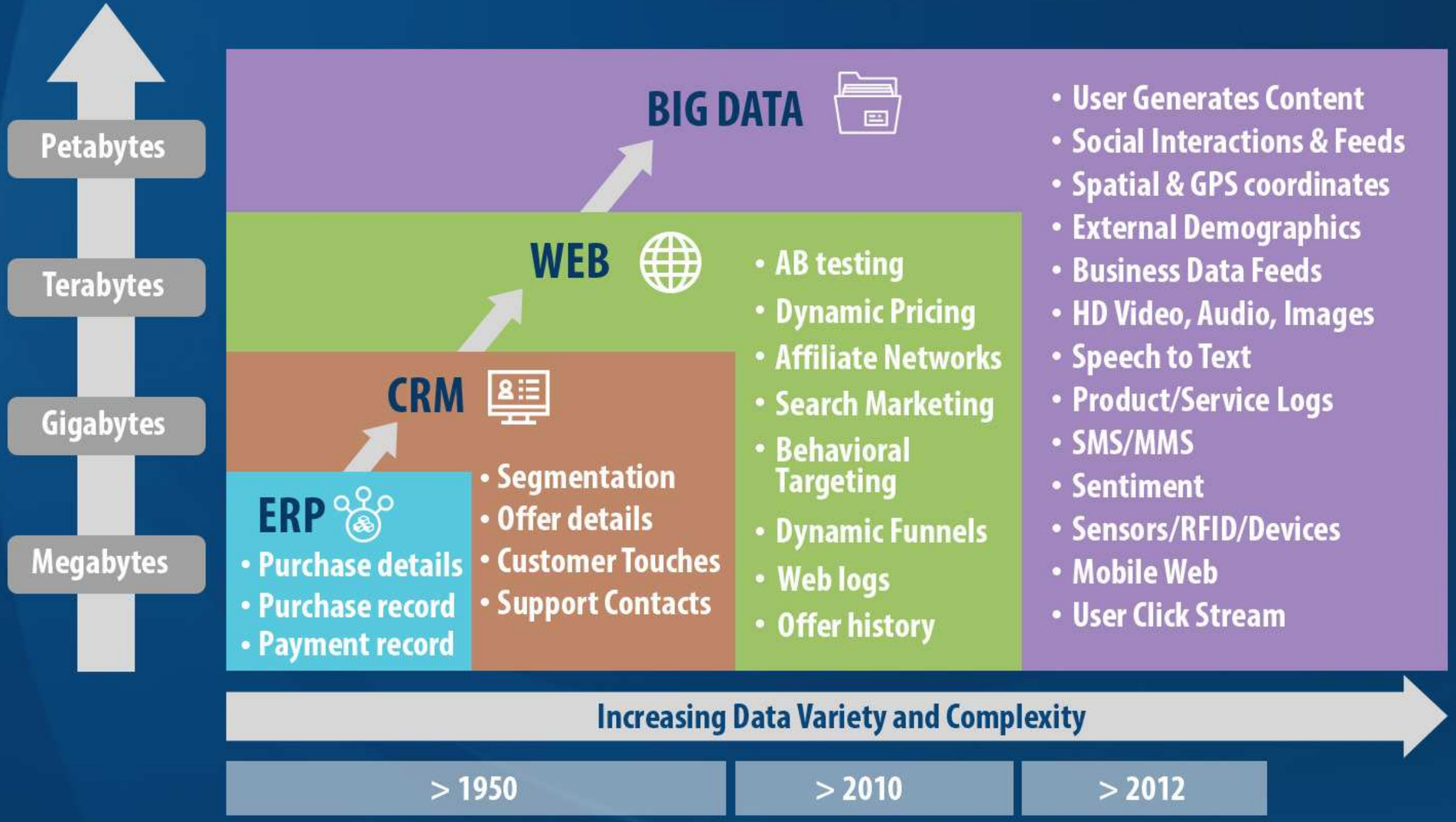
AGENDA

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

DATA LANDSCAPE

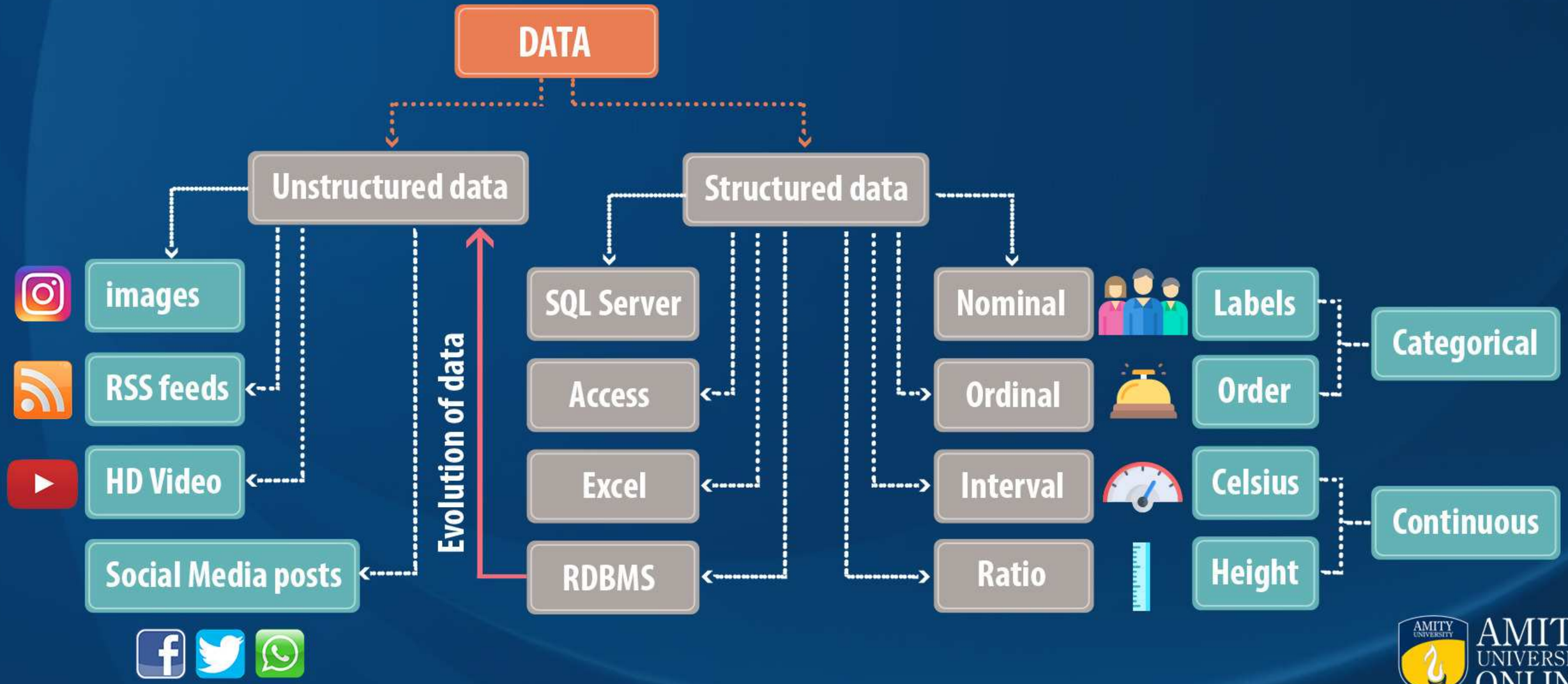


What is Big Data?

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

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

TYPES OF DATA



AMITY UNIVERSITY ONLINE
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

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

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

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

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**

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

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)
- Quantity(mL) of soft drinks consumed in a month

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

UNSTRUCTURED DATA

Text files and documents

Server, website and application logs

Sensor data

Images

Video files

Audio files

Emails

Social media data

GETTING INFORMATION OUT OF DATA

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW