**Understand how data is measured**



Data is measured by the number of **bits** it takes to represent it. All information in a computer can be represented as a binary number consisting solely of 0’s and 1’s. Each 0 or 1 in a number is a bit. A bit is the smallest unit of storage in computers. Since computers work in binary (Base 2), this means that all the important numbers that differentiate between different data sizes will be powers of 2.

A **byte** is a collection of 8 bits. Take a moment to examine the table below to get a feel for the difference between data measurements and their relative sizes to one another.

| **Unit** | **Equivalent to** | **Abbreviation** | **Real-World Example** |
| --- | --- | --- | --- |
| Byte | 8 bits | B | 1 character in a string |
| Kilobyte | 1024 bytes | KB | A page of text (~4 kilobytes) |
| Megabyte | 1024 Kilobytes | MB | 1 song in MP3 format (~2-3 megabytes) |
| Gigabyte | 1024 Megabytes | GB | ~300 songs in MP3 format |
| Terabyte | 1024 Gigabytes | TB | ~500 hours of HD video |
| Petabyte | 1024 Terabytes | PB | 10 billion Facebook photos |
| Exabyte | 1024 Petabytes | EB | ~500 million hours of HD video |
| Zettabyte | 1024 Exabytes | ZB | All the data on the internet in 2019 (~4.5 ZB) |

The amount of data in the world is exploding and growing at an incredible pace every year. This growth is largely the result of the over 4.6 billion people around the world connected to the Internet. Now that smartphones and other Internet-connected devices have become common, they generate a staggering amount of new data. Many experts believe that the size of all the data on the Internet will swell to 175 ZB by the end of 2025!

The size of the dataset you’re working with usually determines which tool, spreadsheets or SQL, is best suited for the task. Spreadsheets often start to have performance issues as dataset sizes increase beyond a few megabytes. SQL databases are much better at working with larger datasets that have billions of rows with sizes measured in gigabytes. The dataset’s size still matters here--larger datasets will take longer for queries to complete, depending on the query’s content and the number of rows SQL has to process to complete the query.