The CENTRE for EDUCATION
in MATHEMATICS and COMPUTING

Big Data

A CS and Society resource addressing social, environmental, and economic issues within the realm of digital technology and computing

This resource will:

- define Big Data, and explore how it is generated and where it is stored, and
- challenge you to consider both the benefits and potential harms of Big Data.

What Is Big Data?

Big Data
Extremely large, fast growing, and complicated datasets that need special tools and processes in order to be managed efficiently.

Big Data is generated through the use of digital technology and connected devices.

Think about your typical day, from the moment you wake up until you go to sleep.

Which of your daily activities create or share data?

How Big Is Big Data?

Take A Guess ...

LIVE STATS

- How many Google searches did you make today?
- How many Google searches were made worldwide today?
- How many emails were sent worldwide today?

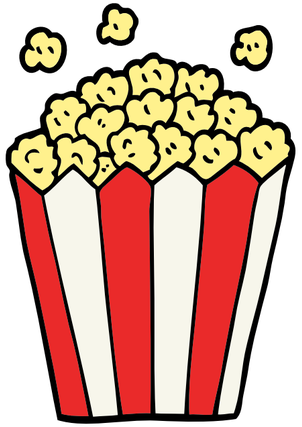
How much data in total do you think is generated worldwide in one day?

Is it the size of ...

10 billion movies?	40 billion movies?	80 billion movies?	100 billion movies?

Data Generated Worldwide in ... ONE DAY

According to Statista, approximately **402 million terabytes** (TB) of data is created daily!



Let's break it down:

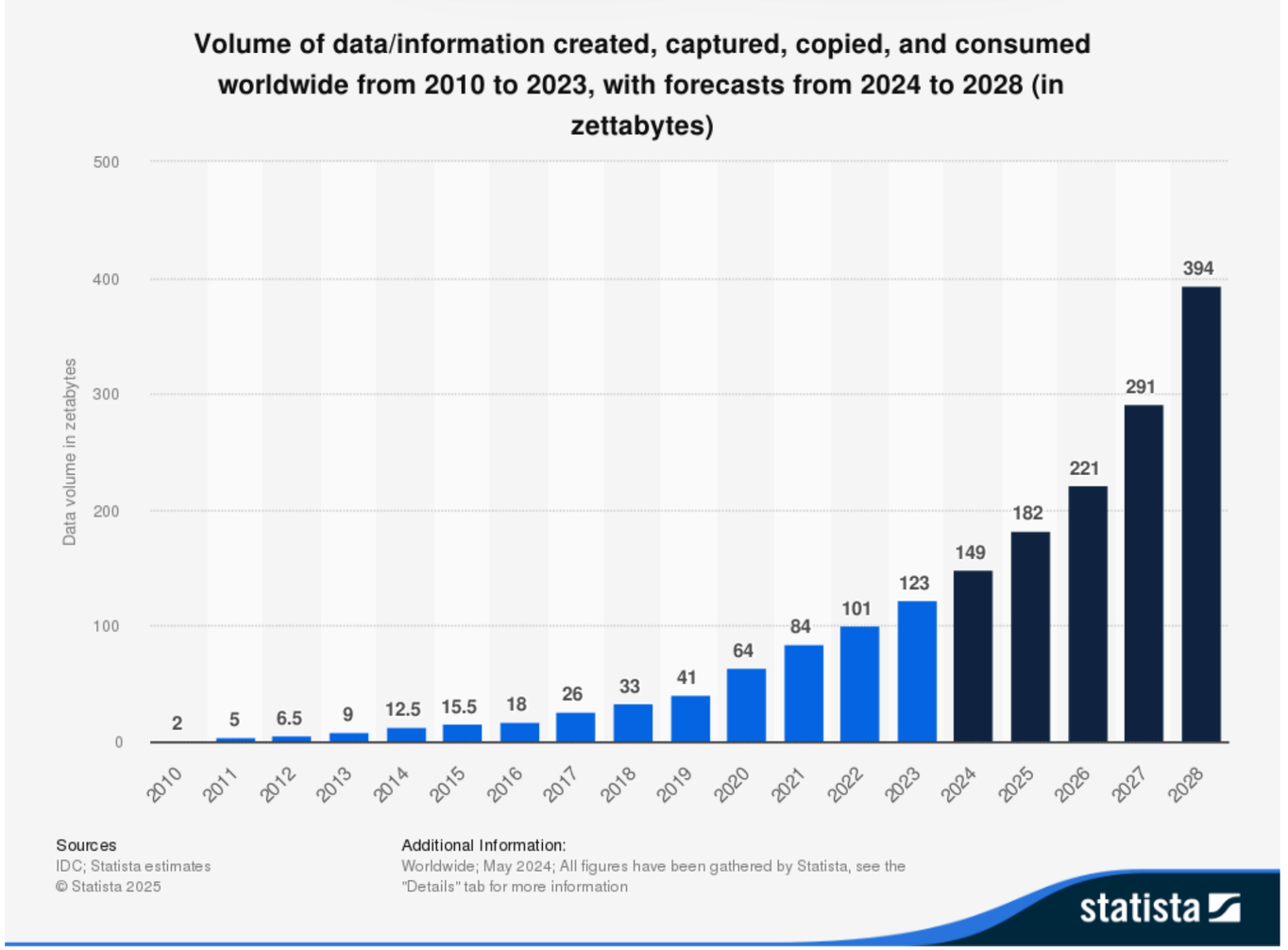
- 1 terabyte (TB) = 1000 gigabytes (GB)
- 402,000,000 TB = 402,000,000(1000) = 402,000,000,000 GB
- Assume 1 movie is approximately 5 GB of data
- Amount of data created daily ≈ 402,000,000,000 / 5 = 80,400,000,000 movies!

If we assume that 1 movie is 2 hours long, how long would it take to watch 80.4 billion movies?

Data Generated Worldwide in ... ONE YEAR

Annual amount of data over time:

- 2020 — 64 zettabytes
- 2022 — 101 zettabytes
- 2025 — 182 zettabytes
- 2028 — 394 zettabytes



Data Generated Worldwide in ... ONE YEAR

Approximately **182 zettabytes** (ZB) of data created in the year 2025!

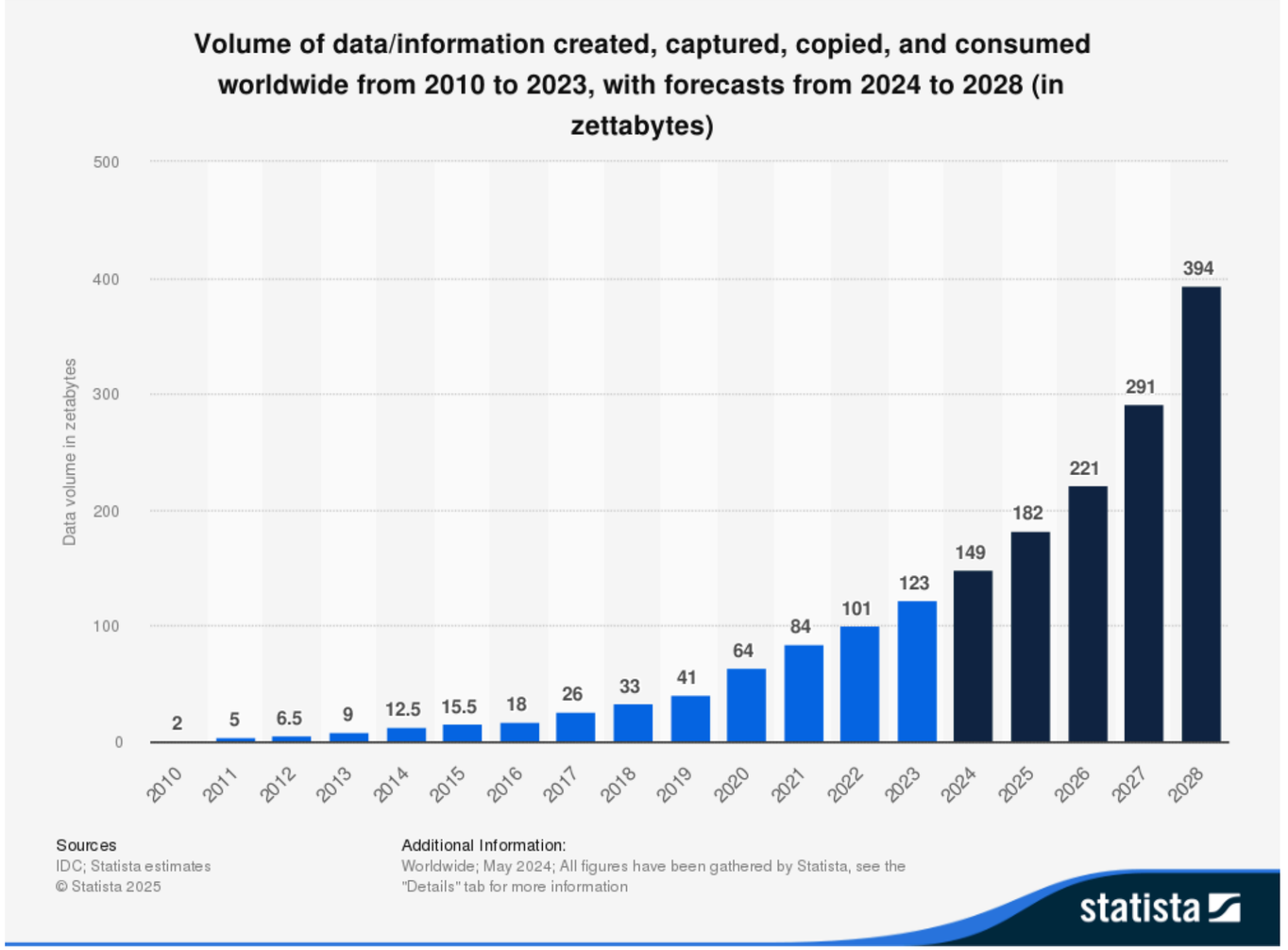
Let's break it down:

- 1 zettabtye (ZB) = 1,000,000,000 terabytes (TB)
- 1 terabyte (TB) = 1000 gigabytes (GB)
- 1 ZB = 1,000,000,000(1000) = 1,000,000,000,000 GB
- Amount of data created in the year 2025 ≈ 182 trillion GB!

A single zettabyte contains more bytes of data than there are grains of sand on all the world's beaches!

Discussion Questions

1. What do you notice about how much data the world is creating each year?
2. Why do you think the amount of data is increasing so quickly?
3. Where do you think all of this data is stored?



Why Is Data Growing So Quickly?

More People Online
An increasing number of people are connecting to the internet

More Content Creation
Photos, videos, and posts are shared constantly

More Smart Devices
Smart devices are gathering information 24/7

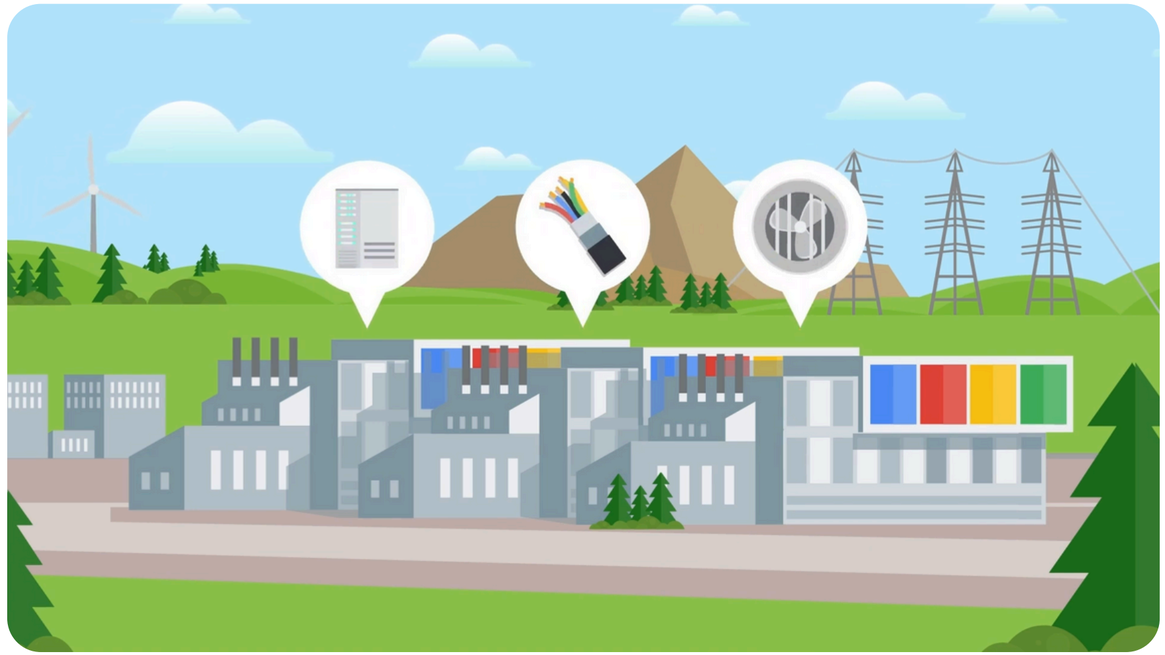
More Artificial Intelligence (AI)
Training and using AI generates a lot of data

More Storage
Cheap cloud storage encourages digitization of information

Where Is All This Data Stored?

All data is ultimately stored on physical devices. Even cloud storage is physical! **Data Centers** are specialized facilities specifically designed to store, manage, and process large volumes of data. The following video explains more.

VIDEO



Data Centers

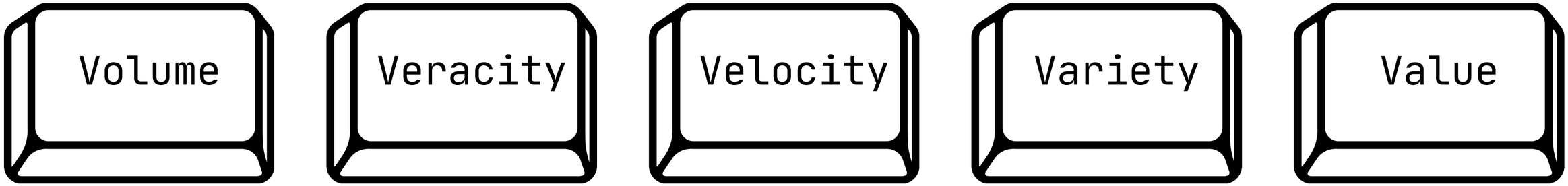
As Big Data grows exponentially, data centers are the backbone enabling real-time analytics, cloud storage, and artificial intelligence applications.

Take a look at Google's data centers worldwide and complete the scavenger hunt found under additional materials.

PHOTOS

The Five Vs of Big Data

Big Data can be categorized according to five defining characteristics:



- [x] Learn more about the Five Vs of Big Data: **ARTICLE**
- [x] Test yourself with a quiz: **QUIZ**
- [x] Complete the **Digital Platform Data** activity found under additional materials.

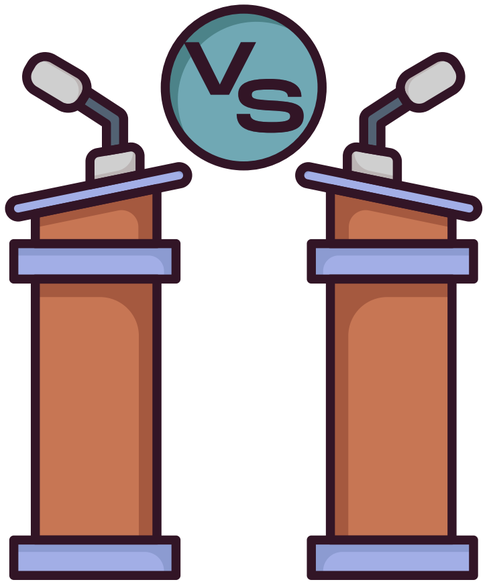
The Big Data Debate

Is Big Data beneficial or harmful to society?

In small groups, choose a **Big Data Debate** topic.

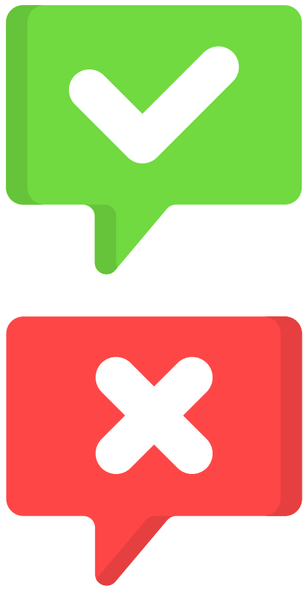
Divide your group into the *Pro* team and the *Con* team.

Each team should prepare 3-4 arguments to support their position.



The Big Data Debate: Topic A

Ownership and Profit



"Companies (such as Instagram) should be allowed to own and monetize the data they collect from people."

"The people who create the data should retain ownership and control."

Things to consider:

- Infrastructure costs
- Summarized data
- Informed consent vs. confusing Terms and Conditions

The Big Data Debate: Topic B

Privacy vs. Innovation

"Allowing access to Big Data drives innovations that benefit the greater good."

"Innovation does not justify surveillance of personal information."

Things to consider:

- Medicine and health care
- Public safety and security
- Biased or incomplete information
- Data misuse or leaks

The Big Data Debate: Topic C

Environmental Concerns

"Insights from Big Data can help environmental sustainability efforts."

"The creation and operation of data centers worsen existing environmental problems."

Things to consider:

- Optimization and efficiency
- Energy and water consumption
- E-Waste

The Big Data Debate: Topic D

Economic Impact

"Big Data creates new industries and job opportunities."

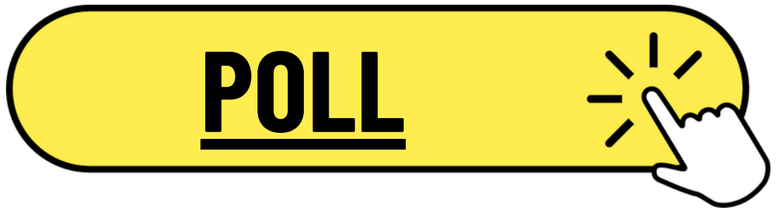
"Big Data causes people to lose their jobs."

Things to consider:

- Tech giants and competition
- Automation
- Data centers
- Wage gaps and reskilling

Closing Thoughts

Knowing what you now know about Big Data, what are your closing thoughts? Share anonymously using the following poll.



More Information and Resources

Big Data

- Google Cloud
- IBM
- Oracle
- Geeks for Geeks
- Data Per Day Stats
- Pros and Cons

Data Centers

- Google's Data Centers
- IBM
- Canada Energy Regulator
- Data Center Map

Your Role

- Your app permissions (Android / iOS)
- Your ad privacy (Chrome)
- Your digital energy consumption (Additional Materials)

Additional Materials

Data Center Hunt

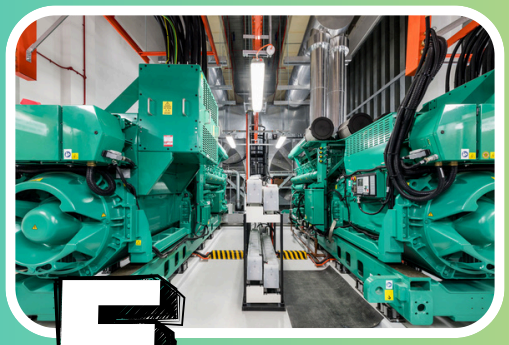
Locate these images in Google's Data Centers Photo Gallery.
Explain the significance of each photo by answering the questions on the following page.


A


B


C


D


E


F


G


H


I

Scavenger Hunt to accompany slide: Data Centers

Data Center Hunt

What equipment allows data centers to communicate with each other?

A

What part of a data center is shown in this photo?

B

What do the blue LEDs on these servers indicate?

C

What is this a photo of? What does it do?

D

What is this a photo of? Why are they important?

E

What do the blue pipes carry? What do the red pipes carry?

F

What components of a data center are shown in this photo?

G

What is this cooling plant using to cool its data center?

H

What is this a photo of? Why would a data center need this?

I

Scavenger Hunt to accompany slide: Data Centers

Data Center Hunt

A

What equipment allows data centers to communicate with each other?

Routers, switches, and fiber optic networks enable communication.

B

What part of a data center is shown in this photo?

This is a photo of a server aisle.

C

What do the blue LEDs on these servers indicate?

The blue LEDs indicate that the servers are running smoothly.

D

What is this a photo of? What does it do?

This is a photo of a transformer which helps distribute power throughout a data center.

E

What is this a photo of? Why are they important?

These are standby generators. They provide backup power to a data center.

F

What do the blue pipes carry? What do the red pipes carry?

The blue pipes supply cold water and the red pipes return warm water.

G

What components of a data center are shown in this photo?

This photo shows water storage tanks and cooling towers.

H

What is this cooling plant using to cool its data center?

This cooling plant in Hamina uses seawater from the Gulf of Finland.

I

What is this a photo of? Why would a data center need this?

This is a sprinkler riser room which is essential in the event of a fire.

Scavenger Hunt to accompany slide: Data Centers

Digital Platform Data

Choose your favourite digital platform (TikTok, Instagram, YouTube, Netflix, Amazon, Spotify, etc.) and complete the following:

1. What platform did you choose?
2. (Volume) Research how much data the platform generates daily.
3. (Velocity) Research how quickly the platform generates new data.
4. (Variety) Identify the types of data found on the platform.
5. (Veracity) Describe any accuracy or misinformation issues with the platform's data.
6. (Value) Describe how the platform's data is valuable to users, advertisers, and the platform itself.
7. Which of the Vs do you think is the biggest challenge for the platform to deal with? What makes it tough?
8. Describe one specific change you would make to this platform in order to improve one of the Vs.

Activity to accompany slide: The Five Vs of Big Data

Energy Estimation

How much energy do your daily digital activities actually consume? Estimate the following to the best of your ability:

Streaming videos = 0.15 kWh x _____ hours per day = _____ kWh

Sending text-only messages = 0.0003 kWh x _____ messages per day = _____ kWh

Sending picture messages = 0.003 kWh x _____ messages per day = _____ kWh

Posting short videos = 0.01 kWh x _____ videos per day = _____ kWh

Browsing social media = 0.06 kWh x _____ hours per day = _____ kWh

Video calling = 0.15 kWh x _____ hours per day = _____ kWh

Online gaming = 0.08 kWh x _____ hours per day = _____ kWh

Streaming music = 0.03 kWh x _____ hours per day = _____ kWh

Total Energy Use = _____ kWh

What small, realistic changes could you make to your daily digital habits to save energy?

Activity to accompany slide: More Information and Resources