

Big Data Introduction

What is Data?

- The quantities, characters, or symbols on which operations are performed by a computer,
- which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.



What is Big Data?

- Big Data is also data but with a huge size.
- Big Data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time.
- In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.



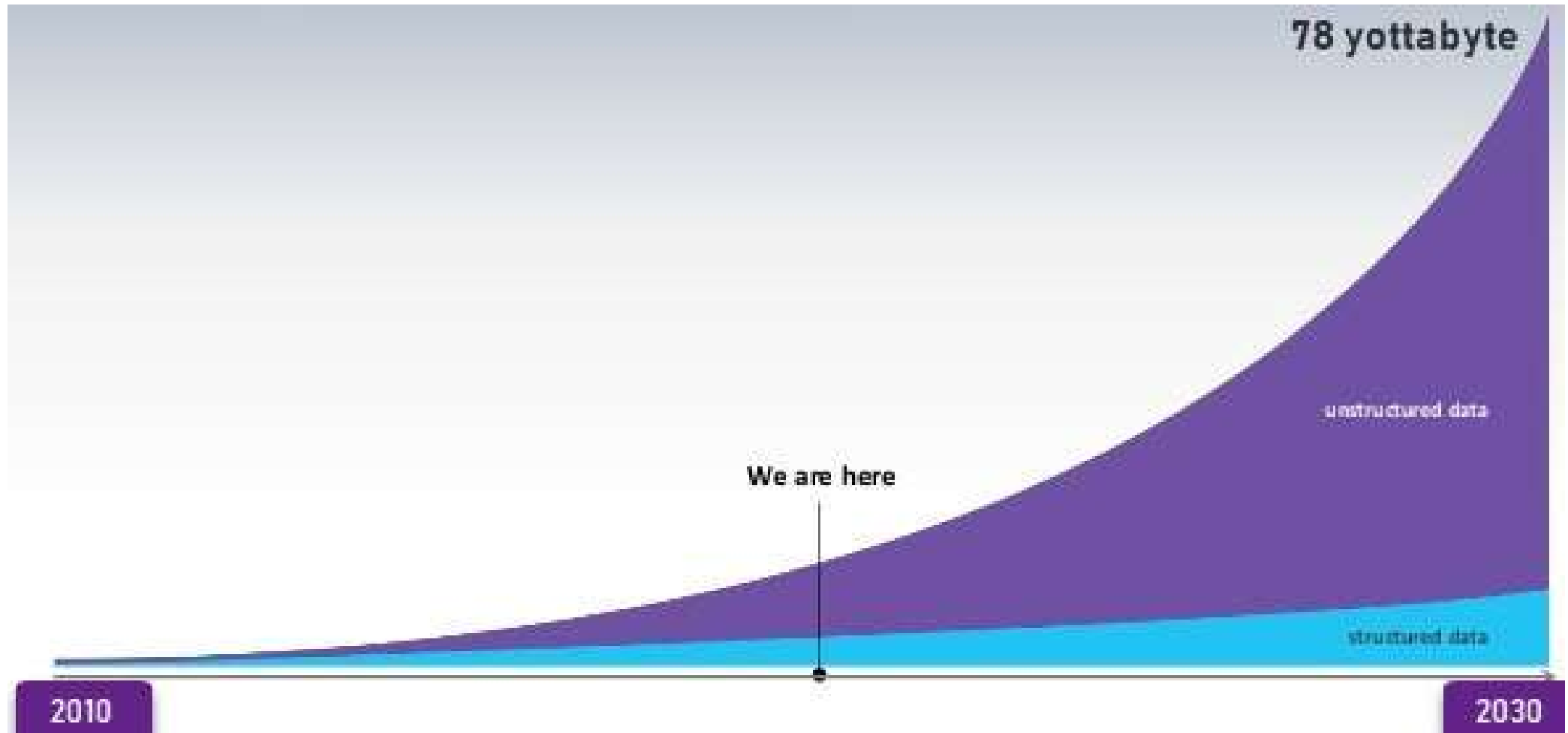
Examples Of Big Data

- The New York Stock Exchange
 - Generates about one terabyte of new trade data per day.
- Social Media
 - The statistic shows that 500+terabytes of new data get ingested into the databases of social media site Facebook, every day.
 - This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.
- A single Jet engine
 - Can generate 10+terabytes of data in 30 minutes of flight time. With many thousand flights per day, generation of data reaches up to many Petabytes.

Types Of Big Data

- BigData could be found in three forms:
 - Structured
 - Any data that can be stored, accessed and processed in the form of fixed format
 - Unstructured
 - A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc.
 - Semi-structured
 - Can contain both the forms of data.

Data Growth over the years



Once upon a time



Total Capacity – 1.44 MB

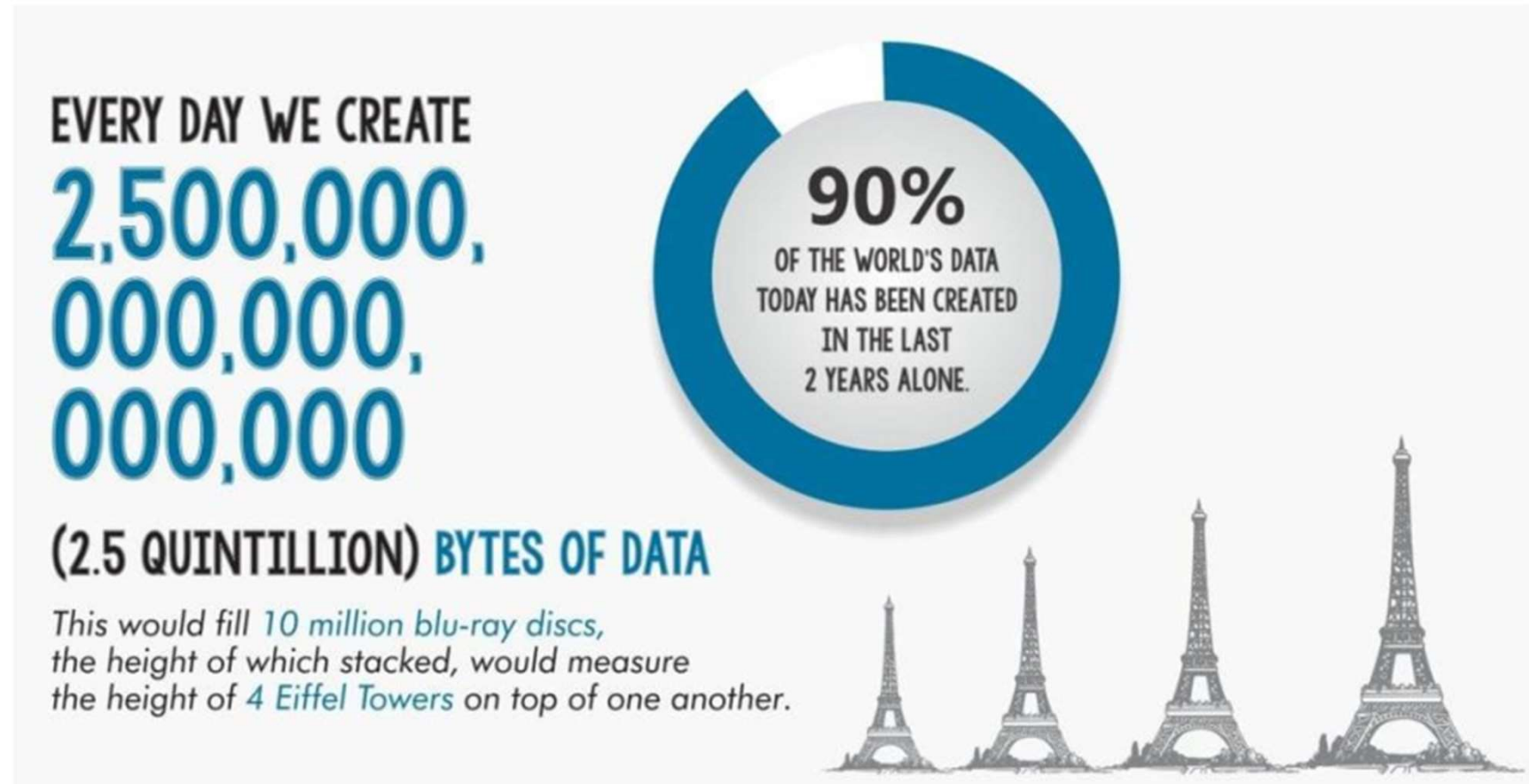
1990

By 2020, it's estimated that 1.7MB of data will be created every second for every person on earth."

— Domo report (6th edition)

2020

How much data we create daily?



1 Quintillion = 1 million terabyte

Data is now new oil

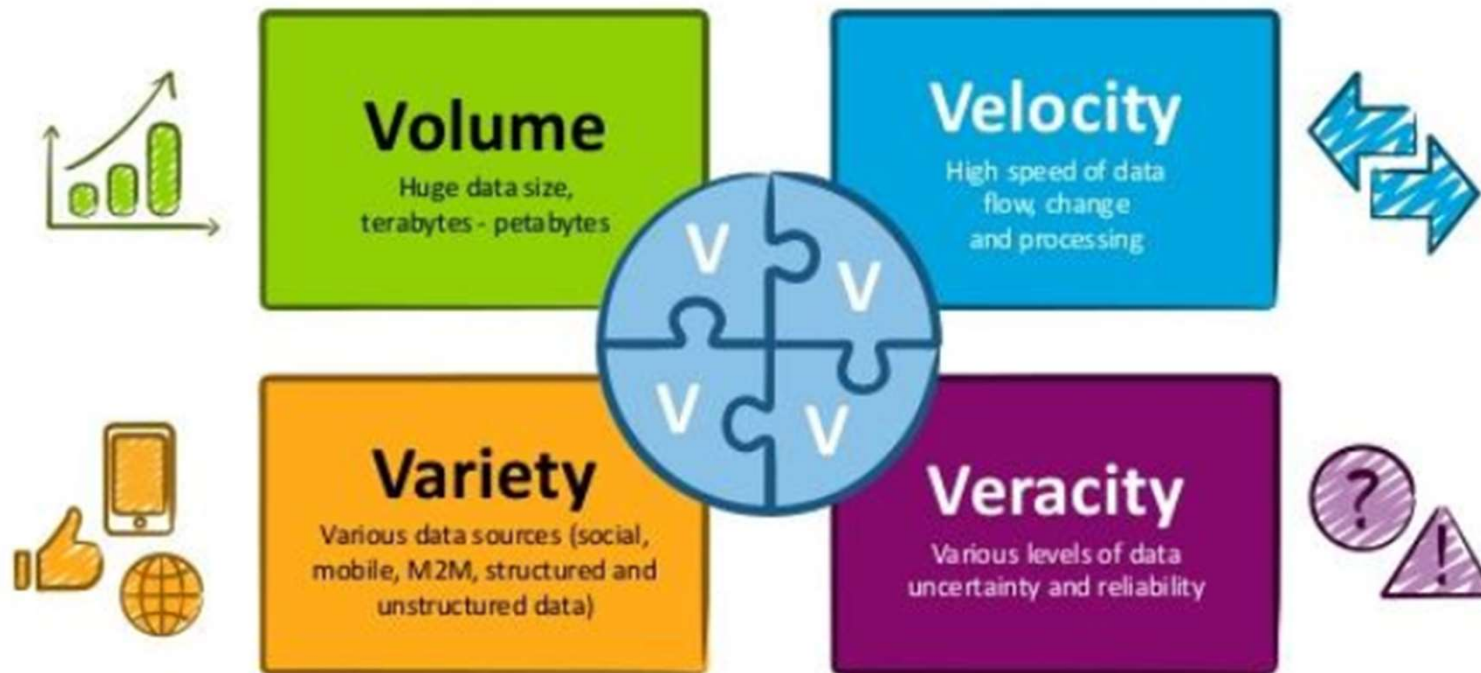
We need to find it, extract it, refine it, distribute it and monetize it

– David Buckingham, Big data expert



Problem statement

- Need a solution which can handle below 4 V's of data.





Thanks