University of South Australia

What is Big Data?

1

What is data?

Raw facts, figures or knowledge

Meaningless without context!

How does ‘regular data’ differ from ‘Big Data’?

2

What is Big Data?

Data sets whose **volume**, **velocity** or **variety** is so large and *complex* that it is difficult to store, manage, process and analyse using traditional database systems and data processing tools.

3

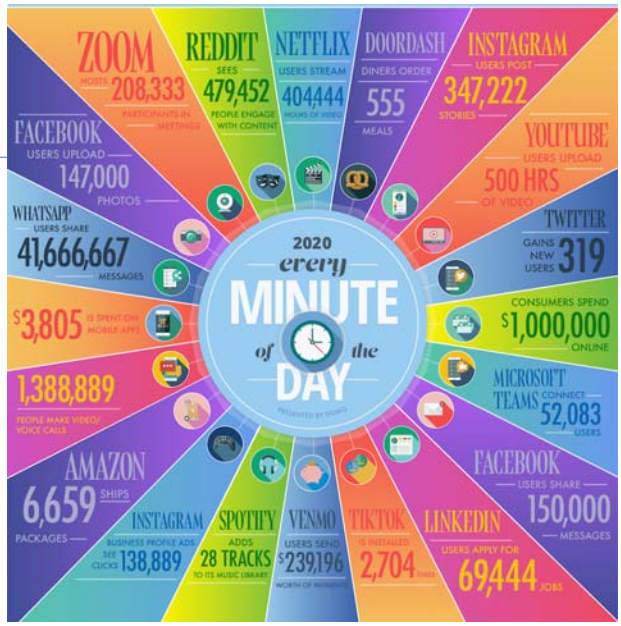
Big Data sources

- Business transaction systems
- Customer databases
- Medical records
- Internet clickstream logs
- Mobile applications
- Social networks
- Scientific research repositories
- Machine sensors and devices (IoT)

4

Data never sleeps 8.0

- Technology and data play significant roles in our daily lives.



https://www.domo.com/learn/data-never-sleeps-8

5

Big Data characteristics – 3Vs

- Doug Laney, analyst at Meta Group Inc. (2001), later popularised by *Gartner*:
- Large **volume** of data in many environments;
- Wide **variety** of data types stored in Big Data systems;
- **Velocity** at which the data is generated, collected and processed.

6

Volume

- No specific size to make data ‘big’
- Big Data deployments can involve terabytes (TB), petabytes (PB) and even exabytes (EB) of data captured over time.

7

Variety

Structured data in SQL-based databases and warehouses

Unstructured data in Hadoop clusters or NoSQL database systems

Semi-structured data, e.g. web server logs

Can be stored together in a data lake, which typically is based on Hadoop or a cloud object storage service

Often include multiple data sources that may not otherwise be integrated

8

Velocity

- Speed at which big data is generated and must be processed and analyzed
- In many cases, sets of big data are updated on a real- or near-real-time basis

9

The six Vs of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume*, *variety and velocity*. Over time, other Vs have been added to descriptions of big data:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.

10

It's Vs all the way down

https://www.m-brain.com/technology/

https://images.xenonstack.com/blog/10-vs-of-big-data.png

The Internet of Things (IoT)

- Source of 'Big' and 'small' data
- Smart devices that can talk to each other and measure, monitor and control the physical world

Medium.com