



Let's Go

CT Scan

Structured

Unstructured



Let's Go

CT Scan

Hospital | Data Type

Image | BLOB/DB specific

Structured

Unstructured



Categories of data

Quantitative: something which can be measured, counted, added

Qualitative: Categories, types, groups, locations.

Travel Data

Store	Type	Qty	Price/Piece	Total sale price
NYC	T-Shirt	3	20	60
SFO	Pants	1	25	25
NYC	Pants	2	30	60
SFO	Scarf	2	15	30
SFO	Jacket	1	60	60
NYC	Overcoat	1	250	250
NYC	Shoes	2	125	300

- Store: Qualitative – Location
- Type: Qualitative – Category
- Qty: Quantitative
- Price/Piece: Quantitative
- Total Sale price: Quantitative

Let's Go

Genre of movies you like?

Quantitative

Qualitative



Let's Go

**Genre of movies you
like?**

Quantitative

Qualitative



Let's Go

Genre of movies you like?

Action, Sci fi, Suspense

Quantitative

Qualitative



Let's Go

**What are the sales
for this quarter?**

Quantitative

Qualitative



Let's Go

**What are the sales for
this quarter?**

Quantitative



Qualitative

Let's Go

**What are the sales for
this quarter?**

It could take any numeric value

Quantitative



Qualitative

Quick Check

Age:

Place:

SSN:

Coffee price:

Type of sports you like:

Which city are you from?

What is your GPA?

What are the sales for this quarter?

What is the email open rate?

Which family of viruses does COVID 19 belong to?

Which material is used to make LCD panels?

Quick Check

Age: Quantitative

Place: Qualitative

SSN: Quantitative

Coffee price: Quantitative

Type of sports you like: Qualitative

Which city are you from? Qualitative

What is your GPA? Quantitative

What are the sales for this quarter? Quantitative

What is the email open rate? Quantitative

Which family of viruses does COVID 19 belong to? Qualitative

Which material is used to make LCD panels? Qualitative

4

Data Classification for Reporting Systems

Qualitative – Dimension

Quantitative – Measure

Store	Type	Qty	Price/Piece	Total sale Price
NYC	T-Shirt	3	20	60
SFO	Pants	1	25	25
NYC	Pants	2	30	60
SFO	Scarf	2	15	30
SFO	Jacket	1	60	60
NYC	Overcoat	1	250	250
NYC	Shoes	2	125	300
		SUM = 12	AVG :75	SUM = 785

Typical Mapping

Store	Dimension
Type	Dimension
Qty	Measure
Price	Measure
Total Sale price	Measure

Let's Go

Time

Dimension

Measure



Let's Go

Time

Year, month , quarter

Dimension

Measure



Let's Go

Sales

Dimension

Measure



Let's Go

Sales

Transaction which can be summarized or simply added by a dimension like time or country or city

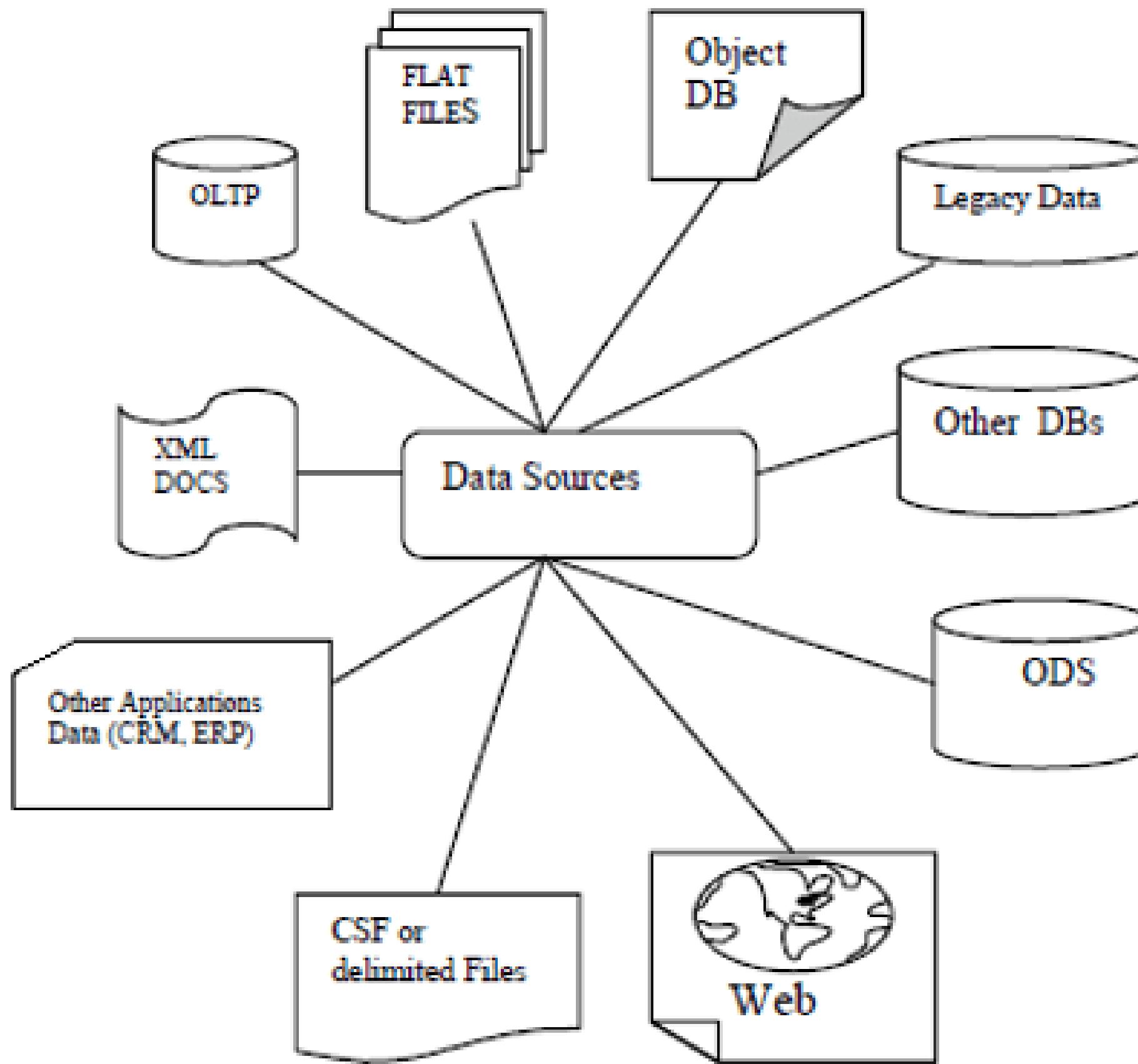
Dimension

Measure

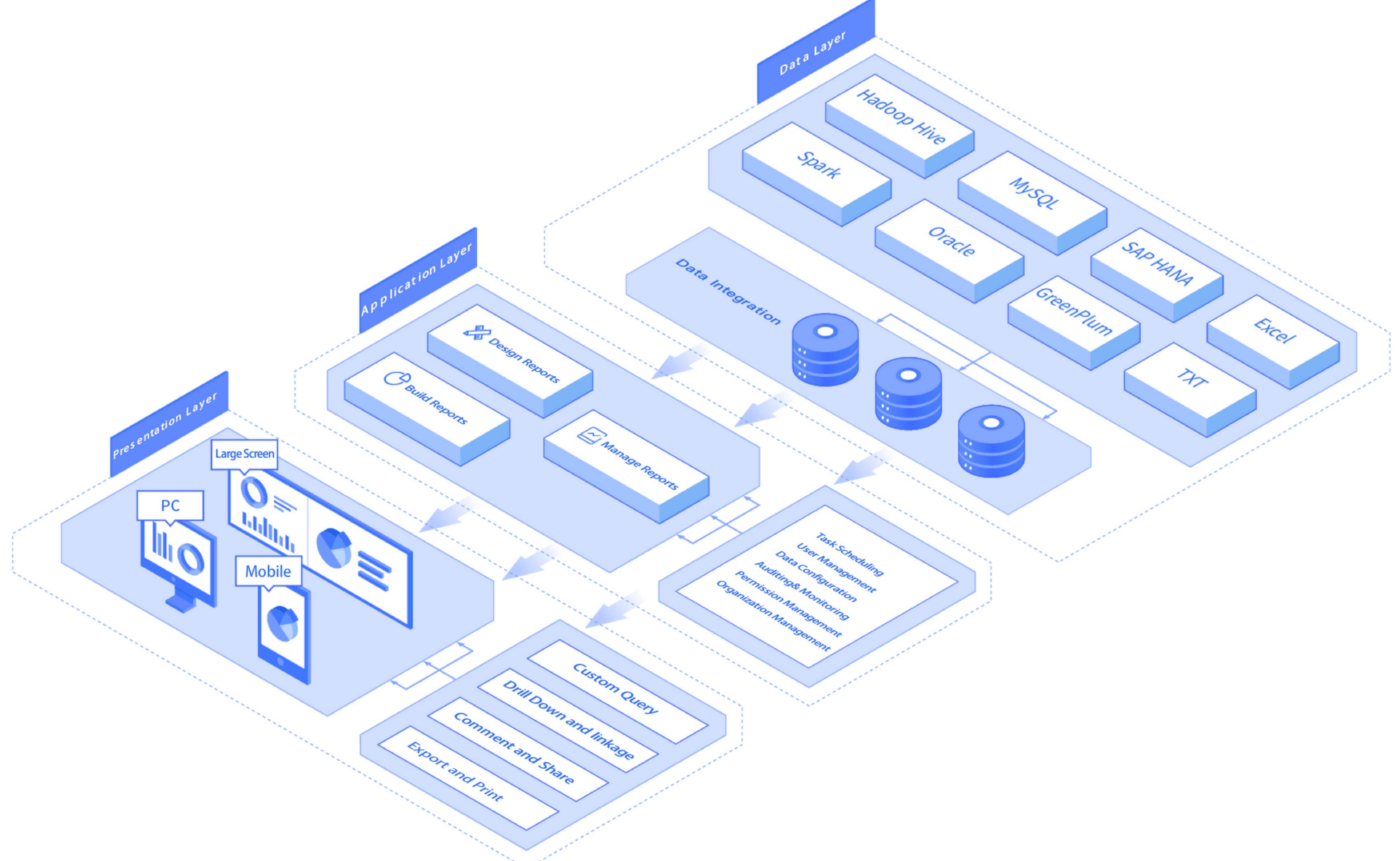


5

Sources of data for an Enterprise







Data Sources

OLTP
Databases



Enterprise
applications



Third-party



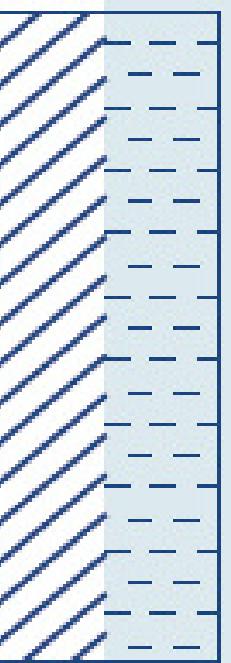
Web apps



Other



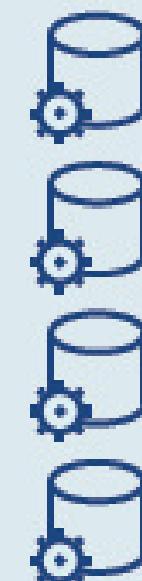
ETL or ELT



EDW



Logical
Datamarts



Data Lake



BI / Analytics