BIG DATA MANAGEMENT

CZ/CE4123

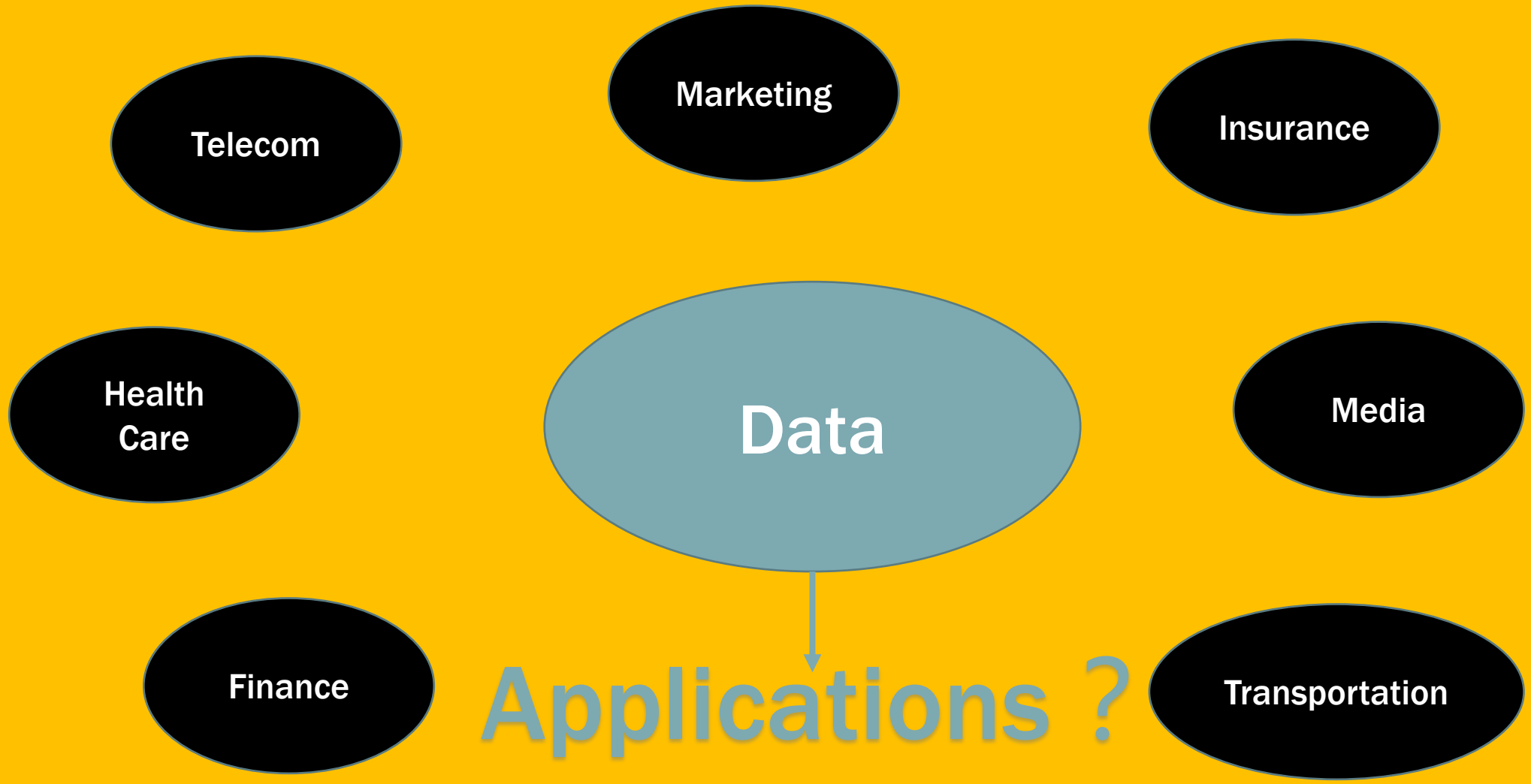
Why big data?

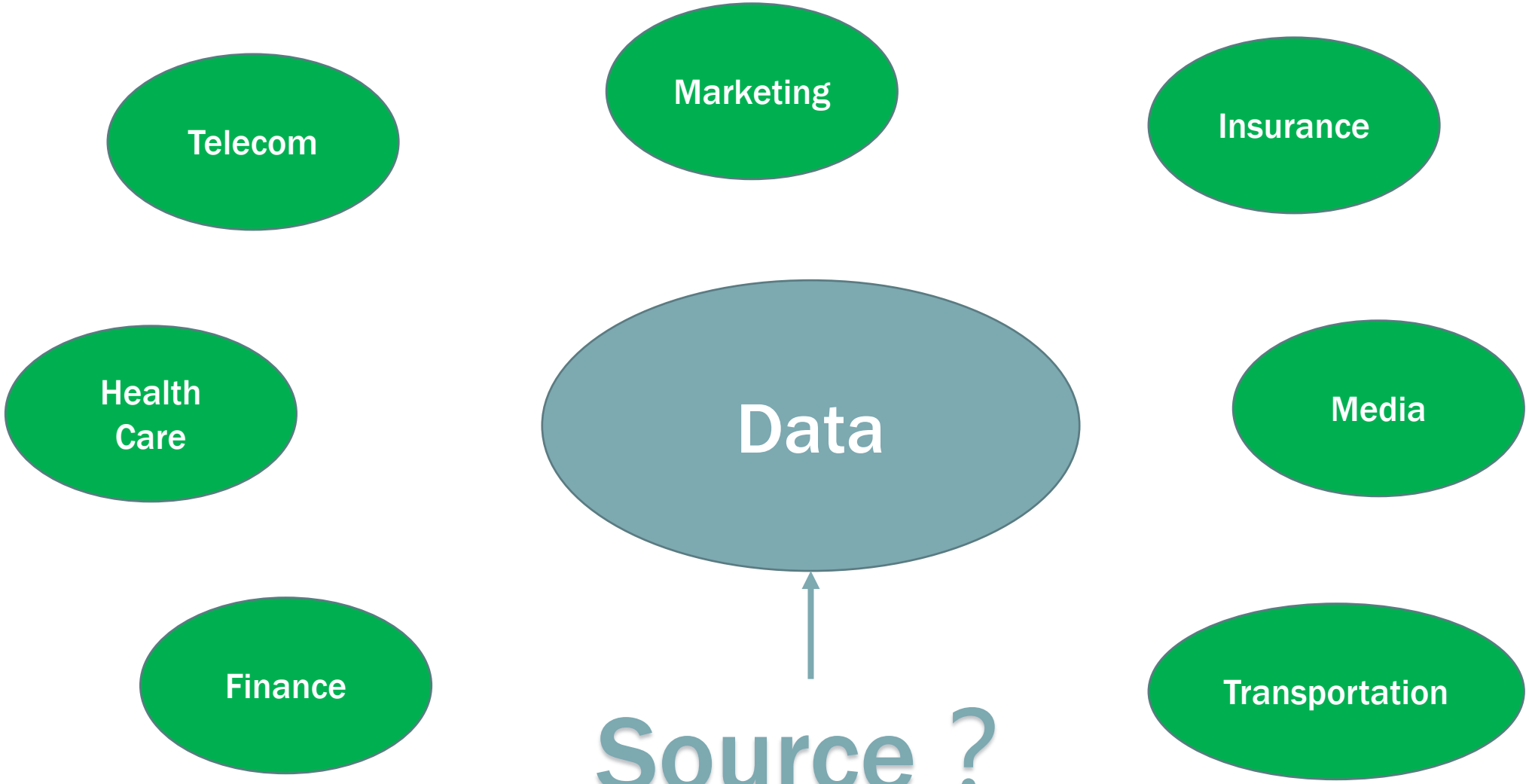
WHY WE NEED TO UNDERSTAND BIG DATA MANAGEMENT

Data

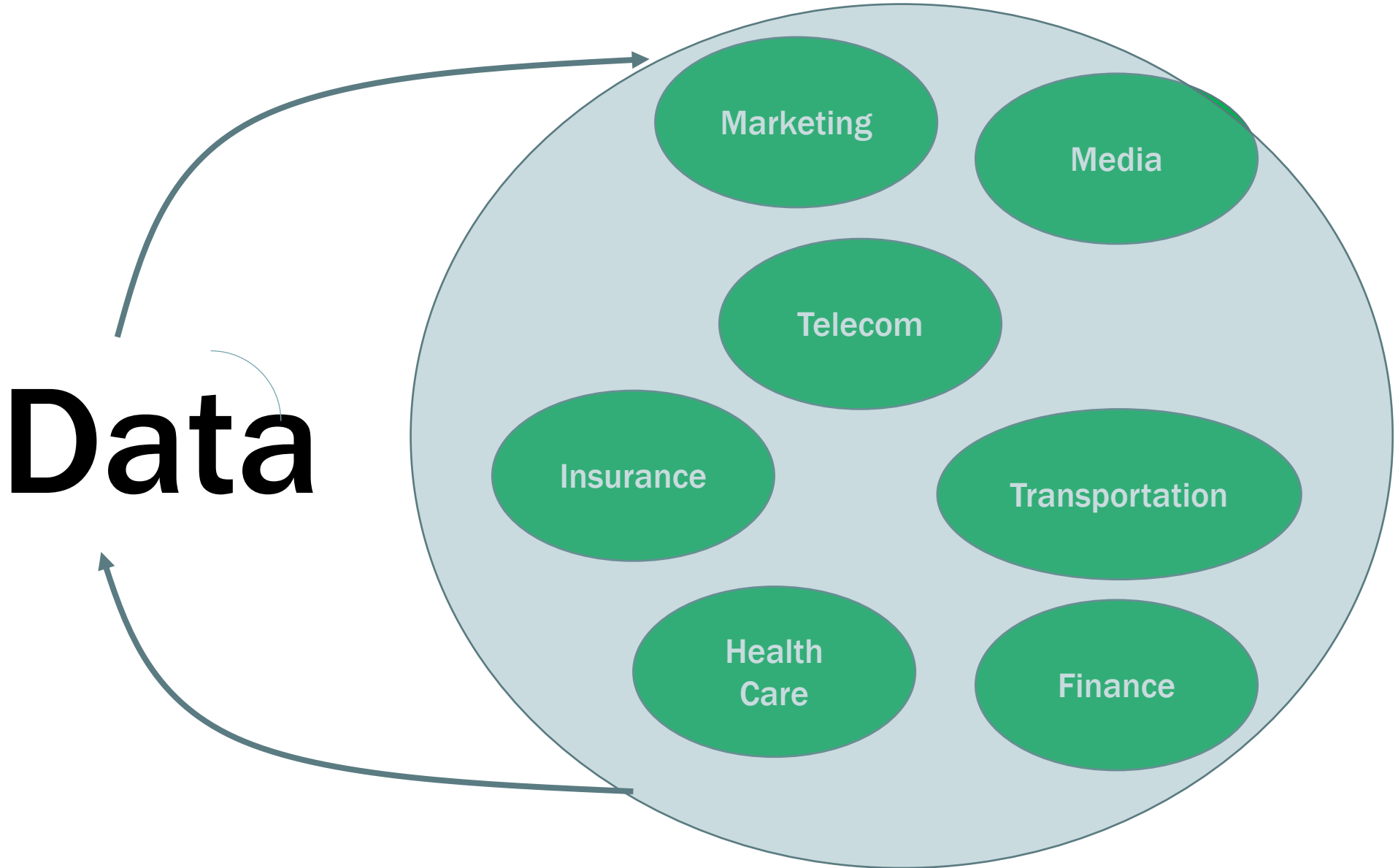
Data

Is at the center of most things


Telecom
Marketing
Insurance
Health Care
Data
Media
Finance
Applications ?
Transportation


Telecom
Marketing
Insurance
Health Care
Data
Media
Finance
Source ?
Transportation

DATA LOOP


Data
Marketing
Media
Telecom
Insurance
Transportation
Health Care
Finance

APPLICATIONS WE USE DAILY

DATA SIZE

1 Bit – 1/8 of a letter

1 Byte – a letter (8 bits)

1 Megabyte – a book (1024 kilobytes)

1 Gigabyte -- ~1000 books (1024 megabytes)

1 Terabyte -- ~ 1million books (1024 gigabytes)

1 Petabyte -- ~ 1billion books (1024 terabytes)

1 Exabyte -- ~ 10^12 books (1024 Petabytes)

Source: Columbia University

BOOKS PILING UP TO REACH MOON

1 Exabyte

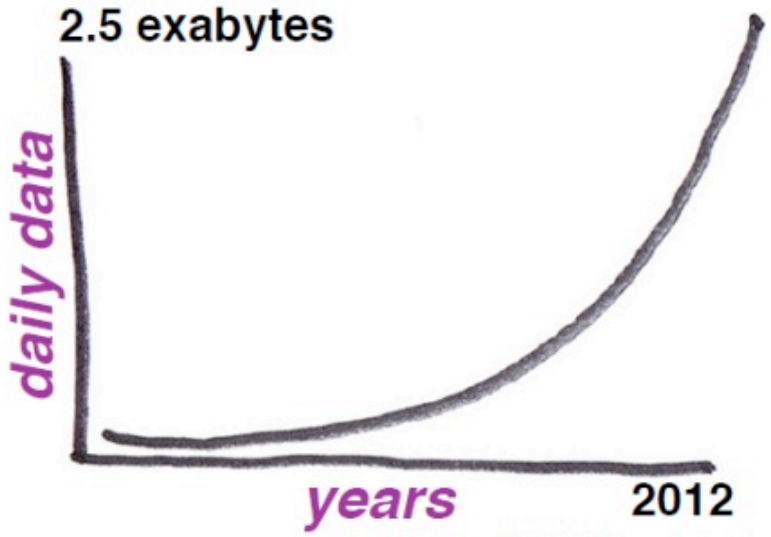
Moon

distance

Earth

HOW MUCH DATA GENERATED?

Nowadays, more than 2.5 exabytes of data are generated in every single date!



[IBMbigdata]

Every two days we create as much data as much we did from the dawn of humanity to 2003.

[Eric Schmidt, Google]

We are in the era of big data!

BIG DATA ECONOMY

>$3 trillion/per year

A report from McKinsey Global Institute estimates that Big Data could generate an additional $3 trillion in value every year in just seven industries. The report also estimated that over half of this value would go to customers in forms such as fewer traffic jams, easier price comparisons, and better matching between educational institutions and students.

>500,000 big data jobs

Data analysis has been called “the sexiest job of the 21st century.” The United States already has an estimated 500,000 Big Data jobs. But McKinsey estimates that there is a shortage of between 140,000 and 190,000 workers with advanced degrees in statistics, computer engineering and other applied fields. Perhaps more important is the shortage of 1.5 million managers and analysts who hold traditional jobs but are capable of integrating Big Data into their decision making.

THE KEY CHALLENGE OF BIG DATA

Data is so big that we need

High-Performance Computation over big data

efficiently store and query big data

CHALLENGES OF BIG DATA

How to store ?

https://www.freeiconspng.com/images/question-mark-icon

CHALLENGES OF BIG DATA

How to query ?

THE FIRST THING WE NEED TO KNOW

What are the characteristics of big data?

Next lecture: