

Supplemental Reading for Data Storage

Data Storage Measurements

In this reading, you will learn about the different names for measurements of data storage capacities and file sizes. Data storage capacity increases in step with the evolution of computer hardware technology. Larger storage capacities allow for dynamic growth in file sizes. These advances make it possible for companies like Netflix and Hulu to store thousands of feature-length films in high video quality formats.

There are standardized sets of terms used to name the ever expanding sizes of data storage and files. For example, the common terms used to describe file sizes and hard drive storage capacity include: bytes, kilobytes, megabytes, gigabytes, and terabytes. However, if you are a computer engineer, you might use a different set of terms.

Data storage measurement nomenclature

Table illustrating decimal values for data storage measurements

- **Decimal nomenclature:** kilobyte, megabyte, gigabyte, terabyte, petabyte, exabyte, zettabyte, yottabyte

The decimal naming system for computer storage uses the metric system of prefixes from the International System of Units: kilo, mega, giga, tera, peta, exa, zetta, and yotta. These prefixes may also be referred to as the decimal system of prefixes. The metric/decimal nomenclature represent a base-10 approximation of the actual amount of data storage bytes. The metric system prefixes were selected to simplify the marketing of computer products.

Table illustrating binary values for data storage measurements

- **Binary nomenclature:** kibibyte, mebibyte, gibibyte, tebibyte, pebibyte, exbibyte, zebibyte, yobibyte

The binary naming system is a standard set by the International Organization for Standardization (ISO) in partnership with the International Electrotechnical Commission (IEC). The ISO 80000 and IEC 80000 guides to units of measurement define the International System of Quantities (ISQ). The prefixes kibi-, mebi-, gibi-, -tebi-, pebi-, exbi-, zebi-, and yobi- were created by the IEC organization. They are a blend of the first two letters of the metric prefix fused with the first two letters of the word “binary” (example: **mega**byte + **binary** + **byte**= mebibyte).

Binary measurements of computer data are more accurate than decimal system measurements. While decimal nomenclature is commonly used to market computers and computer parts to the general public, binary nomenclature is often used in computer engineering for numerical accuracy.

Quantities of storage measurements

As data storage grows, the need for new terminology to describe the exponentially larger byte quantities grows too. The current byte nomenclature, mathematical representations, and storage capacities are as follows:

-

One bit: Also called a binary digit, bits store an electric signal as 1. The absence of an electric signal is stored as 0, which is also the default value of a bit. One bit can store only one value, either 1 or 0. These two possible values are the basis of the binary number system (base-2) that computers use. All numbers in a base-2 system increase exponentially as powers of 2.

- **One byte:** One byte stores eight bits of ones and zeros that translate to a symbol or basic computer instruction. Examples: 01101101 is the byte that translates to the letter “m.” The byte 01111111 tells the computer to delete the character to the right of the cursor.

- **One kilobyte (1 KB):**

- **Kilobyte (KB) decimal format:** $10^3 = 1,000$ bytes

- **Kibibyte (KiB) binary format:** $2^{10} = 1,024$ bytes

- **Decimal inaccuracy:** Off by -2.4% or -24 bytes

- **Name origin:** “Kilo-” is a French derivation from the Ancient Greek word for “thousand” A kilobyte is one thousand bytes.

- **1 KB can hold:** A short text file or a small icon as a 16x16 pixel .gif file.

- **One megabyte (1 MB):**
 - **Megabyte (MB) decimal format:** $10^6 = 1,000,000$ bytes
 - **Mebibyte (MiB) binary format:** $2^{20} = 1,048,576$ bytes
 - **Decimal inaccuracy:** Off by -4.9% or -48,576 bytes
 - **Name origin:** “Mega-” is derived from the Ancient Greek word for “large.” A megabyte is a large number of bytes.
 - **1 MB can hold:** Approximately one minute of music in a lossless .mp3 format or a short novel.

- **One gigabyte (1 GB):**
 - **Gigabyte (GB) decimal format:** $10^9 = 1,000,000,000$ bytes

-

Gibibyte (GiB) binary format: $2^{30} = 1,073,741,824$ bytes

-

Decimal inaccuracy: Off by -7.4% or -73,741,824 bytes

-

Name origin: “Giga-” is derived from the Ancient Greek word for “giant.” A gigabyte is a giant number of bytes.

-

1 GB can hold: Between 2.5-3 hours of music in .mp3 format or 300 high-resolution images.

-

One terabyte (1 TB):

-

Terabyte (TB) decimal format: $10^{12} = 1,000,000,000,000$ bytes

-

Tebibyte (TiB) binary format: $2^{40} = 1,099,511,627,776$ bytes

-

Decimal inaccuracy: Off by -10.0%

-

Name origin: “Tera-” is a shortened form of “tetra-”, which was derived from the Ancient Greek word for the number four. The 10^{12} decimal format can also be written as 1000^4 (one-thousand to the 4th power). “Tera-” in Ancient Greek means “monster.” You might think of the word “terabyte” as a monstrously large number of bytes.

-

1 TB can hold: Approximately 200,000 songs in .mp3 format or 300 hours of video.

-

One petabyte (PB):

-

Petabyte (PB) decimal format: $10^{15} = 1,000,000,000,000,000$ bytes

-

Pebibyte (PiB) binary format: $2^{50} = 1,125,899,906,842,624$ bytes

-

Decimal inaccuracy: Off by -12.6%

-

Name origin: “Peta-” is derived from the Ancient Greek word “penta” meaning five. The 10^{18} decimal format can also be written as 1000^5 (one-thousand to the 5th power).

-

1 PB can hold: The content from 1.5 million CD-ROM discs or 500 billion pages of text.

-

One exabyte (EB):

-

Exabyte (EB) decimal format: $10^{18} = 1,000,000,000,000,000,000$ bytes

-

Exbibyte (EiB) binary format: $2^{60} = 1,152,921,504,606,846,976$ bytes

-

Decimal inaccuracy: Off by -15.3%

-

Name origin: “Exa-” was derived from the Ancient Greek word for six. The 10^{18} decimal format can also be written as 1000^6 (one-thousand to the 6th power).

-

1 EB can hold: Approximately 11 million movies in 4k video resolution or 3,000 copies of the entire United States Library of Congress.

-

One zettabyte (ZB):

- **Zettabyte (ZB) decimal format:** $10^{21} = 1,000,000,000,000,000,000,000$ bytes
- **Zebibyte (ZiB) binary format:** $2^{70} = 1,180,591,620,717,411,303,424$ bytes
- **Decimal inaccuracy:** Off by -18.1%
- **Name origin:** “Zetta” was derived from the Latin word “septem” which means seven. The 10^{21} decimal format can also be written as 1000^7 (one-thousand to the 7th power).
- **1 ZB can hold:** Seagate reports one zettabyte can hold 30 billion movies in 4k video resolution.

- ## One yottabyte (YB):

- **Yottabyte (YB) decimal format:** $10^{24} = 1,000,000,000,000,000,000,000,000$ bytes

- **Yobibyte (YiB) binary format:** $2^{80} = 1,208,925,819,614,629,174,706,176$ bytes
- **Decimal inaccuracy:** Off by -20.9%
- **Name origin:** “Yotta” is Ancient Greek for eight. The 10^{24} decimal format can also be written as 10008 (one-thousand to the 8th power).
- **1 YB can hold:** In 2011, a cloud storage company estimated that one yottabyte could hold the data of one million data centers.
- In 2023 its estimated that the total size of the internet is larger than 1 yottabyte
-
-

Ask ChatGPT4

how large is the internet estimated to be in exabytes?

○

“

The size of the Internet or the total amount of data created globally is not specified directly in exabytes in the sources I found. However, we can infer some information based on the data provided. It was expected that 2020 would see the creation of 44 zettabytes (1 zettabyte = 1 million petabytes = 1 billion terabytes = 1 trillion gigabytes = 1,000 exabytes) of data, and the world was expected to generate 103 zettabytes of data by 2023. The impending age of the yottabyte (1,000 zettabytes) is in sight, implying that we are moving towards generating and managing data volumes in the order of thousands of zettabytes^{1?}.

In terms of storage, advancements in technology are being made to manage this enormous amount of data. For instance, IBM's 'Elastic Storage' technology, which was part of the Watson system, is designed to handle architectural limits stretching into the thousands of yottabytes (1 yottabyte = 1 million petabytes)^{2?}. Moreover, an increasing amount of data is unstructured, showing a compound annual growth rate (CAGR) of 38%, and solutions are being developed to handle this surge in data, including innovations in software, hardware, and algorithms for scale-out storage and the implementation of an intelligent data fabric to manage data across systems, regions, and clouds^{3?}.

In 2022, the combined shipment volume of the world's top five External Controller-Based (ECB) storage vendors was triple the volume in 2012, but it is still insufficient to meet the demands of the upcoming yottabyte era. Huawei projects that this figure will multiply ten times by 2032, surpassing 100 exabytes. However, this figure pertains to the storage industry and not the total size of the Internet or global data volume^{3?}.

In summary, while a specific figure in exabytes for the total size of the Internet in 2023 isn't readily available, we are currently in an era where we are generating and dealing with data in the scale of zettabytes (each of which is equivalent to 1,000 exabytes), and moving towards the yottabyte era. I would need more time to search further to find a more precise estimate in

exabytes for the total size of the Internet in 2023.

Revision #4

Created 2023-05-29 09:29:53 UTC by naruzkurai

Updated 2023-05-29 10:25:50 UTC by naruzkurai