Computers and Big Data

Guide: ShaneHadden
Generated: 2026-11-02 15:56

What is Big Data?

Big Data refers to extremely large datasets that cannot be easily managed, processed, or analyzed using traditional data processing tools. It encompasses data from various sources, including social media, sensors, transactions, and more. Big Data is characterized by its volume, velocity, and variety. Organizations use Big Data analytics to uncover patterns, trends, and insights that can inform decision-making and improve operations.

Why has big data grown over the last few decades?

Big data has grown due to the increase in digital data generation from sources like social media, IoT devices, and online transactions. Advances in technology, such as cloud computing and data storage, have made it easier to collect and analyze vast amounts of information. Additionally, businesses and organizations recognize the value of data for decision-making, leading to more investment in data analytics tools. This combination of factors has contributed to the exponential growth of big data.

What is the difference between structured and unstructured data?

Structured data is organized and easily searchable, typically found in databases and spreadsheets, using a predefined format like tables. Examples include names, dates, and transaction records. Unstructured data lacks a specific format, making it harder to analyze. It includes text, images, videos, and social media posts. While structured data is easier to process and analyze, unstructured data contains valuable insights but requires advanced tools and techniques to extract useful information.

Do computers convert all data to numbers for processing?

Yes, computers convert all data to numbers for processing. This includes text, images, and sounds, which are transformed into binary code (0s and 1s). Each type of data has a specific encoding method that allows computers to interpret and manipulate it. For example, text is encoded using ASCII or Unicode, while images are processed using formats like JPEG or PNG. Ultimately, regardless of the original format, data is represented numerically for computation.

What is binary?

Binary is a base-2 numeral system that uses only two digits: 0 and 1. It represents data in computers, where each digit is a bit. In binary, each position represents a power of 2, allowing computers to process and store information efficiently. For example, the binary number 101 represents the decimal number 5, calculated as (1×2^2) + (0×2^1) + (1×2^0). Binary is fundamental to computer operations, enabling everything from simple calculations to complex data processing.

How are 1s and 0s represented by a computer?

Computers represent 1s and 0s using binary code. Each 1 or 0 is called a bit. Bits are organized into groups, such as bytes (8 bits), to represent more complex data. Electrical signals indicate these bits: a high voltage might represent a 1, while low voltage represents a 0. This binary system allows computers to process and store data efficiently, enabling everything from simple calculations to complex applications.

How are bits stored when the computer is turned off?

When a computer is turned off, bits are stored in non-volatile memory, such as hard drives or solid-state drives (SSDs). These storage types retain data without power. In hard drives, bits are stored magnetically on spinning disks. In SSDs, bits are stored in flash memory cells that maintain their state even when the

power is off. This allows the computer to retrieve data when it is turned back on.

How has the cost of memory changed over time?

The cost of memory has significantly decreased over time due to advancements in technology and manufacturing processes. In the 1970s, a megabyte of RAM cost thousands of dollars, while today, gigabytes can be purchased for just a few dollars. This decline is driven by factors such as increased production efficiency, miniaturization of components, and the development of new materials. As a result, more powerful memory is now accessible, enabling the growth of big data applications and more comple

Can a computer generate a truly random number?

No, computers cannot generate truly random numbers. They use algorithms to produce pseudo-random numbers, which are determined by an initial value called a seed. While these numbers can appear random, they are not truly random because they can be reproduced if the seed is known. True randomness comes from unpredictable physical processes, like radioactive decay or thermal noise.

What is hexadecimal and why is it used?

Hexadecimal is a base-16 number system using digits 0-9 and letters A-F. It is used in computing because it efficiently represents binary data, making it easier for humans to read and write. Each hexadecimal digit corresponds to four binary digits (bits), allowing for a more compact representation of large binary numbers. This is especially useful in programming, memory addresses, and color codes in web design.

Can a computer process a number like 1/3?

Yes, a computer can process the number 1/3. However, it cannot represent it exactly in binary form because 1/3 is a repeating decimal (0.333...). Instead, computers use approximations, typically storing it as a floating-point number. This allows them to perform calculations with it, but there may be slight inaccuracies due to the limitations of binary representation.

What is quantum computing?

Quantum computing is a type of computing that uses quantum bits, or qubits, which can represent and store information in multiple states simultaneously, unlike classical bits that are either 0 or 1. This allows quantum computers to perform complex calculations much faster than traditional computers. They leverage principles of quantum mechanics, such as superposition and entanglement, to solve problems in areas like cryptography, optimization, and big data analysis more efficiently.

What is web scraping?

Web scraping is the process of automatically extracting data from websites. It involves using software or scripts to access web pages, retrieve specific information, and save it in a structured format, like a spreadsheet or database. This technique is commonly used for data analysis, market research, and gathering information from multiple sources. However, it's important to respect website terms of service and legal guidelines when scraping data.

What is an API?

An API, or Application Programming Interface, is a set of rules that allows different software applications to communicate with each other. It defines the methods and data formats that applications can use to request and exchange information. APIs enable developers to access features or data from other services, making it easier to build complex applications by integrating existing tools and resources.