

Description

The first Apple Macintosh, which celebrated its 30th birthday in 2014, came with 128KB (0.13MB) of memory!

Today, we buy laptops and external hard drives with 1Terabyte (1,024GB) of memory as we move forward to even larger memory capacities of *Petabytes* (1,024TB), *Exabytes* (1,024PB), *Zettabytes* (1,024 EB), and *Yottabytes* (1, 024ZB). For comparison, the Large Hadron Collider (LHC) in Switzerland, produces around 15PB of data every year.

A Management Problem

Technology has clearly been generous in providing massive amounts of storage capacity (provided you can afford the infrastructure to physically store it—even the cloud has a physical address). However, for academic researchers, the problem isn't the amount of data that can be stored, it's what you do with that data once you have it.

Global access to multiple individual databases from multiple countries, or from global data aggregators, has the potential to rapidly advance the rate of scientific discovery, but behind that potential lies some critical assumptions about the usability of all this data.

Gigo – Garbage in, Garbage out

Google's path to world dominance as a search engine has been based on the *kaizen* approach with its search algorithm—constant improvement to stay one step ahead of the *black hat* hackers and to provide the most accurate responses to your search requests. Individual researchers, librarians, and database managers are equally dependent on the quality of the search algorithm being used. If the data has been categorized correctly, and the search algorithm aligns with that categorization, you have a fighting chance of generating relevant results. Without it, the sheer volume of wrong data that can be generated from a multi-terabyte database has the potential to derail your literature review or research study by burning hundreds of hours of valuable research time.

The increased availability of such large volumes of data has already given rise to more sophisticated analysis based on patterns of data in huge datasets. The accuracy of such research is totally

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

dependent on the quality of that data, and with such large numbers, the potential to fabricate results by continuing to look until you find the pattern you're looking for may be too tempting to ignore.

More ? Better

The blessing of access to large volumes of data is definitely a mixed one. Since data serves as the foundation for any research work, access to larger datasets for larger studies, or more journals for a more comprehensive literature review, can be a tremendous asset in producing quality research. However, the current rate of increase of data availability exceeds the tools and expertise we have available to manage that data. Institutions are faced with budgetary decisions about purchasing access to data on an invisible cloud, or building their own cloud to ensure reliable access. They must also decide if their in-house data specialists have the skills to navigate through these huge databases or whether they need to hire new specialists with more targeted skills, and even if you find those specialists, will they have access to the latest modeling tools to produce something useful?

Category

1. Industry News
2. Publishing News

Date Created
2016/04/18
Author
admin

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license