Data Analytics

Agenda

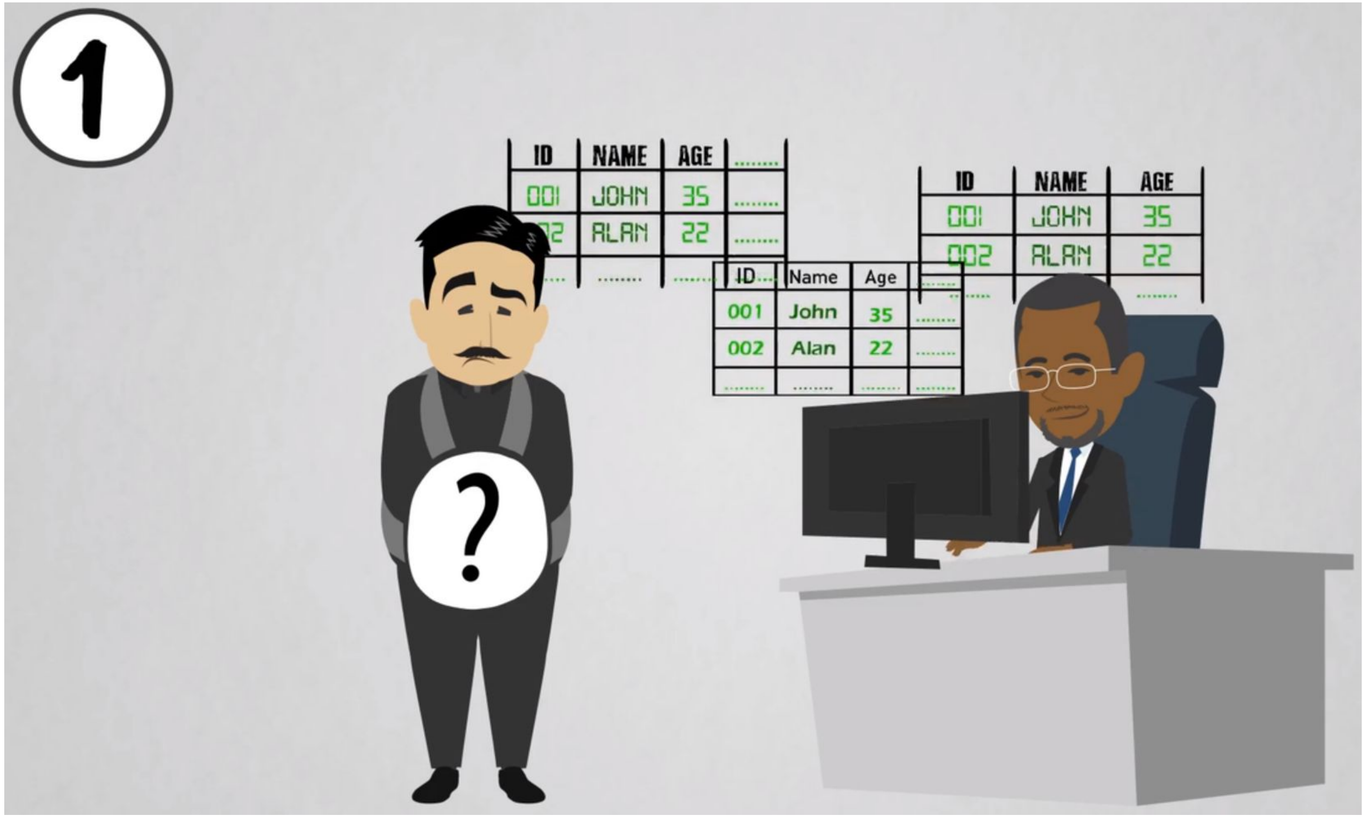
1. Approach a Business Task
2. 365 DataScience Infographic
3. Data
4. Data Science

Approach a Business Task

Scenario 1

The boss

- has read the reports/dashboards
- want you to make some predictions for the firm's outgoing costs over the next year



Scenario 1

The logical way to approach this problem is to:

- gather some relevant data
- then prepare it for analysis

Scenario 2

The boss says

- We have an enormous amount of data
- We don't know what we could do with it but it must be useful
- Can you do something with it, such as:
  - Tell us how we could increase our profit for next year

Scenario 2

In this case

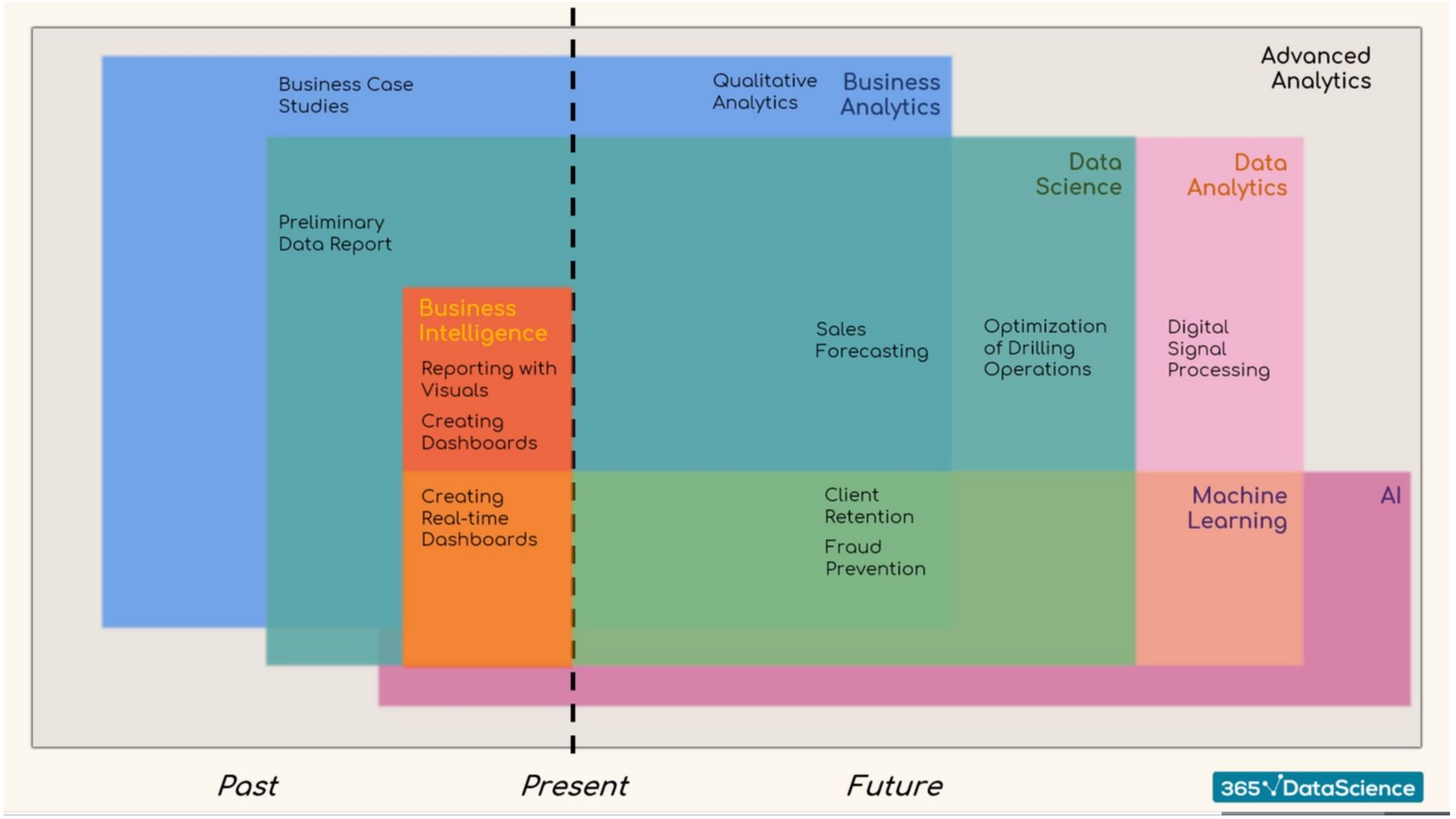
- Having the dataset is the starting point
- You don’t need to collect data to answer a business question
- You can analyse it and apply different analytics tools to extract insights and make forecasts

In both Scenario 1 & Scenario 2

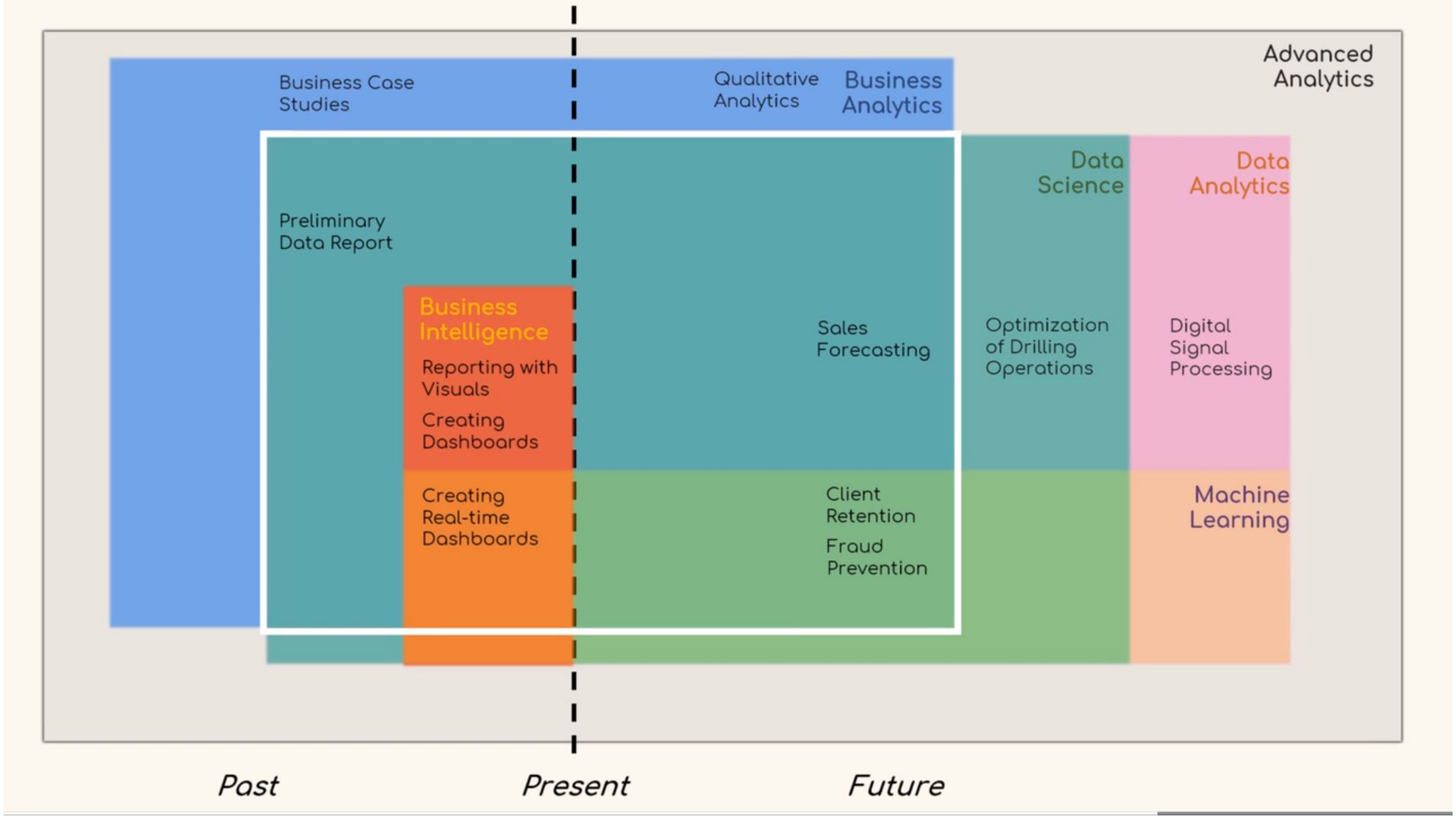
- The solution to any task begins with having a proper dataset
- This must be first on the to-do list
- Only then, we can proceed with
  - further analysis
  - and forecasting

365 Data Science Infographic

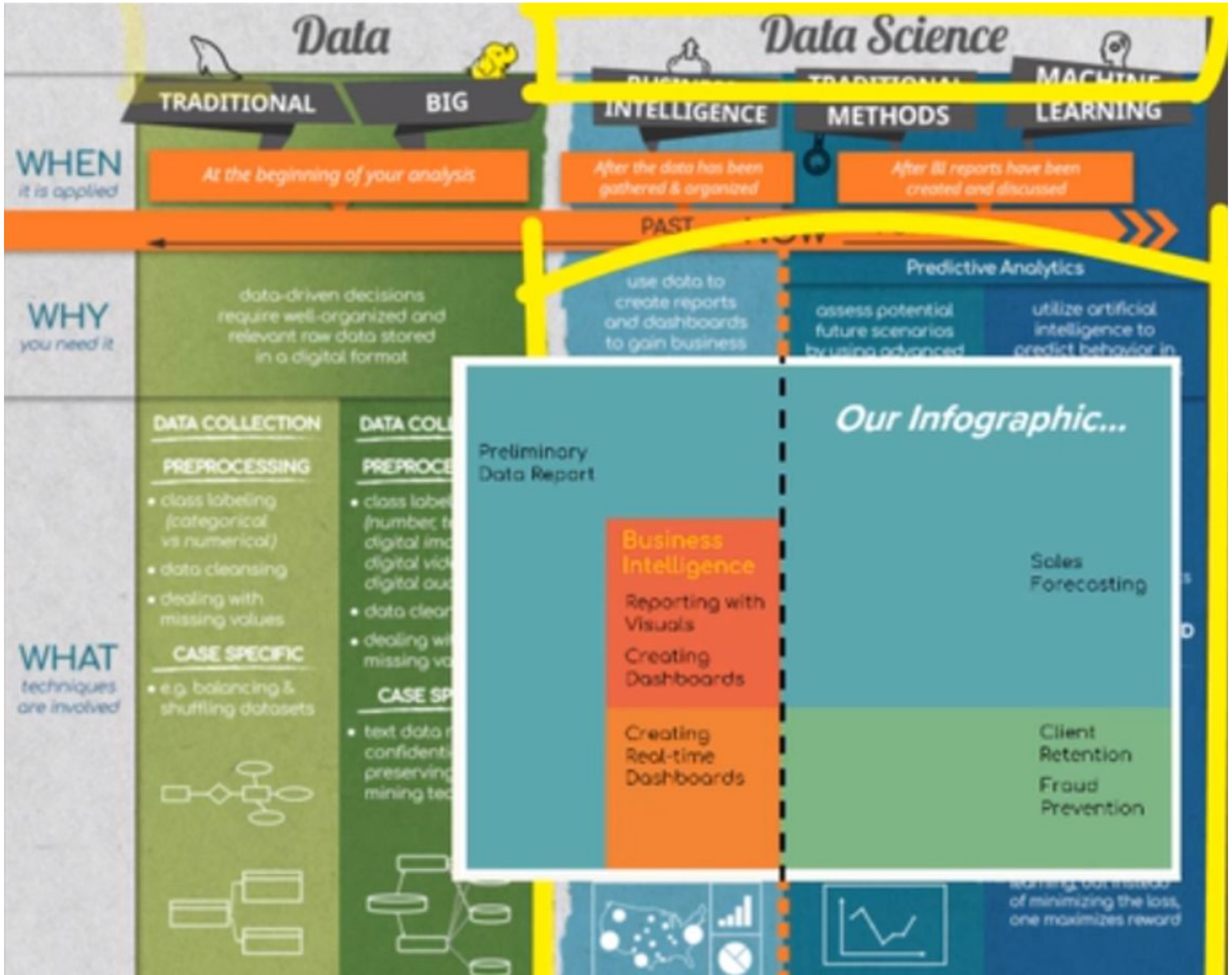
Advanced Data Analytics Diagram



Advanced Analytics - concentrate on Business



Advanced Analytics & its companion Infographic



365 DataScience Infographic Columns

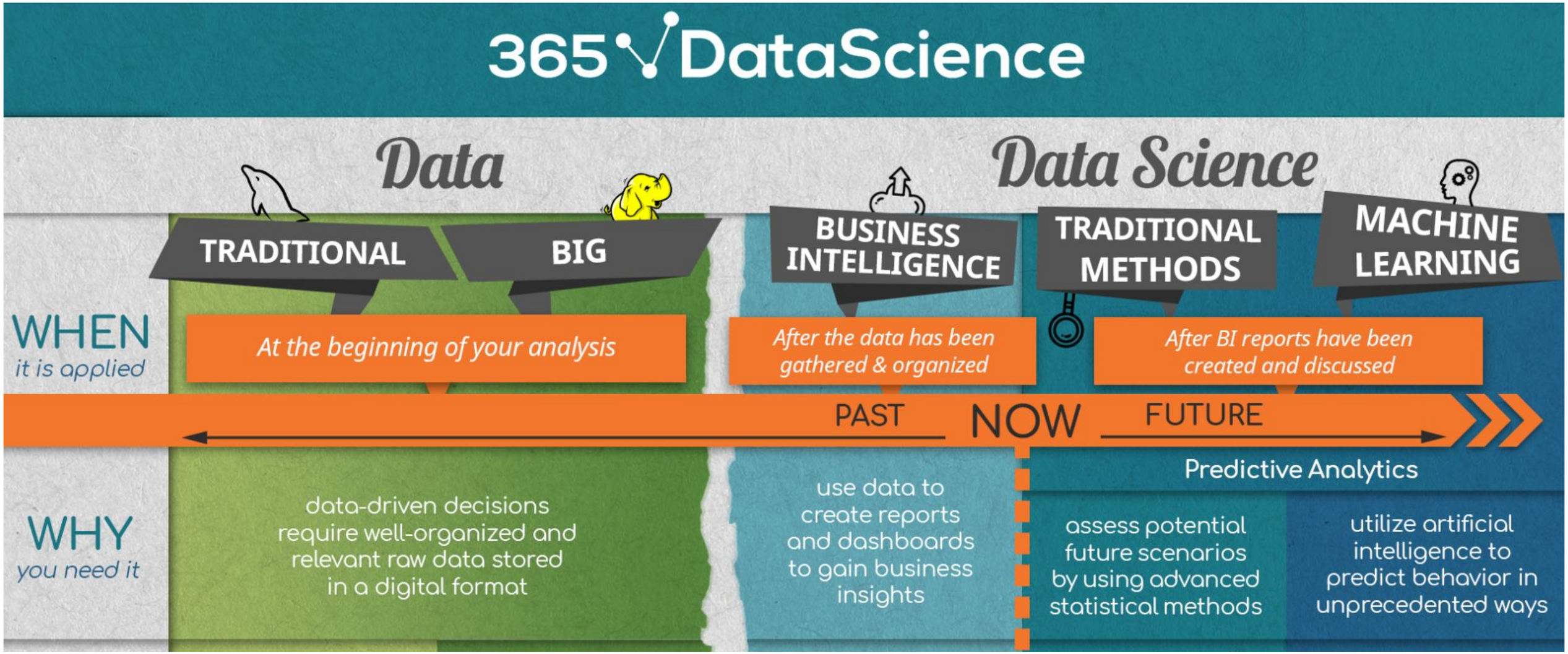
Each describe a stage of solving business task process

1. Working with Traditional Data
2. Working with Big Data
3. Doing Business Intelligence
4. Applying Traditional Data Science Techniques
5. Using Machine Learning Techniques

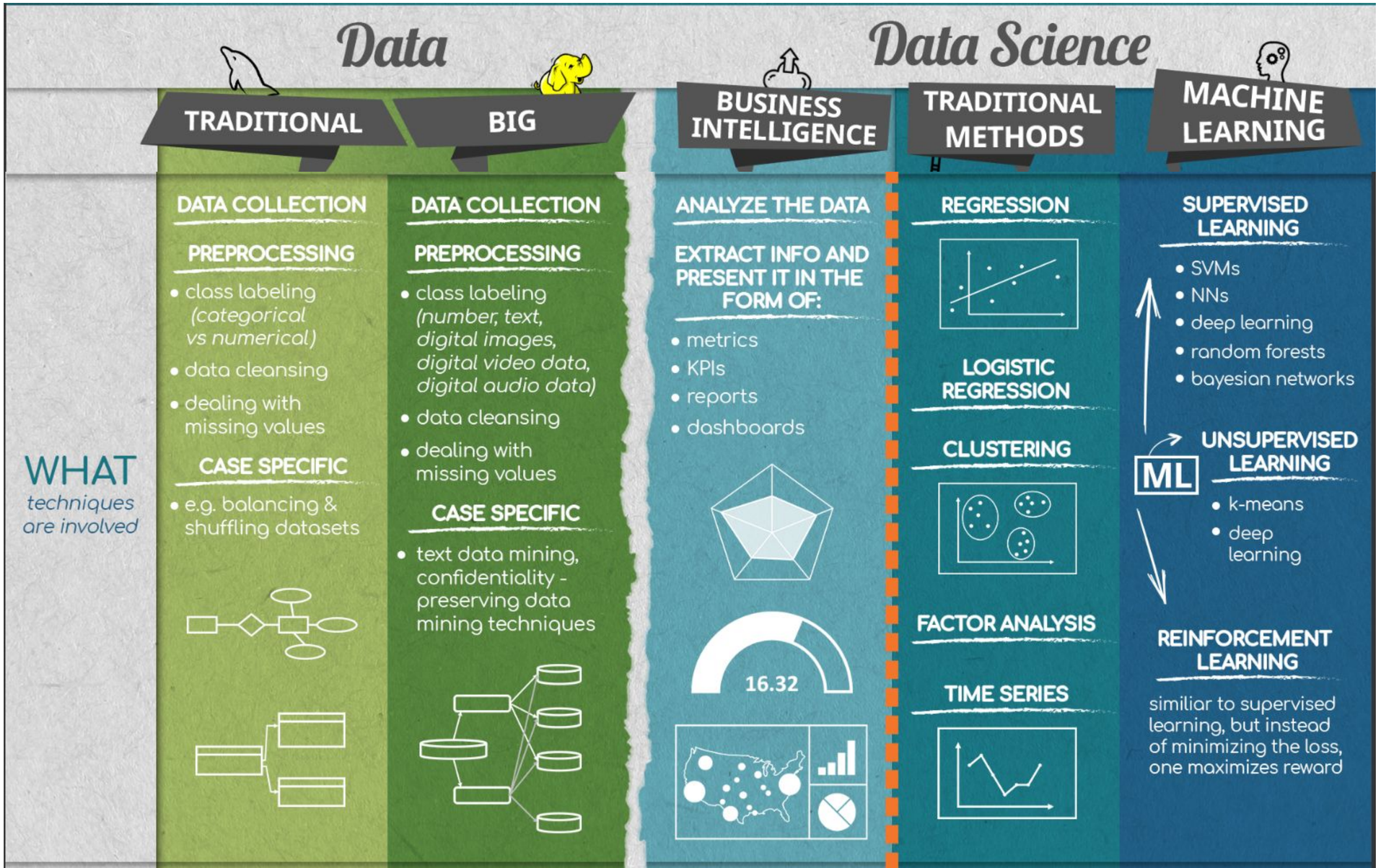
Each Row answer important Question

1. When is this part of the process applied?
2. Why do we need it?
3. What are the techniques related to this activity?
4. Where and in which real-life cases can it be applied?
5. How is it implemented? Using what tools?
6. Who is doing this?
7. What are the common misconceptions about this activity?

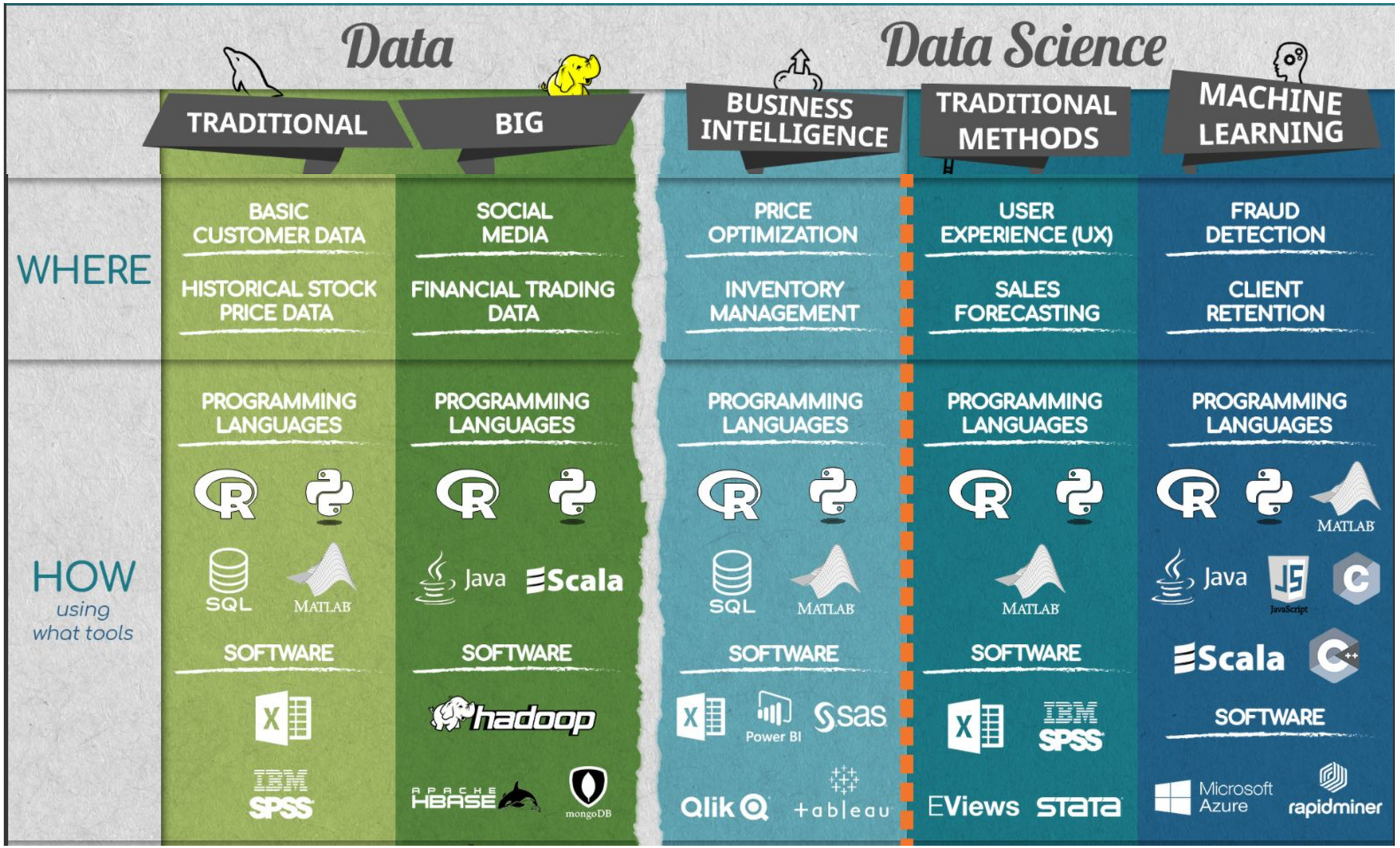
365 DataScience Infographic - Question 1,2



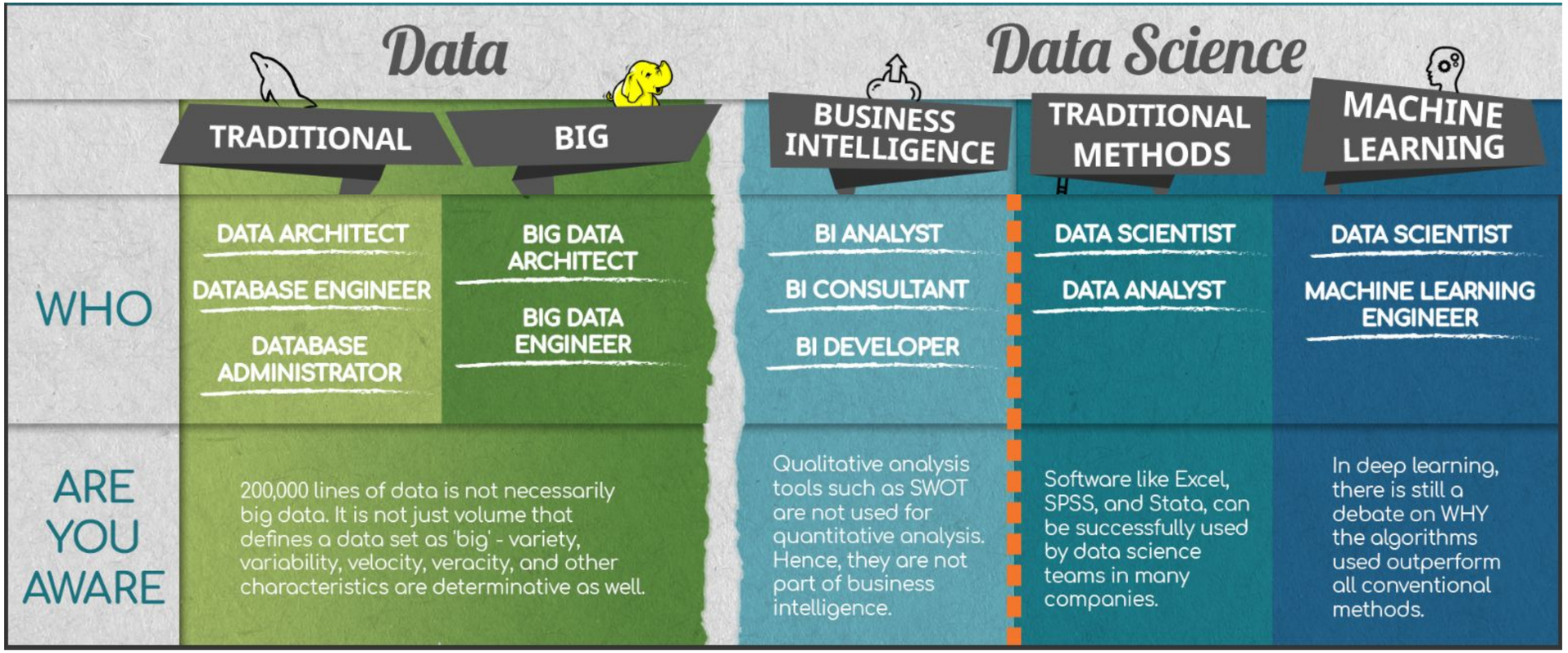
365 DataScience Infographic - Question 3



365 DataScience Infographic - Question 4,5



365 DataScience Infographic - Question 6,7

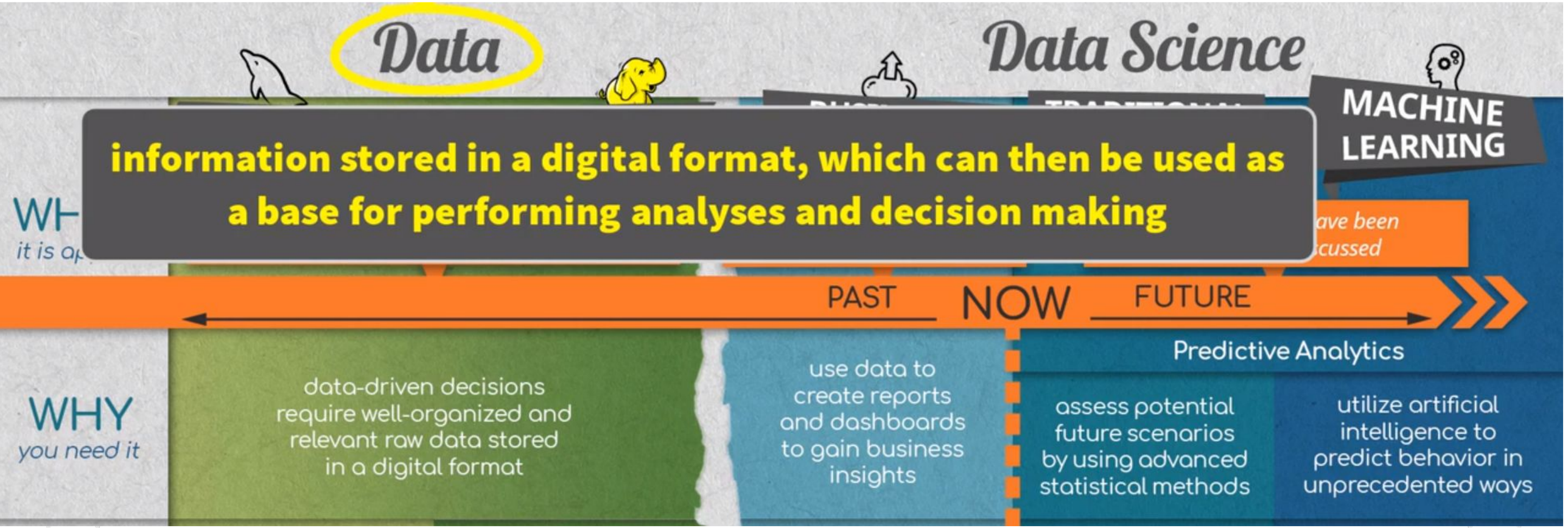


Data

Data Definition

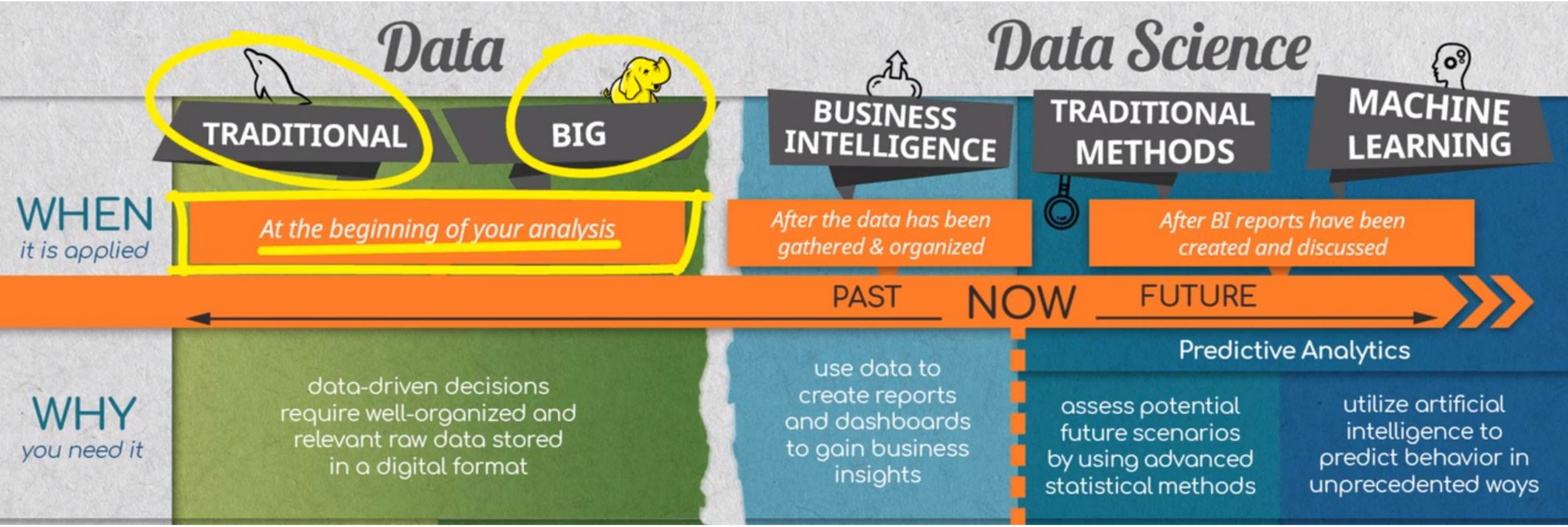
- can be defined as:
  - information stored in a digital format,
  - which can be used as a base for performing **analysis** and decision making

Data



- Dealing with data is the first step
  - when solving a business problem or researching,
  - so it is important to know what you are looking at

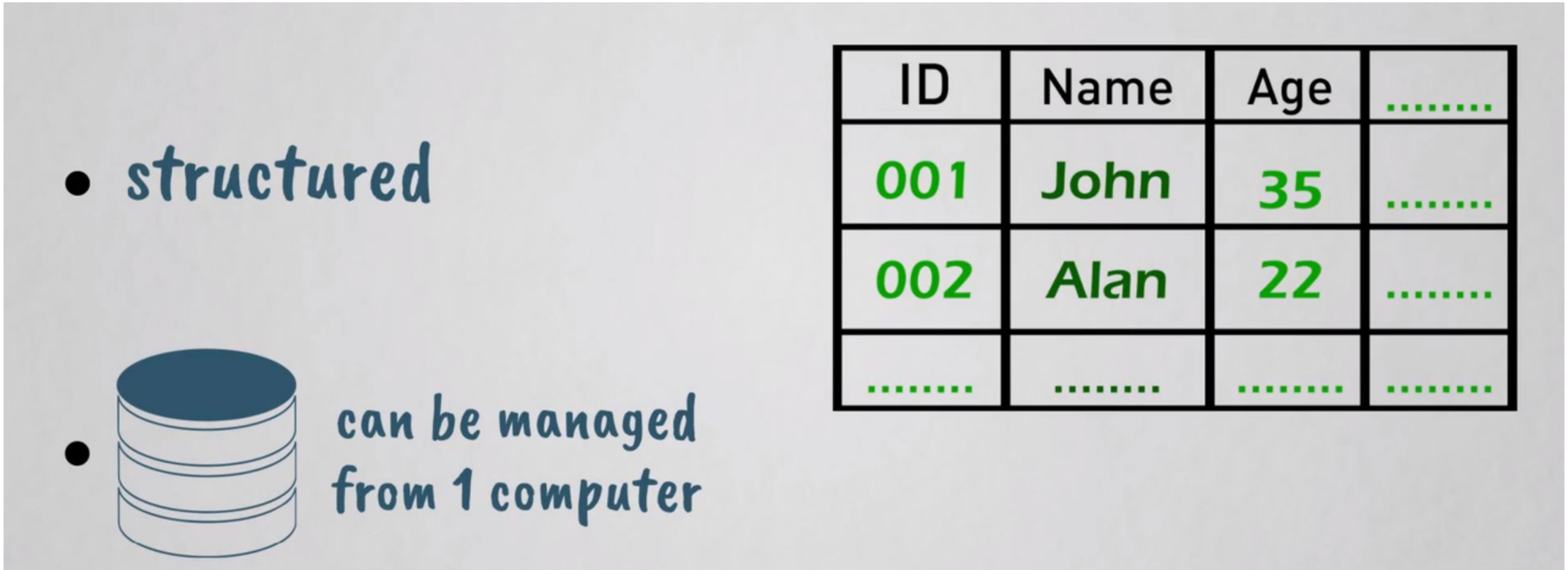
Data (Traditional & Big)



- Either Data or Big Data
  - it is your first port of call for business problem-solving,
  - so it is important to know what you are dealing with

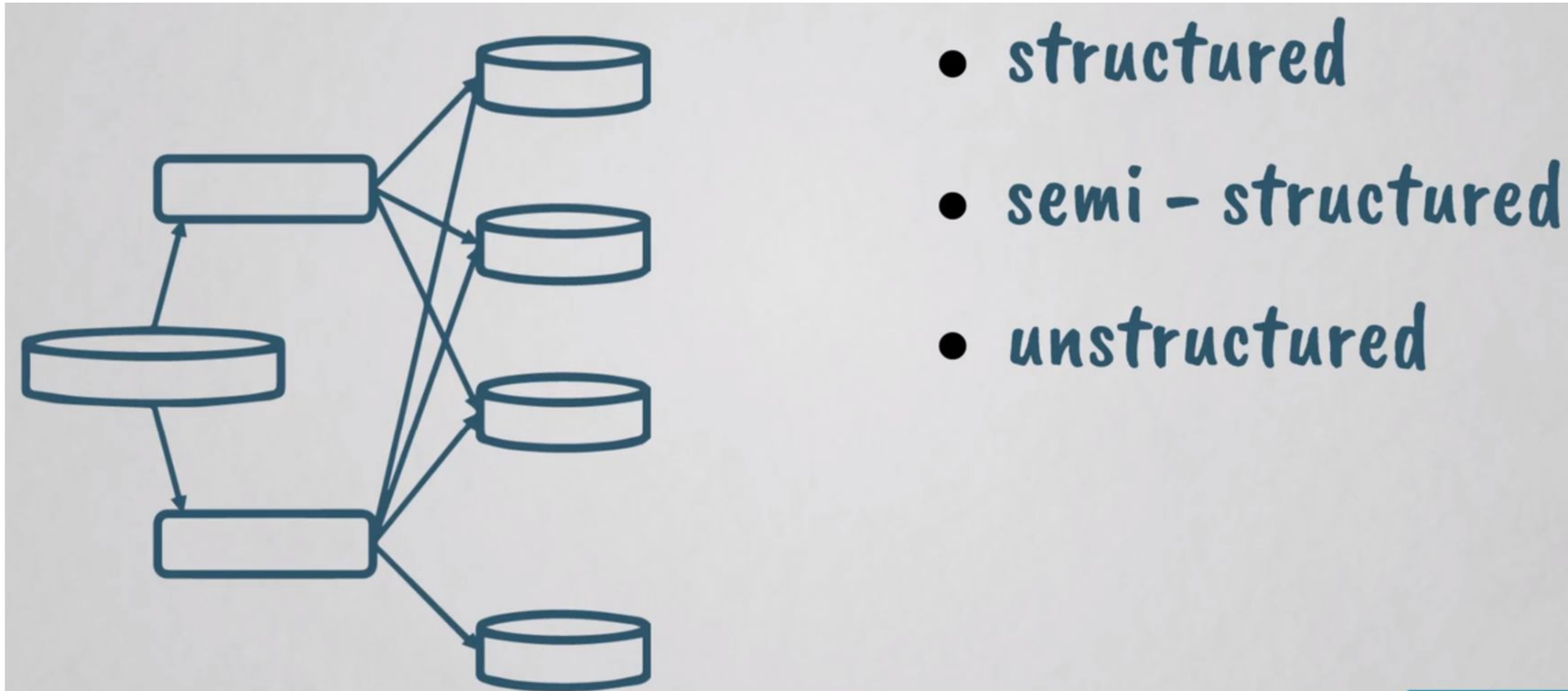
Traditional Data

- is **structured** and stored in **databases**
- in the form of tables containing **numeric** or **text** values
- can be managed from one computer.



Big Data - 1

- a term reserved for extremely **large** data
- not just humongous in terms of **volume**
- could be in various format:
  - **structured**
  - semi-structured
  - unstructured



Big Data - 2

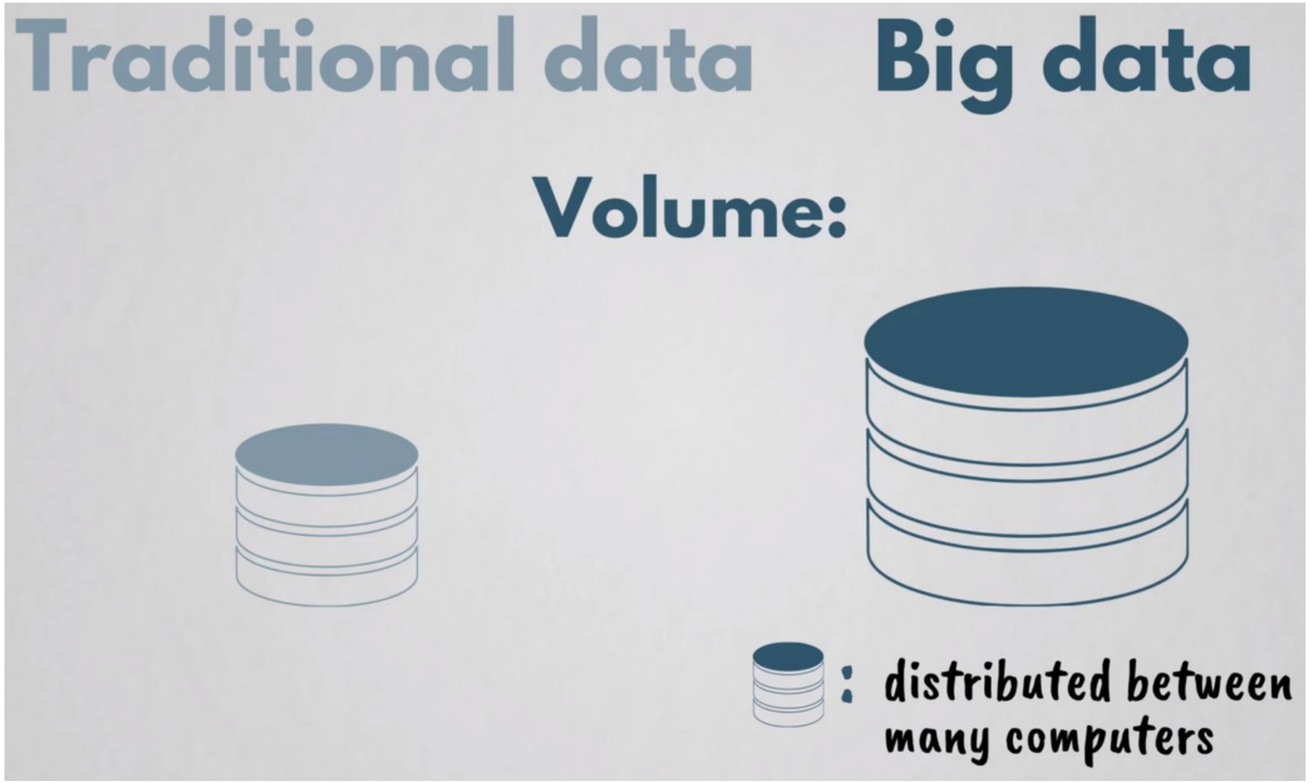
- is often characterized with the letter ‘**V**’
  - Under different frameworks we may have **3**,**5**,**7** and even **11 Vs** of **Big Data**

- The main **Vs**:
  - **volume**: amount of data
  - **variety**: number of data types
  - **velocity**: speed of data
- Some other **Vs**:
  - **vision** about **Big Data**
  - **value** big data carries
  - **visualization** tools used
  - **variability** in the consistency of BI data

Traditional Data vs Big Data - Volume

Big Data

- needs a whopping amount of memory space,
- typically distributed between many computers
- Its size is measured in TB, PB, EB



Traditional Data vs Big Data - Variety

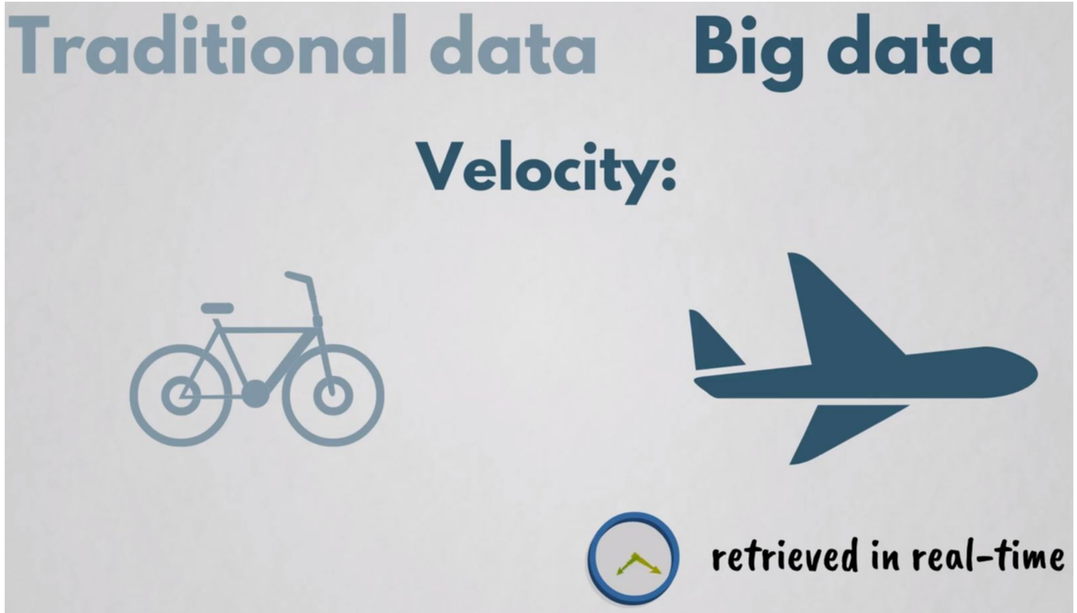
Big Data

- not just numbers and text
- implies dealing with images, audio, video, files, mobile data, and others

Traditional Data vs Big Data - Velocity

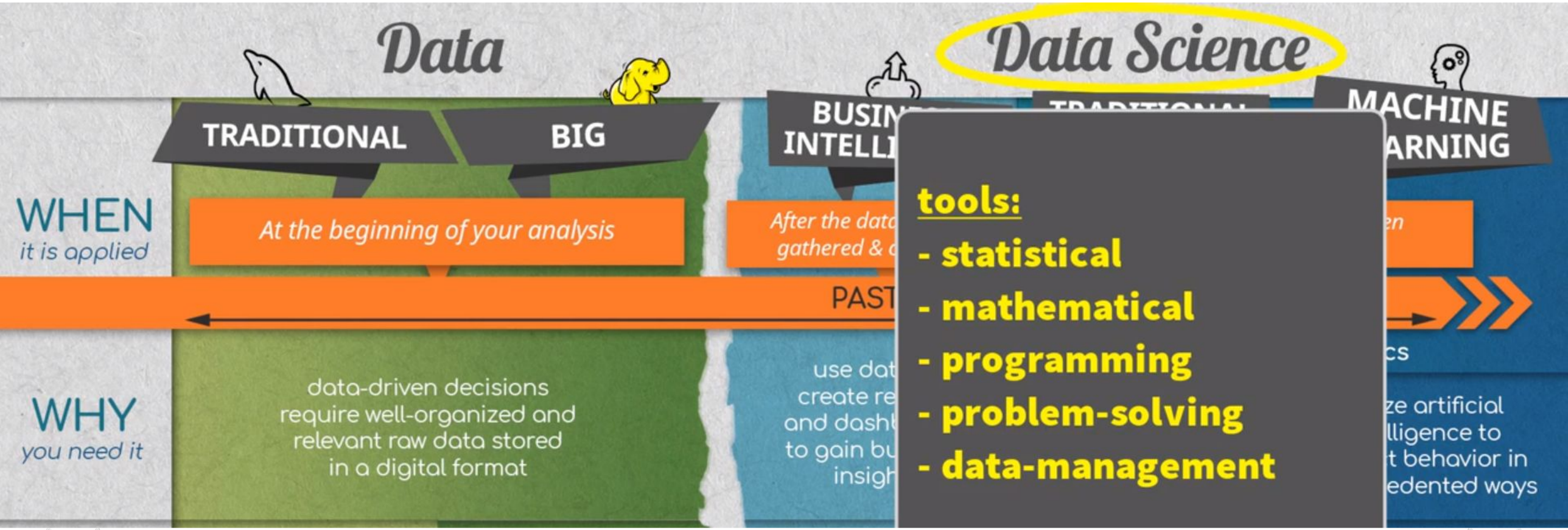
Big Data

- One goal is to make extracting patterns from **Big Data** as quickly as possible
  - the progress that has been made in this area is remarkable
- Outputs from huge datasets can be retrieved in real-time
  - this means they can be extracted so quickly,
  - so results could be computed immediately after source data has been obtained



Data Science

Data Science - 1



- After gathering and organizing all data,
  - it is time to get your hand dirty with **analytics**

Data Science - 2

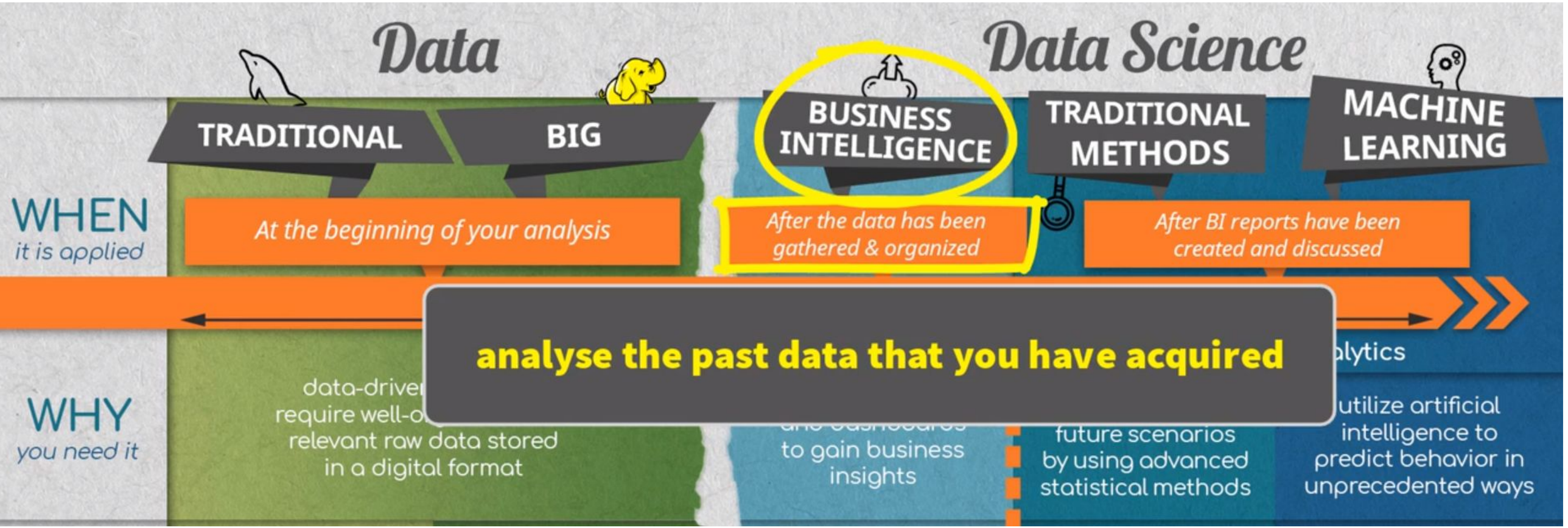
The infographic divides Data Science into three segments:

- BI
- Traditional methods
- ML methods

The infographic divides Data into two segments:

- Traditional Data
- Big Data

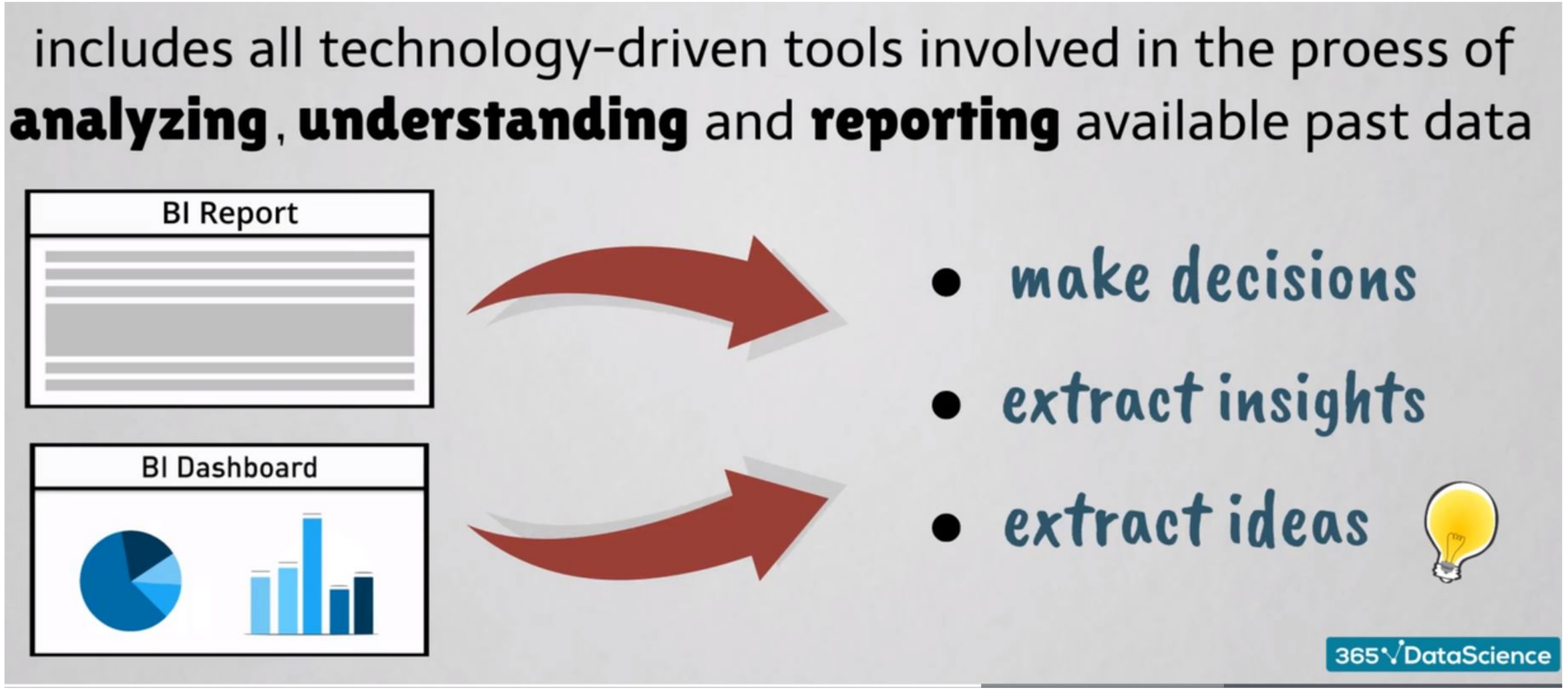
Business Intelligence - 1



1st step of applying data science

- is to analyse the past data that we have acquired
- BI is the discipline we need for this

Business Intelligence - 2



- This would result in having reports or dashboards,
  - which will help in making informed, strategic, and tactical business decisions

BI Questions - 1

BI Questions - 2

- **Example**: BI means understanding **how** your sales grew and **why**:
  - Did competitors lose market share?
  - Was there an increase in the price of your products?
  - Did you sell a mix of more expensive products?
  - `Were there more profitable client accounts?`
  - How did your profitability margins behave in the same time frame of the previous year?

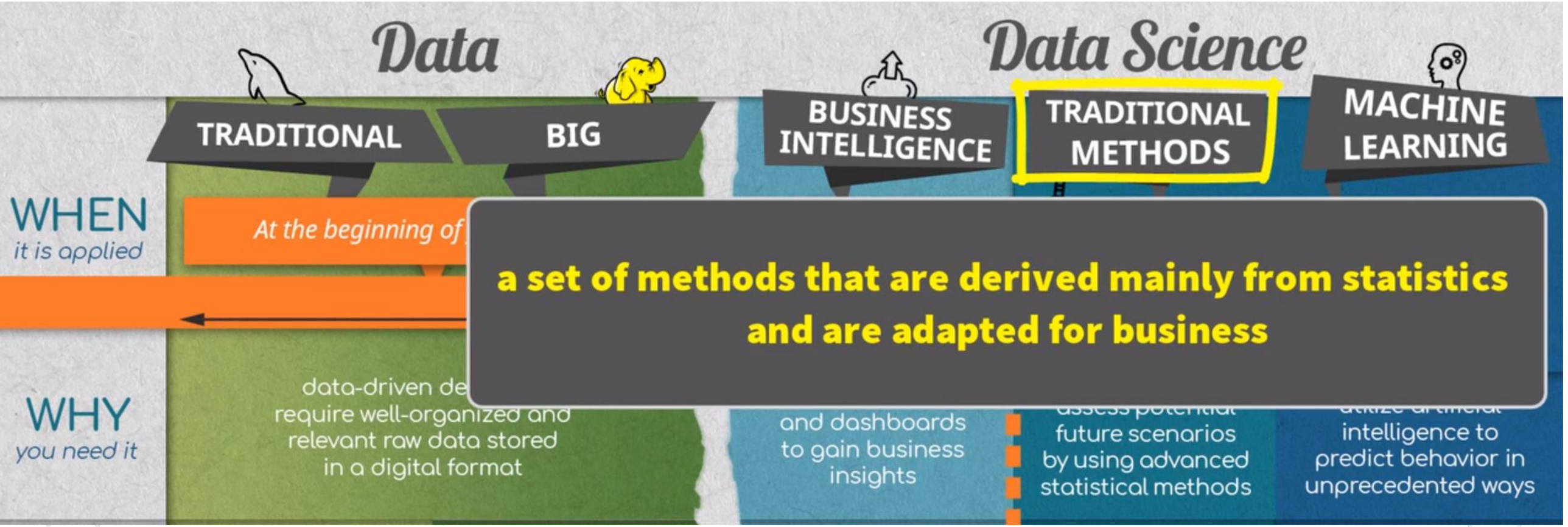
- BI is all about:
  - understanding past business performance to improve future performance

After BI

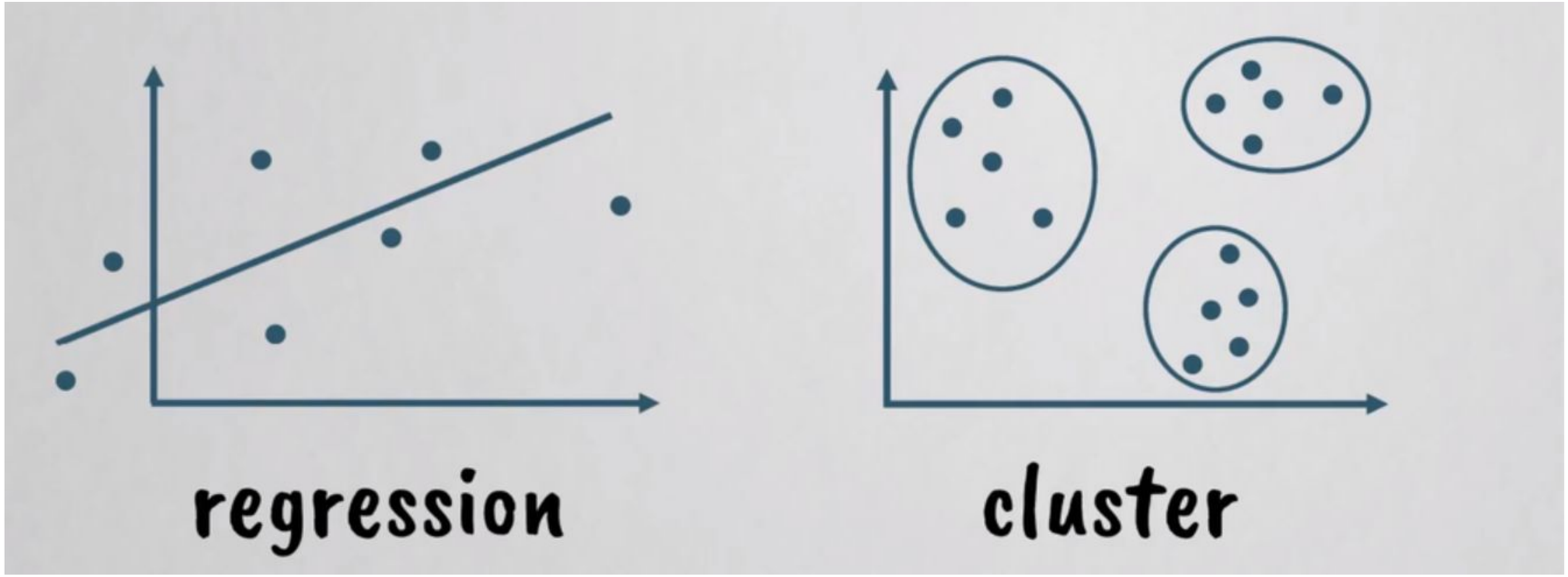
- BI is worth the time in the total process
- BI extracts insights and ideas about business that will
  - help to grow
  - give an edge over competitors, giving added stability

- We want to forecast future sales and profitability, as well as expenses
- Once **BI reports and dashboards** are complete and presented, it is time to apply:
  - Traditional Methods (Traditional Data Science)
  - or ML Methods
    - to develop an idea of what will happen

Traditional Methods

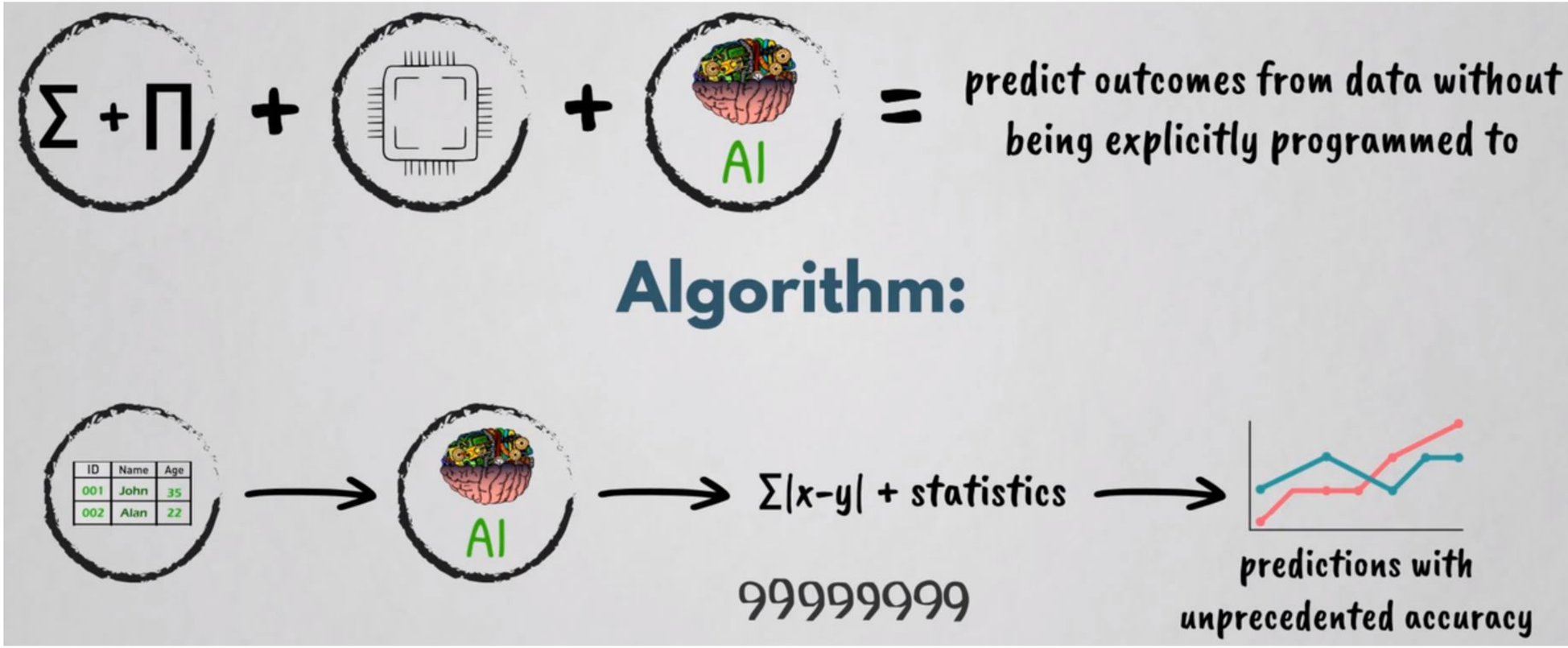


Traditional Methods



- perfect for forecasting future performance with great accuracy
- there is no denying that these tools are absolutely applicable today

Machine Learning



- Through
  - mathematics,
  - a significant amount of computer power, and
  - applying AI,
- the machine can predict outcomes from data
- without being explicitly programmed to

Machine Learning

- In **ML**, the responsibility is left to the machine

- **ML** is all about creating algorithms that let machines
  - receive data,
  - perform calculations, and
  - apply statistical analysis
  - to make predictions with unprecedented accuracy

Note

- The border between Traditional and ML methods can be considered thin (artificial)
- The mathematics behind both is virtually the same
- Nevertheless, we will use this thin boundary to explain better
  - which techniques are considered classical and
  - which are more complex and unconventional

Questions

Links

https://github.com/fcai-b/da

References

1. https://learn.365datascience.com/courses/intro-to-data-and-data-science
    - 365 Data Science - Introduction to Data and Data Science
2. https://www.coursera.org/learn/data-analysis-with-python
    - IBM Coursera Course - Data Analysis with Python