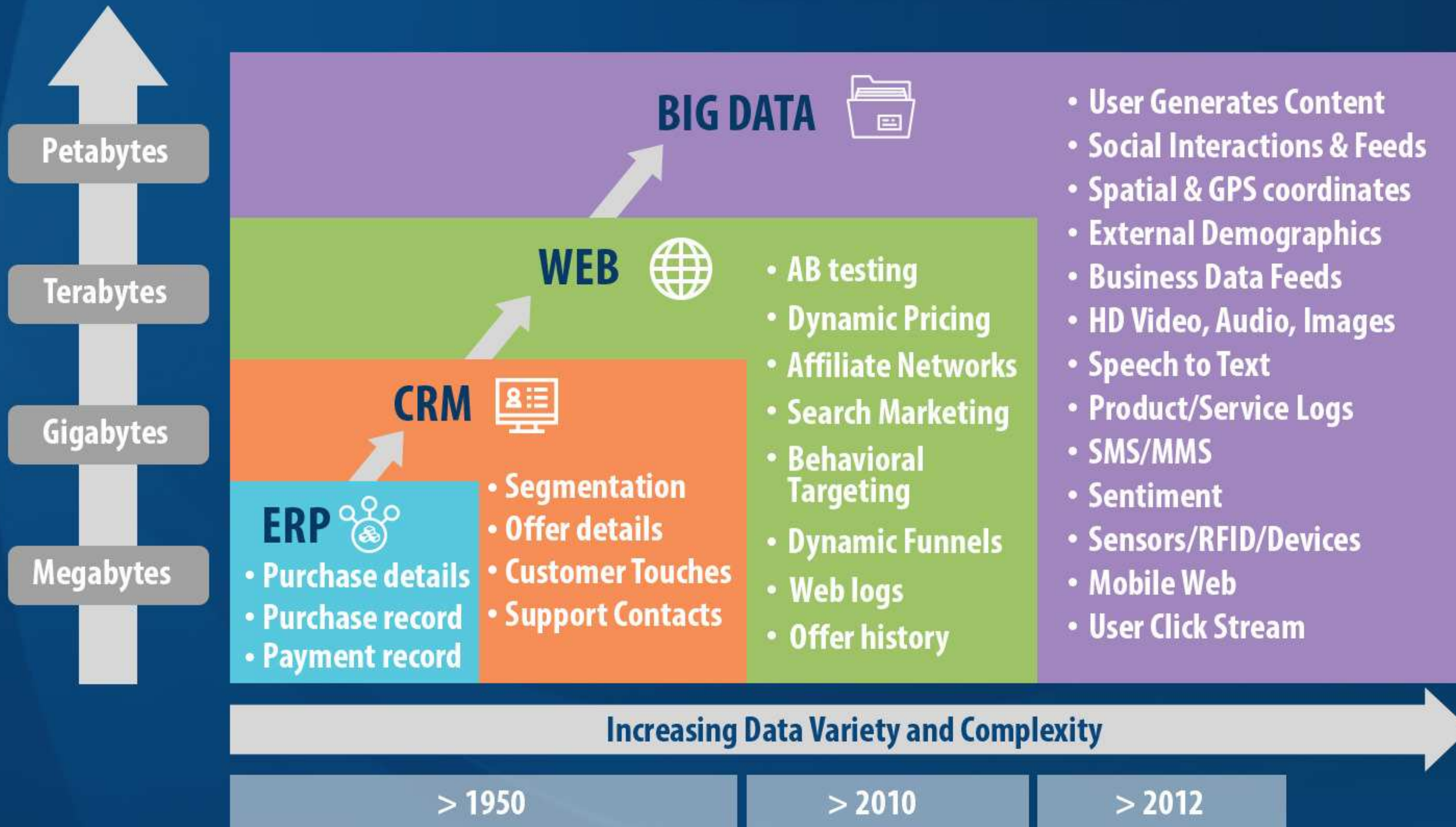


WHAT IS THE DATA?



Big Data

➤ Broad term for data sets so large or complex that traditional data processing applications are inadequate

➤ **Characteristics:**

- Volume
- Variety
- Velocity

WHY IS DATA SCIENCE IMPORTANT?

How can we tap this 'oil'?



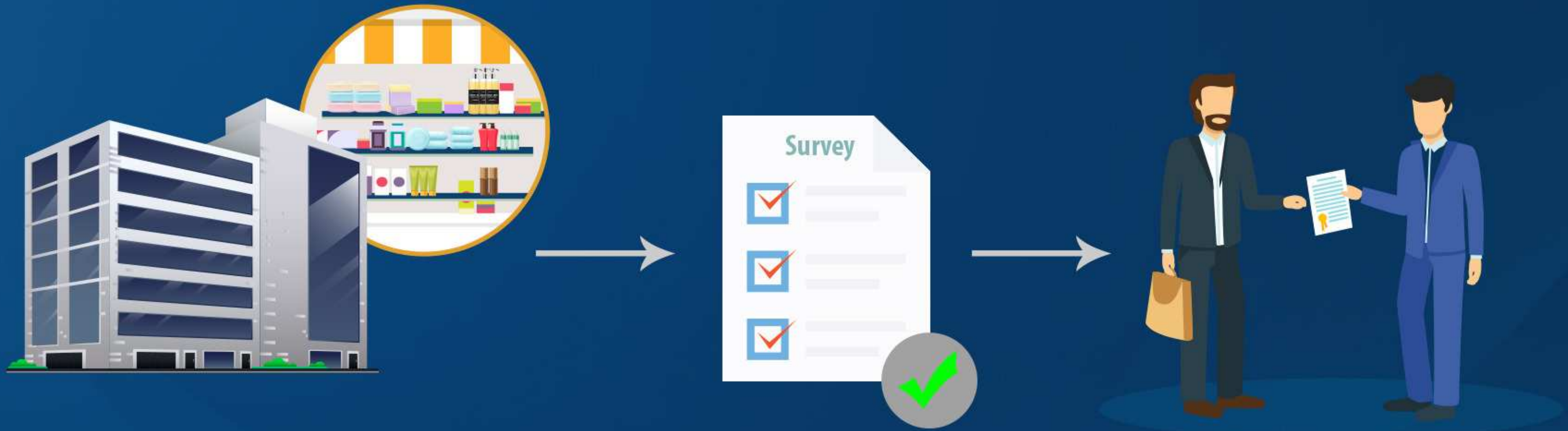
Data is the new oil
Clive Humby



AMITY
UNIVERSITY
ONLINE

CAREERS OF TOMORROW

WHY IS DATA SCIENCE IMPORTANT?



BUSINESS ANALYTICS vs DATA SCIENCE

Basis For Comparison	Data Science	Business Analytics
Coining of Team	DJ Patil and Jeff Hammerbacher who were working in LinkedIn and Facebook respectively, first coined the term Data Scientist in 2008.	Business Analytics has been used since the late 19th Century when it was put in place by Frederick Winslow Taylor.
Concept	Interdisciplinary field of data inference, algorithm building, and systems to gain insights from data.	Use of statistical concepts to extract insights from business data.
Coding	Coding is used widely. The field is a combination of traditional analytics practice with sound knowledge of computer science.	Does not involve much coding. More statistics oriented.
Languages Recommendations	C/C++/C#, Haskell, Java, Julia, Matlab, Python, R, SAS, Scala, SQL, Stata	C/C++/C#, Java, Matlab, Python, R, SAS, Scala, SQL
Statistics	Statistics is used at the end of the analysis following algorithm building and coding.	The whole analysis is based on statistical concepts.
Data Needed	Both structured and unstructured data.	Predominantly structured data.
Future Trends	Machine Learning and Artificial Intelligence.	Cognitive Analytics, Tax Analytics.



WHAT IS DATA SCIENCE?

Data



- (Noun) **Facts** and **statistics** collected together for reference or analysis.
- (Origin) (As a term in philosophy): from Latin, plural of datum. Literally 'something given', neuter past participle of dare 'give'.

+

Science



- (Noun) The intellectual and practical activity encompassing the systematic study of the structure and behaviour of the physical and natural world through **observation** and **experiment**.
- (Origin) Middle English (denoting knowledge): from Old French, from Latin *scientia*, from *scire* 'know'.

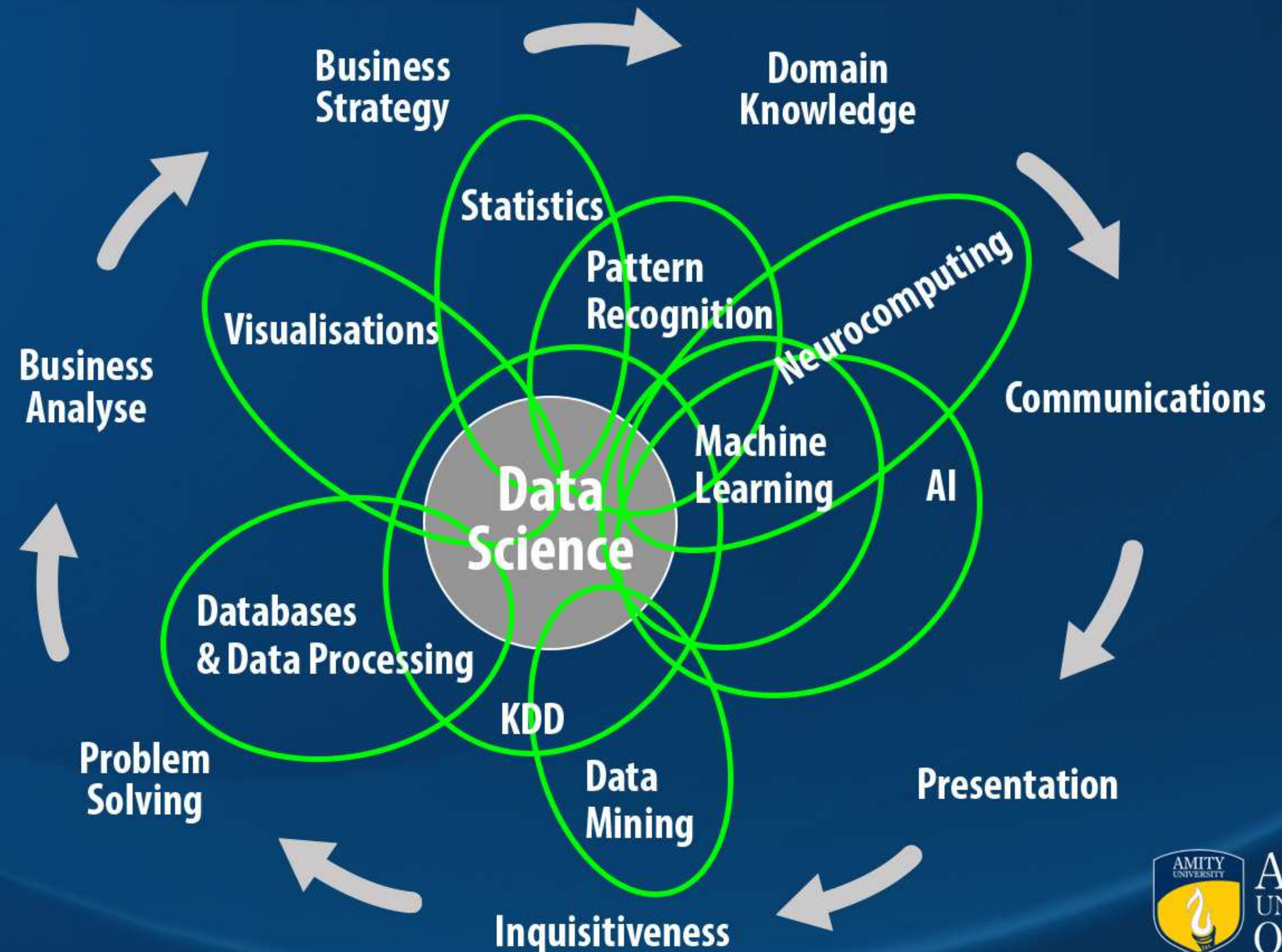


The keyword in “Data Science” is not Data, it is Science!

DATA SCIENCE IS MULTIDISCIPLINARY

What does a data scientist have to do?

Ans: Wear Multidisciplinary hats!



WHAT IS THE SCIENCE?

Computer Science



All studies that form the basis of the design and use of computers

Artificial Intelligence (AI)



Building computers and systems that can do intelligent things that humans naturally do on a daily basis

Machine Learning (ML)



- ✓ Dynamic
- ✓ Data Driven
- ✓ Generalized

Creating machines that are intelligent by:

- Learning the patterns present in data
- Predictions based on the learning

Deep Learning (DL)



- ✓ Complex tasks
- ✓ Self Learning
- ✓ Larger data volumes

1940

> 1950

> 1980

> 2015



AMITY
UNIVERSITY
ONLINE

CAREERS OF TOMORROW

WHAT SKILLSETS MAKE A GOOD DATA SCIENTIST?

Math

- Understanding and deftness into the mathematical and statistical techniques needed to solve real-business problems with data

Tech Skills

- Ability to work on native & new tools to execute on analytical projects
- Need to perform complex data mining and manipulation exercises

Domain Understanding

- Implementing analytical techniques in context of a problem, industry and environment

Problem Solving

- Ability to deploy the right solution for the preferred outcome using first principle thinking, logic, imagination, intuition, and systemic reasoning

THE DATA SCIENCE LIFE CYCLE

