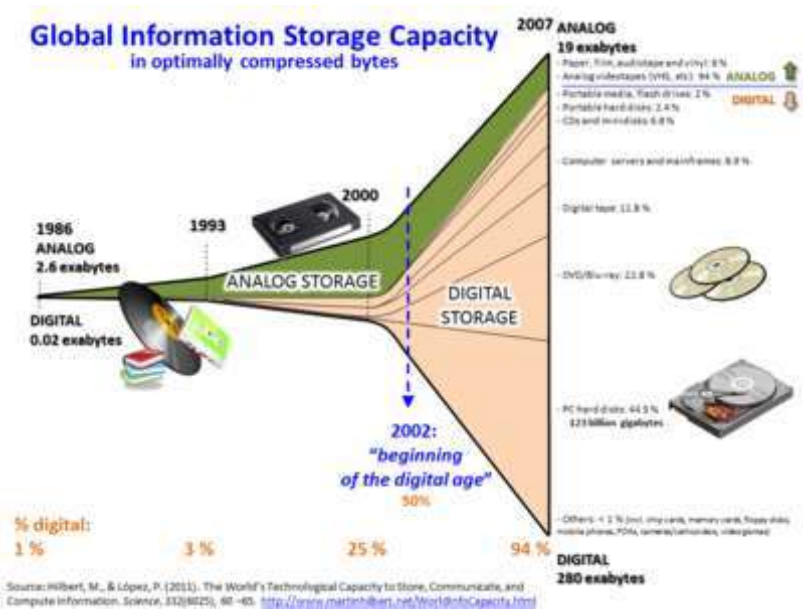
WIKIPEDIA

Big data

Big data refers to data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many fields (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate.[2] Big data analysis challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy, and data source. Big data was originally associated with three key concepts: *volume*, *variety*, and *velocity*.[3] The analysis of big data presents challenges in sampling, and thus previously allowing for only observations and sampling. Thus a fourth concept, *veracity,* refers to the quality or insightfulness of the data. Without sufficient investment in expertise for big data veracity, then the volume and variety of data can produce costs and risks that exceed an organization's capacity to create and capture *value* from big data.[4]



Non-linear growth of digital global information-storage capacity and the waning of analog storage[1]

Current usage of the term *big data* tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from big data, and seldom to a particular size of data set. "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem."[5] Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so on".[6] Scientists, business executives, medical practitioners, advertising and governments alike regularly meet difficulties with large data-sets in areas including Internet searches, fintech, healthcare analytics, geographic information systems, urban informatics, and business informatics. Scientists encounter limitations in e-Science work, including meteorology, genomics,[7] connectomics, complex physics simulations, biology, and environmental research.[8]

The size and number of available data sets have grown rapidly as data is collected by devices such as mobile devices, cheap and numerous information-sensing Internet of things devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks.[9][10] The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s;[11] as of 2012, every day 2.5 exabytes (2.5×2^{60} bytes) of data are generated.[12] Based on an IDC report prediction, the global data volume was predicted to grow exponentially from 4.4 zettabytes to 44 zettabytes between 2013 and 2020. By 2025, IDC predicts there will be 163 zettabytes of data.[13] According to IDC, global spending on big data and business analytics (BDA) solutions is estimated to reach $215.7 billion in 2021.[14][15] While