

Big data

What is it?

- ♥ First of ALL, what is data?
- ♥ Data comes from the Latin verb *dare* which means '*to give*'.
- ♥ Data are **raw facts** given. Facts connote truth. Attempting to extract information from false data leads logically to false information.
- ♥ Singular – datum.
 - However, **data is more regularly used as a collective noun**. i.e. data is more regularly used in the singular to denote a collection or set of items.
 - E.g. not enough **data is** available to arrive at a valid conclusion.

Big data

*...What is
it?*

- 🏰 **Data vs Information vs Knowledge vs Wisdom?**
- 🏰 **Information** == processed data
- 🏰 Information transports **knowledge**
 - e.g. in communication processes
 - Knowledge representation in information in both learning and sharing processes
- 🏰 **Wisdom** == ability to make sound judgement
 - beyond simply knowing
 - but thanks to wealth of knowledge

Big data

...What is it?

Typically defined in the context of digital data (as opposed to analogue).

Refers to a **phenomenon**:

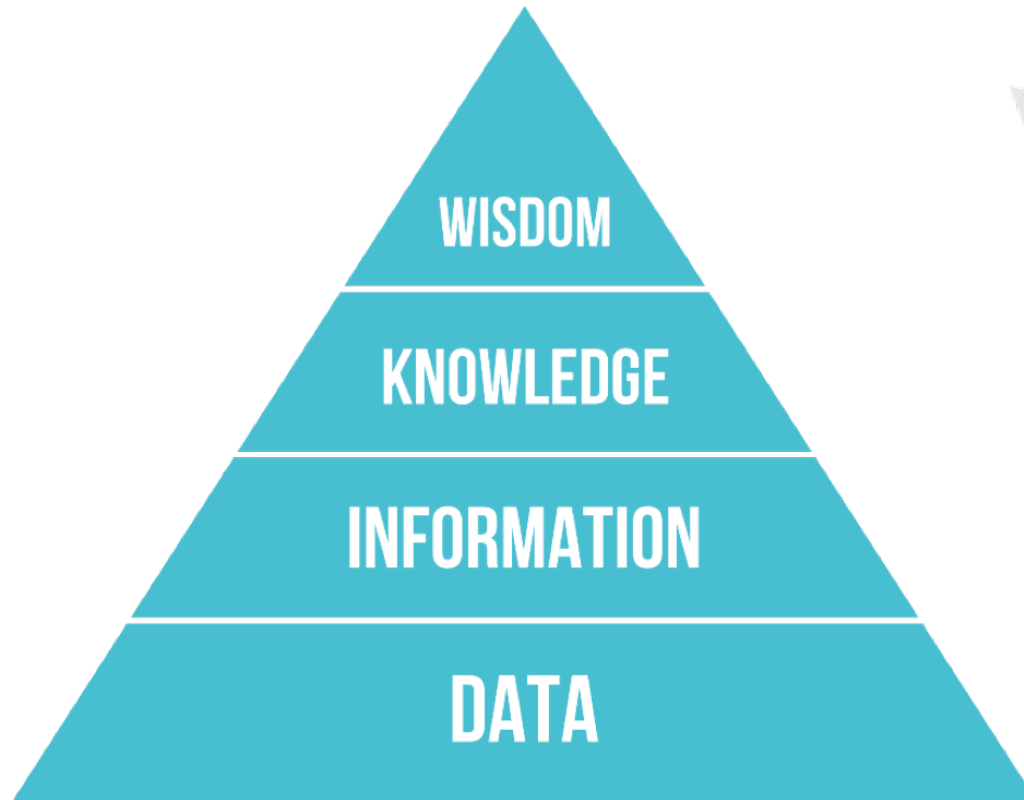
- Big data refers to the **reality** of available data that is not only **massive** in quantity but also that grows exponentially with user and device operations.
- i.e. characterized by:
 - **High volume**—the amount or quantity of data
 - **High velocity**—the rate at which data is created
 - **High variety**—the different types of data
- Additional characteristics (else not useful):
 - **Value** — could be hidden. Science usually required to extract.
 - **Veracity** — accuracy. False data is fake data.

Big data

***...What is
it?***



...Data vs Information vs Knowledge vs Wisdom?



**“In God we trust. All others must bring data.”
(William Edwards Deming)**

Big data

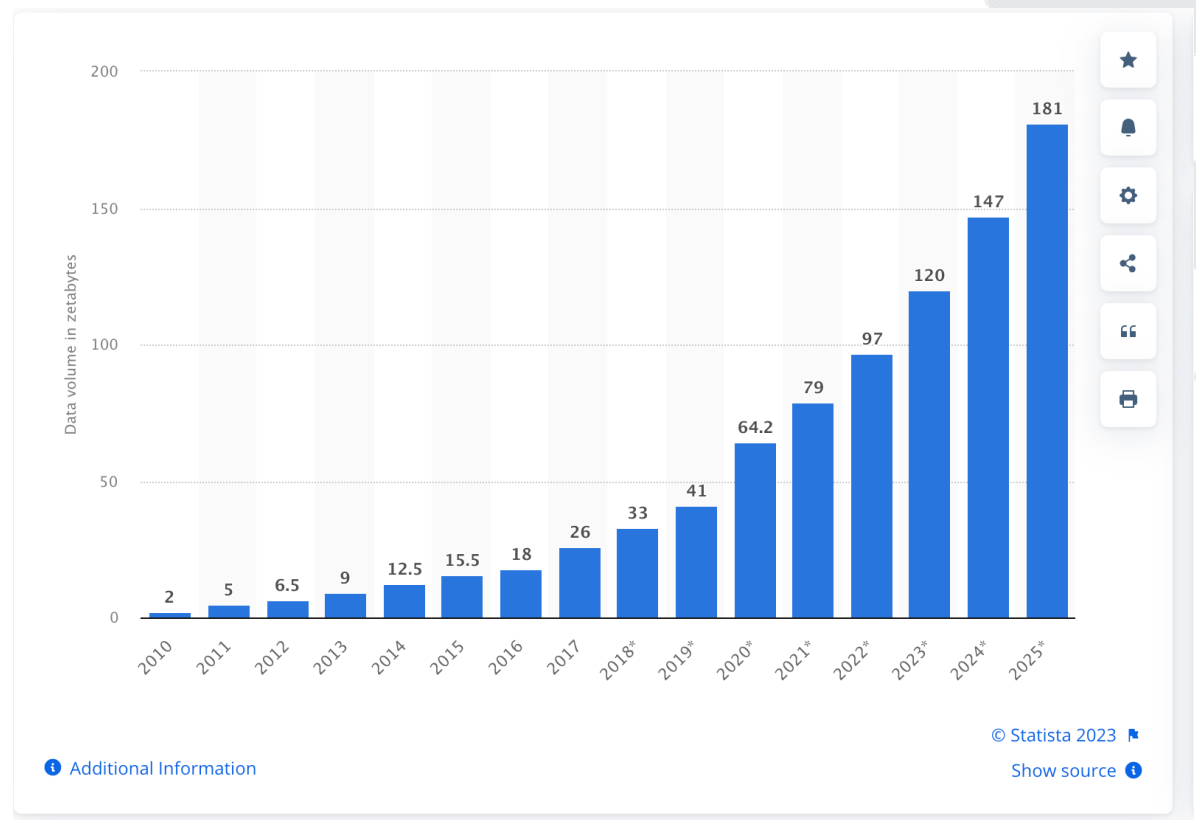
...What is it?



Some statistics



“Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025”



Ref: <https://www.statista.com/statistics/871513/worldwide-data-created/>

Big data

...What is it?

- 🏰 ...Some statistics
- 🏰 “The total amount of data created, captured, copied, and consumed globally is forecast to increase rapidly, reaching **64.2 zettabytes** in 2020.”
- 🏰 “Over the next five years up to 2025, global data creation is projected to grow to more than **180 zettabytes.**”

Ref: <https://www.statista.com/statistics/871513/worldwide-data-created/>

Big data

...What is it?

- 🏰 ...Some statistics
- 🏰 “In 2020, the amount of data created and replicated reached a new high. The growth was higher than previously expected caused by the increased demand due to the COVID-19 pandemic, as more people worked and learned from home and used home entertainment options more often. ”

Ref: <https://www.statista.com/statistics/871513/worldwide-data-created/>

Big data

...What is it?



Units

Unit	Value	Example
Kilobytes (KB)	1,000 bytes	a paragraph of a text document
Megabytes (MB)	1,000 Kilobytes	a small novel
Gigabytes (GB)	1,000 Megabytes	Beethoven's 5th Symphony
Terabytes (TB)	1,000 Gigabytes	all the X-rays in a large hospital
Petabytes (PB)	1,000 Terabytes	half the contents of all US academic research libraries
Exabytes (EB)	1,000 Petabytes	about one fifth of the words people have ever spoken
Zettabytes (ZB)	1,000 Exabytes	as much information as there are grains of sand on all the world's beaches
Yottabytes (YB)	1,000 Zettabytes	as much information as there are atoms in 7,000 human bodies

Ref: <https://www.eecis.udel.edu/~amer/Table-Kilo-Mega-Giga---YottaBytes.html>

Big data

...What is it?

🏰 Shapes

- Structured
- Semi-structured
- Unstructured

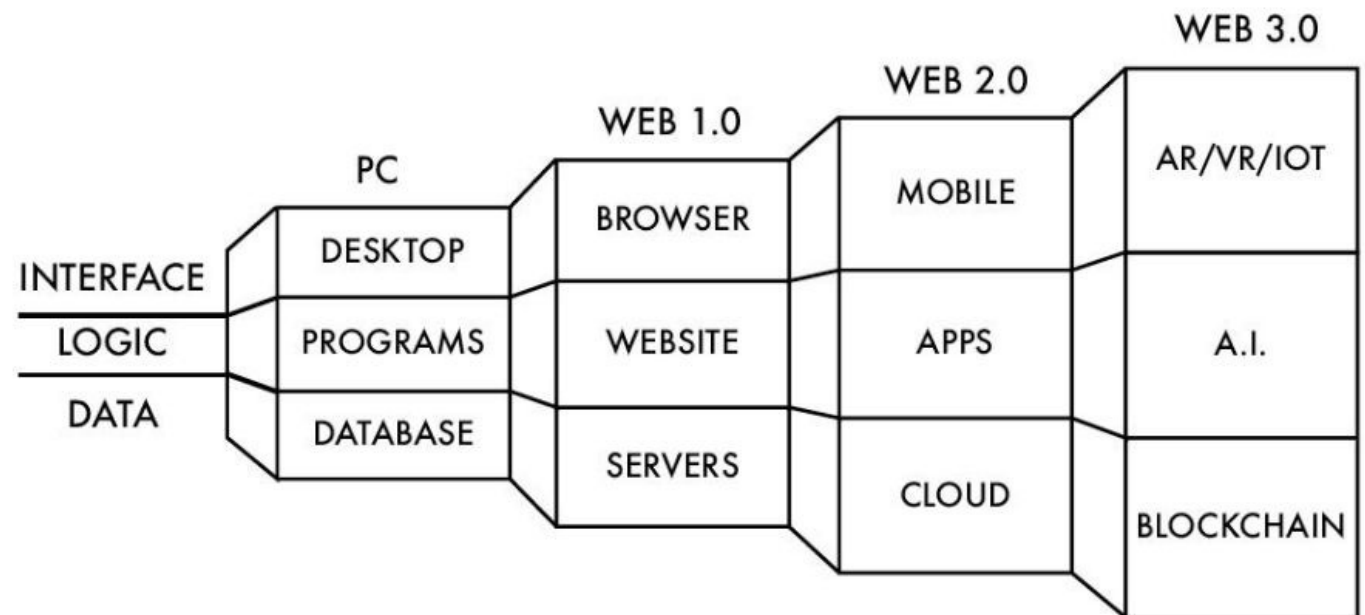
🏰 Sources

- Web
 - Regular
 - Social media
- Mobile
- IoT
- Etc.
- In summary,
 - Heavily facilitated by technologies that are layered on the Internet.

Big data

...*What is it?*

Web 1 → Web 2 → Web 3



Source: Web 3.0 will be an Autonomous Spatial Web. (2021, September 9). InWeb3. <https://www.inweb3.com/web-3-0-will-be-an-autonomous-spatial-web/>

Big data

...*What is it?*

- 🏛 In summary,
- 🏛 “Big data is a term that is used to describe data that is high volume, high velocity, and/or high variety; requires new technologies and techniques to capture, store, and analyze it; ...”

Ref: Watson, Hugh J. (2014) "Tutorial: Big Data Analytics: Concepts, Technologies, and Applications," Communications of the Association for Information Systems: Vol. 34 , Article 65.

Big data

*Why
should I
care?*

- ...In summary (extension),
- “Big data is a term that is used to describe data that is high volume, high velocity, and/or high variety; requires new technologies and techniques to capture, store, and analyze it; **and is used to enhance decision making, provide insight and discovery, as well as support and optimize processes.**”

Ref: Watson, Hugh J. (2014) "Tutorial: Big Data Analytics: Concepts, Technologies, and Applications," Communications of the Association for Information Systems: Vol. 34 , Article 65.