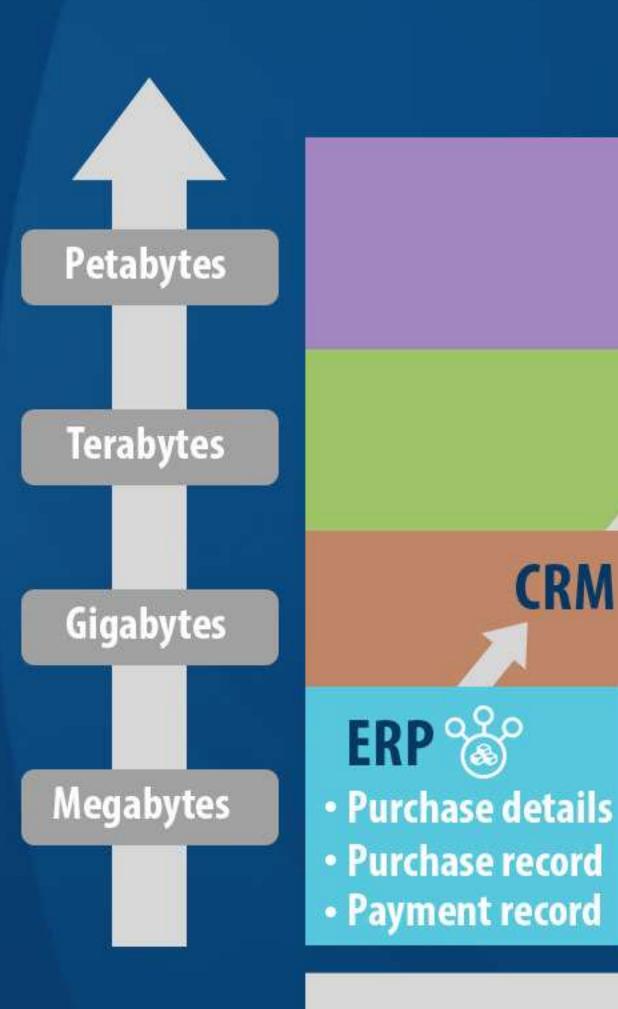
AGENDA

- Data Landscape
- Types of Data
- Classification
- Extracting Information from Data

DATA LANDSCAPE





- Customer Touches
- Support Contacts

AB testing

- Dynamic Pricing
- Affiliate Networks
- Search Marketing
- Behavioral Targeting
- Dynamic Funnels
- Web logs
- Offer history

- User Generates Content
- Social Interactions & Feeds
- Spatial & GPS coordinates
- External Demographics
- Business Data Feeds
- HD Video, Audio, Images
- Speech to Text
- Product/Service Logs
- SMS/MMS
- Sentiment
- Sensors/RFID/Devices
- Mobile Web
- User Click Stream

What is Big Data?

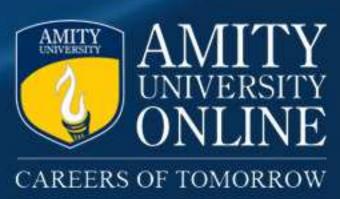
Broad term for datasets so large or complex, that traditional data processing applications are inadequate.



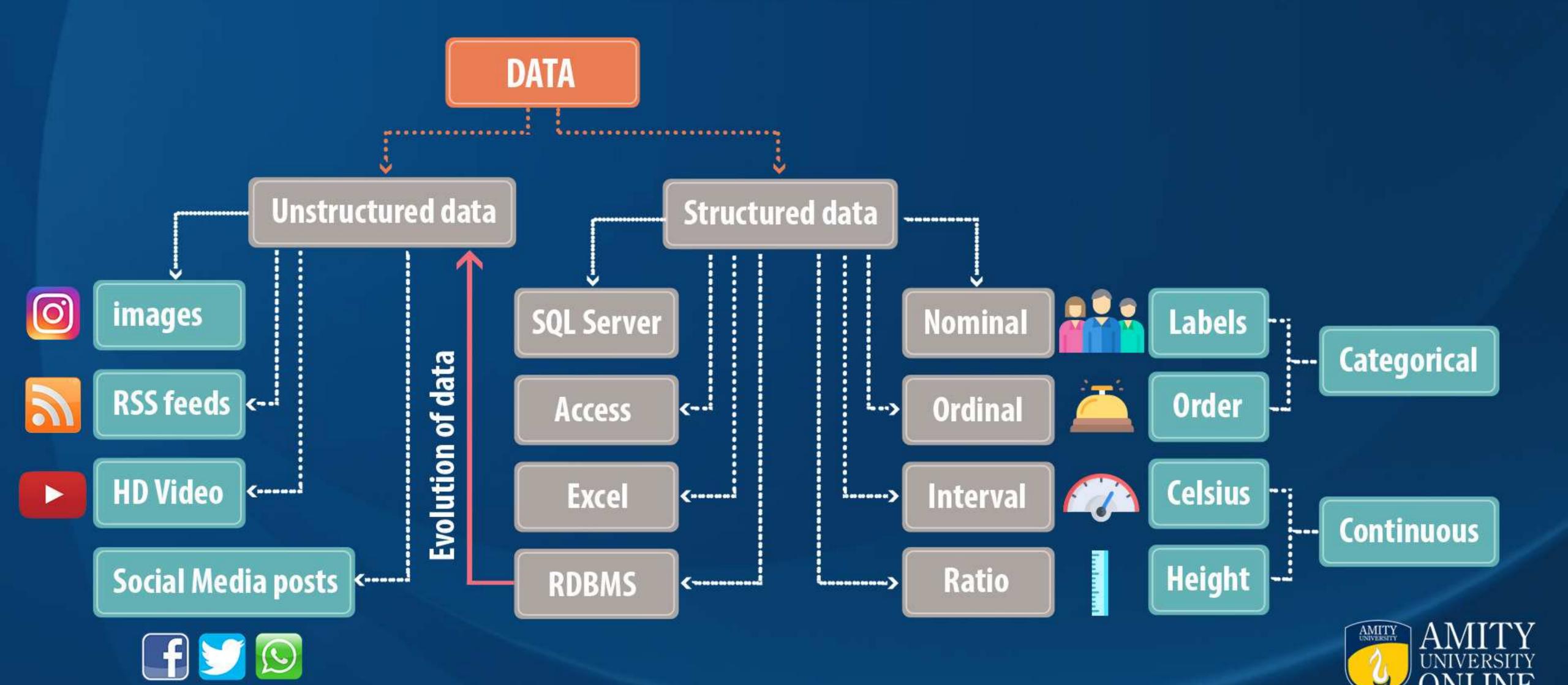
> 1950

> 2010

> 2012



TYPES OF DATA



CAREERS OF TOMORROW

STRUCTURED DATA: THE NOMINAL SCALE

Nominal scales are used for labeling variables, without any quantitative value. "Nominal" scales could simply be called "labels."

Here are some examples, below:

What is your gender?

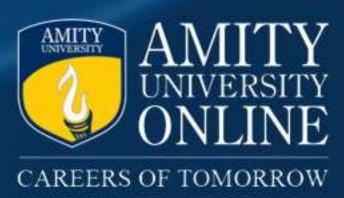
- Male
- Female

What is your hair color?

- Black
- Blonde
- Gray
- Other

Where do you live?

- North of Equator
- South of Equator
- International Space Station
- **Other**



STRUCTURED DATA: THE ORDINAL SCALE

With ordinal scales, it is the order of the values is what's important and significant, but the differences between each one is not really known.

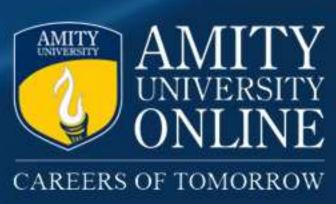
For example, is the difference between "Very Hot" and "Hot" the same as the difference between "Hot" and "Warm"? We can't say!

Socio Economic Class

- Upper
- Upper Middle
- Middle
- Lower Middle

How hot is the water?

- Very hot
- Hot
- Warm
- Room Temperature



STRUCTURED DATA: THE INTERVAL SCALE

Interval scales are numeric scales in which we know not only the order, but also the exact differences between the values.

Example: Celsius scale

The difference between 60 and 50 degrees is a measurable 10 degrees, as is the difference between 80 and 70 degrees.

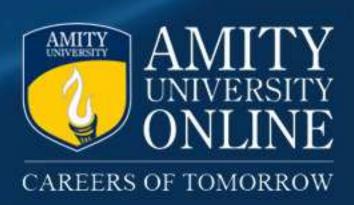


Food for thought

Is 20°C twice as hot as 10°C?

Drawbacks of Interval scale

- No such thing as True Zero
- Ratios are not possible



STRUCTURED DATA: THE RATIO SCALE

Ratio scales are the ultimate nirvana when it comes to measurement scales because they tell us about the order, they tell us the exact value between units, AND they also have an absolute zero which allows for a wide range of both descriptive and inferential statistics to be applied.



Food for thought

Is 20 kg twice as heavy as 10 kg?

Yes!

Example:

- Height
- Weight
- Temperature (on Kelvin scale)

CAREERS OF TOMORROW

 Quantity(mL) of soft drinks consumed in a month

UNSTRUCTURED DATA



Text files and documents

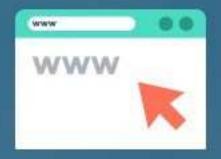


Sensor data





Emails



Server, website and application logs







Social media data

GETTING INFORMATION OUT OF DATA

How many **Frequency Distribution How varied** Dispersion studies Clustering Grouping Data **Time Series Trends** Prediction **Regression Methods Decision Making** Decision Trees, Random forests

