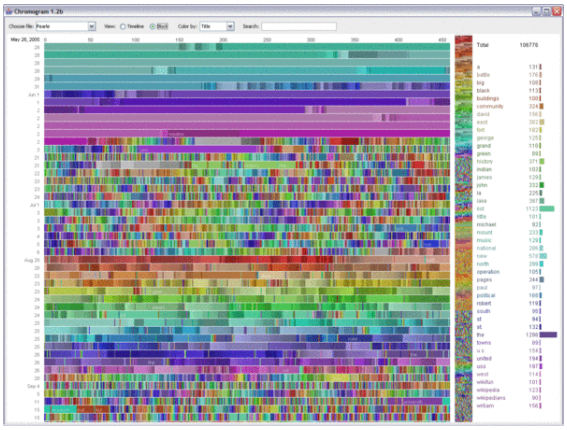
Big data

This article is about large collections of data. For the graph database, see Graph database. For the band, see Big Data (band).

Big data is an all-encompassing term for any collection



A visualization of Wikipedia edits created by IBM. At multiple terabytes in size, the text and images of Wikipedia are a classic example of big data.

of data sets so large and complex that it becomes difficult to process using traditional data processing applications.

The challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and privacy violations. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, prevent diseases, combat crime and so on."[1]

Scientists regularly encounter limitations due to large data sets in many areas, including meteorology, genomics,[2] connectomics, complex physics simulations,[3] and biological and environmental research.[4] The limitations also affect Internet search, finance and business informatics. Data sets grow in size in part because they are increasingly being gathered by ubiquitous information-sensing mobile devices, aerial sensory technologies (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks.[5][6][7] The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s;[8] as of 2012, every day 2.5 exabytes (2.5×10^{18}) of data were created.[9] The challenge for large enterprises is determining who should own big data initiatives that straddle the entire organization.[10]

Big data is difficult to work with using most relational database management systems and desktop statistics and visualization packages, requiring instead "massively parallel software running on tens, hundreds, or even thousands of servers".[11] What is considered "big data" varies depending on the capabilities of the organization managing the set, and on the capabilities of the applications that are traditionally used to process and analyze the data set in its domain. Big Data is a moving target; what is considered to be "Big" today will not be so years ahead. "For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration."[12]

1 Definition

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time.[13] Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data.

In a 2001 research report[14] and related lectures, META Group (now Gartner) analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional, i.e. increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). Gartner, and now much of the industry, continue to use this "3Vs" model for describing big data.[15] In 2012, Gartner updated its definition as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."[16] Additionally, a new V "Veracity" is added by some organizations to describe it.[17]

If Gartner's definition (the 3Vs) is still widely used, the growing maturity of the concept fosters a more sound difference between big data and Business Intelligence, regarding data and their use:[18]

- Business Intelligence uses descriptive statistics with data with high information density to measure things, detect trends etc.;
- Big data uses inductive statistics and concepts from

nonlinear system identification [19] to infer laws (regressions, nonlinear relationships, and causal effects) from large sets of data with low information density[20] to reveal relationships, dependencies and perform predictions of outcomes and behaviors.[19][21]

Big data can also be defined as "Big data is a large volume unstructured data which can not be handled by standard database management systems like DBMS, RDBMS or ORDBMS".

2 Examples

2.1 Big science

The Large Hadron Collider experiments represent about 150 million sensors delivering data 40 million times per second. There are nearly 600 million collisions per second. After filtering and refraining from recording more than 99.999% of these streams, there are 100 collisions of interest per second.[22][23][24]

- As a result, only working with less than 0.001% of the sensor stream data, the data flow from all four LHC experiments represents 25 petabytes annual rate before replication (as of 2012). This becomes nearly 200 petabytes after replication.
- If all sensor data were to be recorded in LHC, the data flow would be extremely hard to work with. The data flow would exceed 150 million petabytes annual rate, or nearly 500 exabytes per day, before replication. To put the number in perspective, this is equivalent to 500 quintillion (5×10^{20}) bytes per day, almost 200 times more than all the other sources combined in the world.

The Square Kilometre Array is a telescope which consists of millions of antennas and is expected to be operational by 2024. Collectively, these antennas are expected to gather 14 exabytes and store one petabyte per day.[25][26] It is considered to be one of the most ambitious scientific projects ever undertaken.

2.2 Science and research

- When the Sloan Digital Sky Survey (SDSS) began collecting astronomical data in 2000, it amassed more in its first few weeks than all data collected in the history of astronomy. Continuing at a rate of about 200 GB per night, SDSS has amassed more than 140 terabytes of information. When the Large Synoptic Survey Telescope, successor to SDSS, comes online in 2016 it is anticipated to acquire that amount of data every five days.[1]
- Decoding the human genome originally took 10 years to process, now it can be achieved in less than a day: the DNA sequencers have divided the sequencing cost by 10,000 in the last ten years, which is 100 times cheaper than the reduction in cost predicted by Moore's Law.[27]
- The NASA Center for Climate Simulation (NCCS) stores 32 petabytes of climate observations and simulations on the Discover supercomputing cluster.[28]

2.3 Government

- In 2012, the Obama administration announced the Big Data Research and Development Initiative, to explore how big data could be used to address important problems faced by the government.[29] The initiative is composed of 84 different big data programs spread across six departments.[30]
- Big data analysis played a large role in Barack Obama's successful 2012 re-election campaign.[31]
- The United States Federal Government owns six of the ten most powerful supercomputers in the world.[32]
- The Utah Data Center is a data center currently being constructed by the United States National Security Agency. When finished, the facility will be able to handle a large amount of information collected by the NSA over the Internet. The exact amount of storage space is unknown, but more recent sources claim it will be on the order of a few exabytes.[33][34][35]

2.4 Private sector

Bus wrapped with SAP Big data parked outside IDF13.

- eBay.com uses two data warehouses at 7.5 petabytes and 40PB as well as a 40PB Hadoop cluster for search, consumer recommendations, and merchandising. Inside eBay's 90PB data warehouse

- Amazon.com handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers. The core technology that keeps Amazon running is Linux-based and as of 2005 they had the world's three largest Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB.[36]
- Walmart handles more than 1 million customer transactions every hour, which are imported into databases estimated to contain more than 2.5 petabytes (2560 terabytes) of data – the equivalent of 167 times the information contained in all the books in the US Library of Congress.[1]
- Facebook handles 50 billion photos from its user base.[37]
- FICO Falcon Credit Card Fraud Detection System protects 2.1 billion active accounts world-wide.[38]
- The volume of business data worldwide, across all companies, doubles every 1.2 years, according to estimates.[39][40]
- Windermere Real Estate uses anonymous GPS signals from nearly 100 million drivers to help new home buyers determine their typical drive times to and from work throughout various times of the day.[41]

2.5 International development

Research on the effective usage of information and communication technologies for development (also known as ICT4D) suggests that big data technology can make important contributions but also present unique challenges to International development.[42][43] Advancements in big data analysis offer cost-effective opportunities to improve decision-making in critical development areas such as health care, employment, economic productivity, crime, security, and natural disaster and resource management.[44][45] However, longstanding challenges for developing regions such as inadequate technological infrastructure and economic and human resource scarcity exacerbate existing concerns with big data such as privacy, imperfect methodology, and interoperability issues.[44]

3 Characteristics

Big data can be described by the following characteristics:

Volume – The quantity of data that is generated is very important in this context.It is the size of the data which determines the value and potential of the data under consideration and whether it can actually be considered as Big Data or not.The name 'Big Data' itself contains a term which is related to size and hence the characteristic.

Variety - The next aspect of Big Data is its variety.This means that the category to which Big Data belongs to is also a very essential fact that needs to be known by the data analysts.This helps the people, who are closely analyzing the data and are associated with it, to effectively use the data to their advantage and thus upholding the importance of the Big Data.

Velocity - The term 'velocity' in the context refers to the speed of generation of data or how fast the data is generated and processed to meet the demands and the challenges which lie ahead in the path of growth and development.

Variability - This is a factor which can be a problem for those who analyse the data. This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

Complexity - Data management can become a very complex process,especially when large volumes of data come from multiple sources.These data need to be linked,connected and correlated in order to be able to grasp the information that is supposed to be conveyed by these data.This situation,is therefore,termed as the 'complexity' of Big Data.

4 Market

Big data has increased the demand of information management specialists in that Software AG, Oracle Corporation, IBM, FICO, Microsoft, SAP, EMC, HP and Dell have spent more than $15 billion on software firms specializing in data management and analytics. In 2010, this industry was worth more than $100 billion and was growing at almost 10 percent a year: about twice as fast as the software business as a whole.[1]

Developed economies make increasing use of data-intensive technologies. There are 4.6 billion mobile-phone subscriptions worldwide and between 1 billion and 2 billion people accessing the internet.[1] Between 1990 and 2005, more than 1 billion people worldwide entered the middle class which means more and more people who gain money will become more literate which in turn leads to information growth. The world's effective capacity to exchange information through telecommunication networks was 281 petabytes in 1986, 471 petabytes in 1993, 2.2 exabytes in 2000, 65 exabytes in 2007[8] and it is predicted that the amount of traffic flowing over the internet will reach 667 exabytes annually by 2014.[1] It is esti-

mated that one third of the globally stored information is in the form of alphanumeric text and still image data,[46] which is the format most useful for most big data applications. This also shows the potential of yet unused data (i.e. in the form of video and audio content).

While many vendors offer off-the-shelf solutions for Big Data, experts recommend the development of in-house solutions custom-tailored to solve the companies problem at hand if the company has sufficient technical capabilities.[47]

5 Architecture

In 2000, Seisint Inc. develops C++ based distributed file sharing framework for data storage and querying. Structured, semi-structured and/or unstructured data is stored and distributed across multiple servers. Querying of data is done by modified C++ called ECL which uses apply scheme on read method to create structure of stored data during time of query. In 2004 LexisNexis acquired Seisint Inc.[48] and 2008 acquired ChoicePoint, Inc.[49] and their high speed parallel processing platform. The two platforms were merged into HPCC Systems and in 2011 was open sourced under Apache v2.0 License. Currently HPCC and Quantcast File System[50] are the only publicly available platforms capable of analyzing multiple exabytes of data.

In 2004, Google published a paper on a process called MapReduce that used such an architecture. The MapReduce framework provides a parallel processing model and associated implementation to process huge amount of data. With MapReduce, queries are split and distributed across parallel nodes and processed in parallel (the Map step). The results are then gathered and delivered (the Reduce step). The framework was very successful,[51] so others wanted to replicate the algorithm. Therefore, an implementation of the MapReduce framework was adopted by an Apache open source project named Hadoop.[52]

MIKE2.0 is an open approach to information management that acknowledges the need for revisions due to big data implications in an article titled "Big Data Solution Offering".[53] The methodology addresses handling big data in terms of useful permutations of data sources, complexity in interrelationships, and difficulty in deleting (or modifying) individual records.[54]

Recent studies show that the use of a multiple layer architecture is an option for dealing with big data. The Distributed Parallel architecture distributes data across multiple processing units and parallel processing units provide data much faster, by improving processing speeds. This type of architecture inserts data into a parallel DBMS, which implements the use of MapReduce and Hadoop frameworks. This type of framework looks to make the processing power transparent to the end user by using a front end application server.[55]

6 Technologies

Big data requires exceptional technologies to efficiently process large quantities of data within tolerable elapsed times. A 2011 McKinsey report[56] suggests suitable technologies include A/B testing, crowdsourcing, data fusion and integration, genetic algorithms, machine learning, natural language processing, signal processing, simulation, time series analysis and visualisation. Multidimensional big data can also be represented as tensors, which can be more efficiently handled by tensor-based computation,[57] such as multilinear subspace learning.[58] Additional technologies being applied to big data include massively parallel-processing (MPP) databases, search-based applications, data-mining grids, distributed file systems, distributed databases, cloud based infrastructure (applications, storage and computing resources) and the Internet.

Some but not all MPP relational databases have the ability to store and manage petabytes of data. Implicit is the ability to load, monitor, back up, and optimize the use of the large data tables in the RDBMS.[59]

DARPA's Topological Data Analysis program seeks the fundamental structure of massive data sets and in 2008 the technology went public with the launch of a company called Ayasdi.[60]

The practitioners of big data analytics processes are generally hostile to slower shared storage,[61] preferring direct-attached storage (DAS) in its various forms from solid state drive (SSD) to high capacity SATA disk buried inside parallel processing nodes. The perception of shared storage architectures—Storage area network (SAN) and Network-attached storage (NAS) —is that they are relatively slow, complex, and expensive. These qualities are not consistent with big data analytics systems that thrive on system performance, commodity infrastructure, and low cost.

Real or near-real time information delivery is one of the defining characteristics of big data analytics. Latency is therefore avoided whenever and wherever possible. Data in memory is good—data on spinning disk at the other end of a FC SAN connection is not. The cost of a SAN at the scale needed for analytics applications is very much higher than other storage techniques.

There are advantages as well as disadvantages to shared storage in big data analytics, but big data analytics practitioners as of 2011 did not favour it.[62]

7 Research activities

Encrypted search and cluster formation in big data was demonstrated in March 2014 at the American Society of Engineering Education. Gautam Siwach engaged at *Tackling the challenges of Big Data* by MIT Computer Science and Artificial Intelligence Laboratory and Dr. Amir Esmailpour at UNH Research Group investigated the key features of big data as formation of clusters and their interconnections. They focused on the security of big data and the actual orientation of the term towards the presence of different type of data in an encrypted form at cloud interface by providing the raw definitions and real time examples within the technology. Moreover, they proposed an approach for identifying the encoding technique to advance towards an expedited search over encrypted text leading to the security enhancements in big data.[63]

In March 2012, The White House announced a national "Big Data Initiative" that consisted of six Federal departments and agencies committing more than $200 million to big data research projects.[64]

The initiative included a National Science Foundation "Expeditions in Computing" grant of $10 million over 5 years to the AMPLab[65] at the University of California, Berkeley.[66] The AMPLab also received funds from DARPA, and over a dozen industrial sponsors and uses big data to attack a wide range of problems from predicting traffic congestion[67] to fighting cancer.[68]

The White House Big Data Initiative also included a commitment by the Department of Energy to provide $25 million in funding over 5 years to establish the Scalable Data Management, Analysis and Visualization (SDAV) Institute,[69] led by the Energy Department's Lawrence Berkeley National Laboratory. The SDAV Institute aims to bring together the expertise of six national laboratories and seven universities to develop new tools to help scientists manage and visualize data on the Department's supercomputers.

The U.S. state of Massachusetts announced the Massachusetts Big Data Initiative in May 2012, which provides funding from the state government and private companies to a variety of research institutions.[70] The Massachusetts Institute of Technology hosts the Intel Science and Technology Center for Big Data in the MIT Computer Science and Artificial Intelligence Laboratory, combining government, corporate, and institutional funding and research efforts.[71]

The European Commission is funding the 2-year-long Big Data Public Private Forum through their Seventh Framework Program to engage companies, academics and other stakeholders in discussing big data issues. The project aims to define a strategy in terms of research and innovation to guide supporting actions from the European Commission in the successful implementation of the big data economy. Outcomes of this project will be used as input for Horizon 2020, their next framework program.[72]

The British government announced in March 2014 the founding of the Alan Turing Institute, named after the computer pioneer and code-breaker, which will focus on new ways of collecting and analysing large sets of data.[73]

At the University of Waterloo Stratford Campus Canadian Open Data Experience (CODE) Inspiration Day, it was demonstrated how using data visualization techniques can increase the understanding and appeal of big data sets in order to communicate a story to the world.[74]

In order to make manufacturing more competitive in the United States (and globe), there is a need to integrate more American ingenuity and innovation into manufacturing ; Therefore, National Science Foundation has granted the Industry University cooperative research center for Intelligent Maintenance Systems (IMS) at university of Cincinnati to focus on developing advanced predictive tools and techniques to be applicable in a big data environment.[75][76] In May 2013, IMS Center held an industry advisory board meeting focusing on big data where presenters from various industrial companies discussed their concerns, issues and future goals in Big Data environment.

Computational social sciences — Anyone can use Application Programming Interfaces (APIs) provided by Big Data holders, such as Google and Twitter, to do research in the social and behavioral sciences.[77] Often these APIs are provided for free.[77] Tobias Preis *et al.* used Google Trends data to demonstrate that Internet users from countries with a higher per capita gross domestic product (GDP) are more likely to search for information about the future than information about the past. The findings suggest there may be a link between online behaviour and real-world economic indicators.[78][79][80] The authors of the study examined Google queries logs made by ratio of the volume of searches for the coming year ('2011') to the volume of searches for the previous year ('2009'), which they call the 'future orientation index'.[81] They compared the future orientation index to the per capita GDP of each country and found a strong tendency for countries in which Google users enquire more about the future to exhibit a higher GDP. The results hint that there may potentially be a relationship between the economic success of a country and the information-seeking behavior of its citizens captured in big data.

Tobias Preis and his colleagues Helen Susannah Moat and H. Eugene Stanley introduced a method to identify online precursors for stock market moves, using trading strategies based on search volume data provided by Google Trends.[82] Their analysis of Google search volume for 98 terms of varying financial relevance, published in *Scientific Reports*,[83] suggests that increases in search volume for financially relevant search terms tend to precede large losses in financial markets.[84][85][86][87][88][89][90][91]

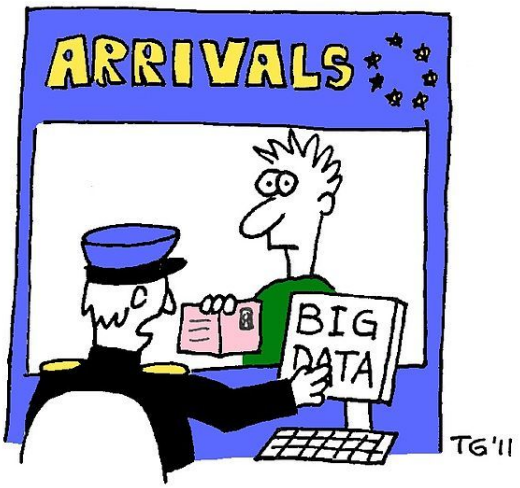
8 Applications

8.1 Manufacturing

Based on TCS 2013 Global Trend Study, improvements in supply planning and product quality provide the greatest benefit of big data for manufacturing.[92] Big data provides an infrastructure for transparency in manufacturing industry, which is the ability to unravel uncertainties such as inconsistent component performance and availability. Predictive manufacturing as an applicable approach toward near-zero downtime and transparency requires vast amount of data and advanced prediction tools for a systematic process of data into useful information.[93] A conceptual framework of predictive manufacturing begins with data acquisition where different type of sensory data is available to acquire such as acoustics, vibration, pressure, current, voltage and controller data. Vast amount of sensory data in addition to historical data construct the big data in manufacturing. The generated big data acts as the input into predictive tools and preventive strategies such as Prognostics and Health Management (PHM).[75]

9 Critique

Critiques of the big data paradigm come in two flavors, those that question the implications of the approach itself, and those that question the way it is currently done.



"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."

Cartoon critical of big data application, by T. Gregorius

9.1 Critiques of the big data paradigm

"A crucial problem is that we do not know much about the underlying empirical micro-processes that lead to the emergence of the[se] typical network characteristics of Big Data".[13] In their critique, Snijders, Matzat, and Reips point out that often very strong assumptions are made about mathematical properties that may not at all reflect what is really going on at the level of micro-processes. Mark Graham has leveled broad critiques at Chris Anderson's assertion that big data will spell the end of theory: focusing in particular on the notion that big data will always need to be contextualized in their social, economic and political contexts.[94] Even as companies invest eight- and nine-figure sums to derive insight from information streaming in from suppliers and customers, less than 40% of employees have sufficiently mature processes and skills to do so. To overcome this insight deficit, "big data", no matter how comprehensive or well analyzed, needs to be complemented by "big judgment", according to an article in the Harvard Business Review.[95]

Much in the same line, it has been pointed out that the decisions based on the analysis of big data are inevitably "informed by the world as it was in the past, or, at best, as it currently is".[44] Fed by a large number of data on past experiences, algorithms can predict future development if the future is similar to the past. If the systems dynamics of the future change, the past can say little about the future. For this, it would be necessary to have a thorough understanding of the systems dynamic, which implies theory.[96] As a response to this critique it has been suggested to combine big data approaches with computer simulations, such as agent-based models.[44] Agent-based models are increasingly getting better in predicting the outcome of social complexities of even unknown future scenarios through computer simulations that are based on a collection of mutually interdependent algorithms.[97][98] In addition, use of multivariate methods that probe for the latent structure of the data, such as factor analysis and cluster analysis, have proven useful as analytic approaches that go well beyond the bi-variate approaches (cross-tabs) typically employed with smaller data sets.

In health and biology, conventional scientific approaches are based on experimentation. For these approaches, the limiting factor are the relevant data that can confirm or refute the initial hypothesis.[99] A new postulate is accepted now in biosciences: the information provided by the data in huge volumes (omics) without prior hypothesis is complementary and sometimes necessary to conventional approaches based on experimentation. In the massive approaches it is the formulation of a relevant hypothesis to explain the data that is the limiting factor. The search logic is reversed and the limits of induction ("Glory of Science and Philosophy scandal", C. D. Broad, 1926) to be considered.

Privacy advocates are concerned about the threat to privacy represented by increasing storage and integration of personally identifiable information; expert panels have released various policy recommendations to conform practice to expectations of privacy.[100][101][102]

9.2 Critiques of big data execution

Big data has been called a "fad" in scientific research and its use was even made fun of as an absurd practice in a satirical example on "pig data".[77] Researcher danah boyd has raised concerns about the use of big data in science neglecting principles such as choosing a representative sample by being too concerned about actually handling the huge amounts of data.[103] This approach may lead to results bias in one way or another. Integration across heterogeneous data resources — some that might be considered "big data" and others not — presents formidable logistical as well as analytical challenges, but many researchers argue that such integrations are likely to represent the most promising new frontiers in science.[104] In the provocative article "Critical Questions for Big Data",[105] the authors title big data a part of mythology: "large data sets offer a higher form of intelligence and knowledge [...], with the aura of truth, objectivity, and accuracy". Users of big data are often "lost in the sheer volume of numbers", and "working with Big Data is still subjective, and what it quantifies does not necessarily have a closer claim on objective truth".[105] Recent developments in BI domain, such as pro-active reporting especially target improvements in usability of Big Data, through automated filtering of non-useful data and correlations.[106]

Big data analysis is often shallow compared to analysis of smaller data sets.[107] In many big data projects, there is no large data analysis happening, but the challenge is the extract, transform, load part of data preprocessing.[107]

Big data is a buzzword and a "vague term",[108] but at the same time an "obsession"[108] with entrepreneurs, consultants, scientists and the media. Big data showcases such as Google Flu Trends failed to deliver good predictions in recent years, overstating the flu outbreaks by a factor of two. Similarly, Academy awards and election predictions solely based on Twitter were more often off than on target. Big data often poses the same challenges as small data; and adding more data does not solve problems of bias, but may emphasize other problems. In particular data sources such as Twitter are not representative of the overall population, and results drawn from such sources may then lead to wrong conclusions. Google Translate - which is based on big data statistical analysis of text - does a remarkably good job at translating web pages, but for specialized domains the results may be badly off. On the other hand, big data may also introduce new problems, such as the multiple comparisons problem: simultaneously testing a large set of hypotheses is likely to produce many false results that mistakenly appear to be significant. Ioannidis argued that "most published research findings are false" [109] due to essentially the same effect: when many scientific teams and researchers each perform many experiments (i.e. process a big amount of scientific data; although not with big data technology), the likelihood of a "significant" result being actually false grows fast - even more so, when only positive results are published.

10 See also

- Apache Accumulo
- Apache Hadoop
- Big Data to Knowledge
- Big structure
- Data Defined Storage
- Cloudera
- HPCC Systems
- Internet of Things
- MapReduce
- Hortonworks
- Nonlinear system identification
- Operations research
- Programming with Big Data in R (a series of R packages)
- Sqrrl
- Supercomputer
- Transreality gaming
- Tuple space
- Unstructured data

11 References

[1] "Data, data everywhere". *The Economist*. 25 February 2010. Retrieved 9 December 2012.

[2] "Community cleverness required". *Nature* **455** (7209): 1. 4 September 2008. doi:10.1038/455001a.

[3] "Sandia sees data management challenges spiral". *HPC Projects*. 4 August 2009.

[4] Reichman, O.J.; Jones, M.B.; Schildhauer, M.P. (2011). "Challenges and Opportunities of Open Data in Ecology". *Science* **331** (6018): 703–5. doi:10.1126/science.1197962. PMID 21311007.

[5] "Data Crush by Christopher Surdak". Retrieved 14 February 2014.

[6] Hellerstein, Joe (9 November 2008). "Parallel Programming in the Age of Big Data". *Gigaom Blog*.

[7] Segaran, Toby; Hammerbacher, Jeff (2009). *Beautiful Data: The Stories Behind Elegant Data Solutions*. O'Reilly Media. p. 257. ISBN 978-0-596-15711-1.

[8] Hilbert & López 2011

[9] "IBM What is big data? — Bringing big data to the enterprise". www.ibm.com. Retrieved 2013-08-26.

[10] Oracle and FSN, "Mastering Big Data: CFO Strategies to Transform Insight into Opportunity", December 2012

[11] Jacobs, A. (6 July 2009). "The Pathologies of Big Data". *ACMQueue*.

[12] Magoulas, Roger; Lorica, Ben (February 2009). "Introduction to Big Data". *Release 2.0* (Sebastopol CA: O'Reilly Media) (11).

[13] Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big Data': Big gaps of knowledge in the field of Internet. *International Journal of Internet Science, 7*, 1-5. http://www.ijis.net/ijis7_1/ijis7_1_editorial.html

[14] Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner. Retrieved 6 February 2001.

[15] Beyer, Mark. "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". Gartner. Archived from the original on 10 July 2011. Retrieved 13 July 2011.

[16] Laney, Douglas. "The Importance of 'Big Data': A Definition". Gartner. Retrieved 21 June 2012.

[17] "What is Big Data?". Villanova University.

[18] http://www.bigdataparis.com/presentation/mercredi/PDelort.pdf?PHPSESSID=tv7k70pcr3egpi2r6fi3qbjtj6#page=4

[19] Billings S.A. "Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains". Wiley, 2013

[20] Delort P., Big data Paris 2013 http://www.andsi.fr/tag/dsi-big-data/

[21] Delort P., Big Data car Low-Density Data ? La faible densité en information comme facteur discriminant http://lecercle.lesechos.fr/entrepreneur/tendances-innovation/221169222/big-data-low-density-data-faible-densite-information-com

[22] "LHC Brochure, English version. A presentation of the largest and the most powerful particle accelerator in the world, the Large Hadron Collider (LHC), which started up in 2008. Its role, characteristics, technologies, etc. are explained for the general public.". *CERN-Brochure-2010-006-Eng. LHC Brochure, English version*. CERN. Retrieved 20 January 2013.

[23] "LHC Guide, English version. A collection of facts and figures about the Large Hadron Collider (LHC) in the form of questions and answers.". *CERN-Brochure-2008-001-Eng. LHC Guide, English version*. CERN. Retrieved 20 January 2013.

[24] Brumfiel, Geoff (19 January 2011). "High-energy physics: Down the petabyte highway". *Nature* **469**. pp. 282–83. doi:10.1038/469282a.

[25] http://www.zurich.ibm.com/pdf/astron/CeBIT%202013%20Background%20DOME.pdf

[26] http://arstechnica.com/science/2012/04/future-telescope-array-drives-development-of-exabyte-processing/

[27] Delort P., OECD ICCP Technology Foresight Forum, 2012. http://www.oecd.org/sti/ieconomy/Session_3_Delort.pdf#page=6

[28] Webster, Phil. "Supercomputing the Climate: NASA's Big Data Mission". *CSC World*. Computer Sciences Corporation. Retrieved 2013-01-18.

[29] Kalil, Tom. "Big Data is a Big Deal". White House. Retrieved 26 September 2012.

[30] Executive Office of the President (March 2012). "Big Data Across the Federal Government". White House. Retrieved 26 September 2012.

[31] Lampitt, Andrew. "The real story of how big data analytics helped Obama win". *Infoworld*. Retrieved 31 May 2014.

[32] Hoover, J. Nicholas. "Government's 10 Most Powerful Supercomputers". *Information Week*. UBM. Retrieved 26 September 2012.

[33] Bamford, James (15 March 2012). "The NSA Is Building the Country's Biggest Spy Center (Watch What You Say)". *Wired Magazine*. Retrieved 2013-03-18.

[34] "Groundbreaking Ceremony Held for $1.2 Billion Utah Data Center". National Security Agency Central Security Service. Retrieved 2013-03-18.

[35] Hill, Kashmir. "TBlueprints Of NSA's Ridiculously Expensive Data Center In Utah Suggest It Holds Less Info Than Thought". *Forbes*. Retrieved 2013-10-31.

[36] Layton, Julia. "Amazon Technology". Money.howstuffworks.com. Retrieved 2013-03-05.

[37] "Scaling Facebook to 500 Million Users and Beyond". Facebook.com. Retrieved 2013-07-21.

[38] "FICO® Falcon® Fraud Manager". Fico.com. Retrieved 2013-07-21.

[39] "eBay Study: How to Build Trust and Improve the Shopping Experience". Knowwpcarey.com. 2012-05-08. Retrieved 2013-03-05.

[40] Leading Priorities for Big Data for Business and IT. eMarketer. October 2013. Retrieved January 2014.

[41] Wingfield, Nick (2013-03-12). "Predicting Commutes More Accurately for Would-Be Home Buyers - NYTimes.com". Bits.blogs.nytimes.com. Retrieved 2013-07-21.

[42] UN GLobal Pulse (2012). Big Data for Development: Opportunities and Challenges (White p. by Letouzé, E.). New York: United Nations. Retrieved from http://www.unglobalpulse.org/projects/BigDataforDevelopment

[43] WEF (World Economic Forum), & Vital Wave Consulting. (2012). Big Data, Big Impact: New Possibilities for International Development. World Economic Forum. Retrieved 24 August 2012, from http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development

[44] "Big Data for Development: From Information- to Knowledge Societies", Martin Hilbert (2013), SSRN Scholarly Paper No. ID 2205145). Rochester, NY: Social Science Research Network; http://papers.ssrn.com/abstract=2205145

[45] "Elena Kvochko, Four Ways To talk About Big Data (Information Communication Technologies for Development Series)". worldbank.org. Retrieved 2012-05-30.

[46] "What Is the Content of the World's Technologically Mediated Information and Communication Capacity: How Much Text, Image, Audio, and Video?", Martin Hilbert (2014), The Information Society; free access to the article through this link: martinhilbert.net/WhatsTheContent_Hilbert.pdf

[47] Rajpurohit, Anmol (2014-07-11). "Interview: Amy Gershkoff, Director of Customer Analytics & Insights, eBay on How to Design Custom In-House BI Tools". *KDnuggets*. Retrieved 2014-07-14. "Dr. Amy Gershkoff: "Generally, I find that off-the-shelf business intelligence tools do not meet the needs of clients who want to derive custom insights from their data. Therefore, for medium-to-large organizations with access to strong technical talent, I usually recommend building custom, in-house solutions.""

[48] "LexisNexis To Buy Seisint For $775 Million". Washington Post. Retrieved 15 July 2004.

[49] "LexisNexis Parent Set to Buy ChoicePoint". Washington Post. Retrieved 22 February 2008.

[50] "Quantcast Opens Exabyte-Ready File System". www.datanami.com. Retrieved 1 October 2012.

[51] Bertolucci, Jeff "Hadoop: From Experiment To Leading Big Data Platform", "Information Week", 2013. Retrieved on 14 November 2013.

[52] Webster, John. "MapReduce: Simplified Data Processing on Large Clusters", "Search Storage", 2004. Retrieved on 25 March 2013.

[53] "Big Data Solution Offering". MIKE2.0. Retrieved 8 Dec 2013.

[54] "Big Data Definition". MIKE2.0. Retrieved 9 March 2013.

[55] Boja, C; Pocovnicu, A; Bătăgan, L. (2012). "Distributed Parallel Architecture for Big Data". *Informatica Economica* **16** (2): 116–127.

[56] Manyika, James; Chui, Michael; Bughin, Jaques; Brown, Brad; Dobbs, Richard; Roxburgh, Charles; Byers, Angela Hung (May 2011). *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

[57] "Future Directions in Tensor-Based Computation and Modeling". May 2009.

[58] Lu, Haiping; Plataniotis, K.N.; Venetsanopoulos, A.N. (2011). "A Survey of Multilinear Subspace Learning for Tensor Data". *Pattern Recognition* **44** (7): 1540–1551. doi:10.1016/j.patcog.2011.01.004.

[59] Monash, Curt (30 April 2009). "eBay's two enormous data warehouses".
Monash, Curt (6 October 2010). "eBay followup — Greenplum out, Teradata > 10 petabytes, Hadoop has some value, and more".

[60] "Resources on how Topological Data Analysis is used to analyze big data". Ayasdi.

[61] CNET News (1 April 2011). "Storage area networks need not apply".

[62] "How New Analytic Systems will Impact Storage". September 2011.

[63] http://asee-ne.org/proceedings/2014/Student%20Papers/210.pdf March 2014.

[64] "Obama Administration Unveils "Big Data" Initiative: Announces $200 Million In New R&D Investments". The White House.

[65] "AMPLab at the University of California, Berkeley". Amplab.cs.berkeley.edu. Retrieved 2013-03-05.

[66] "NSF Leads Federal Efforts In Big Data". National Science Foundation (NSF). 29 March 2012.

[67] Timothy Hunter; Teodor Moldovan; Matei Zaharia; Justin Ma; Michael Franklin; Pieter Abbeel; Alexandre Bayen (October 2011). "Scaling the Mobile Millennium System in the Cloud".

[68] David Patterson (5 December 2011). "Computer Scientists May Have What It Takes to Help Cure Cancer". The New York Times.

[69] "Secretary Chu Announces New Institute to Help Scientists Improve Massive Data Set Research on DOE Supercomputers". "energy.gov".

[70] "Governor Patrick announces new initiative to strengthen Massachusetts' position as a World leader in Big Data". Commonwealth of Massachusetts.

[71] "Big Data @ CSAIL". Bigdata.csail.mit.edu. 2013-02-22. Retrieved 2013-03-05.

[72] "Big Data Public Private Forum". Cordis.europa.eu. 2012-09-01. Retrieved 2013-03-05.

[73] "Alan Turing Institute to be set up to research big data". BBC News. 19 March 2014. Retrieved 2014-03-19.

[74] "Inspiration day at University of Waterloo, Stratford Campus". http://www.betakit.com/. Retrieved 2014-02-28.

[75] "Center for Intelligent Maintenance Systems (IMS Center)".

[76] Lee, Jay; Lapira, Edzel; Bagheri, Behrad; Kao, Hung-An (2013). "Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment". *Manufacturing Letters* **1** (1). doi:10.1016/j.mfglet.2013.09.005.

[77] Reips, Ulf-Dietrich; Matzat, Uwe (2014). "Mining "Big Data" using Big Data Services". *International Journal of Internet Science* **1** (1): 1–8.

[78] Preis, Tobias; Moat,, Helen Susannah; Stanley, H. Eugene; Bishop, Steven R. (2012). "Quantifying the Advantage of Looking Forward". *Scientific Reports* **2**: 350. doi:10.1038/srep00350. PMC 3320057. PMID 22482034.

[79] Marks, Paul (5 April 2012). "Online searches for future linked to economic success". *New Scientist*. Retrieved 9 April 2012.

[80] Johnston, Casey (6 April 2012). "Google Trends reveals clues about the mentality of richer nations". *Ars Technica*. Retrieved 9 April 2012.

[81] Tobias Preis (2012-05-24). "Supplementary Information: The Future Orientation Index is available for download". Retrieved 2012-05-24.

[82] Philip Ball (26 April 2013). "Counting Google searches predicts market movements". *Nature*. Retrieved 9 August 2013.

[83] Tobias Preis, Helen Susannah Moat and H. Eugene Stanley (2013). "Quantifying Trading Behavior in Financial Markets Using Google Trends". *Scientific Reports* **3**: 1684. doi:10.1038/srep01684.

[84] Nick Bilton (26 April 2013). "Google Search Terms Can Predict Stock Market, Study Finds". *New York Times*. Retrieved 9 August 2013.

[85] Christopher Matthews (26 April 2013). "Trouble With Your Investment Portfolio? Google It!". *TIME Magazine*. Retrieved 9 August 2013.

[86] Philip Ball (26 April 2013). "Counting Google searches predicts market movements". *Nature*. Retrieved 9 August 2013.

[87] Bernhard Warner (25 April 2013). "'Big Data' Researchers Turn to Google to Beat the Markets". *Bloomberg Businessweek*. Retrieved 9 August 2013.

[88] Hamish McRae (28 April 2013). "Hamish McRae: Need a valuable handle on investor sentiment? Google it". *The Independent* (London). Retrieved 9 August 2013.

[89] Richard Waters (25 April 2013). "Google search proves to be new word in stock market prediction". *Financial Times*. Retrieved 9 August 2013.

[90] David Leinweber (26 April 2013). "Big Data Gets Bigger: Now Google Trends Can Predict The Market". *Forbes*. Retrieved 9 August 2013.

[91] Jason Palmer (25 April 2013). "Google searches predict market moves". *BBC*. Retrieved 9 August 2013.

[92] "Manufacturing: Big Data Benefits and Challenges". *TCS Big Data Study*. Mumbai, India: Tata Consultancy Services Limited. Retrieved 2014-06-03.

[93] Lee, Jay; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L (Jan 2013). "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications". *Mechanical Systems and Signal Processing* **42** (1).

[94] Graham M. (9 March 2012). "Big data and the end of theory?". *The Guardian* (London).

[95] "Good Data Won't Guarantee Good Decisions. Harvard Business Review". *Shah, Shvetank; Horne, Andrew; Capellá, Jaime;*. HBR.org. Retrieved 8 September 2012.

[96] Anderson, C. (2008, 23 June). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired Magazine, (Science: Discoveries). http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

[97] Rauch, J. (2002). Seeing Around Corners. The Atlantic, (April), 35–48. http://www.theatlantic.com/magazine/archive/2002/04/seeing-around-corners/302471/

[98] Epstein, J. M., & Axtell, R. L. (1996). Growing Artificial Societies: Social Science from the Bottom Up. A Bradford Book.

[99] Delort P., Big data in Biosciences, Big Data Paris, 2012 http://www.bigdataparis.com/documents/Pierre-Delort-INSERM.pdf#page=5

[100] Ohm, Paul. "Don't Build a Database of Ruin". Harvard Business Review.

[101] Darwin Bond-Graham, *Iron Cagebook - The Logical End of Facebook's Patents,* Counterpunch.org, 2013.12.03

[102] Darwin Bond-Graham, *Inside the Tech industry's Startup Conference,* Counterpunch.org, 2013.09.11

[103] danah boyd (2010-04-29). "Privacy and Publicity in the Context of Big Data". *WWW 2010 conference*. Retrieved 2011-04-18.

[104] Jones, MB; Schildhauer, MP; Reichman, OJ; Bowers, S (2006). "The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere" (PDF). *Annual Review of Ecology, Evolution, and Systematics* **37** (1): 519–544. doi:10.1146/annurev.ecolsys.37.091305.110031.

[105] Boyd, D.; Crawford, K. (2012). "Critical Questions for Big Data". *Information, Communication & Society* **15** (5): 662. doi:10.1080/1369118X.2012.678878.

[106] Failure to Launch: From Big Data to Big Decisions, Forte Wares.

[107] Gregory Piatetsky (2014-08-12). "Interview: Michael Berthold, KNIME Founder, on Research, Creativity, Big Data, and Privacy, Part 2". KDnuggets. Retrieved 2014-08-13.

[108] Harford, Tim (2014-03-28). "Big data: are we making a big mistake?". *Financial Times*. Financial Times. Retrieved 2014-04-07.

[109] Ioannidis, J. P. A. (2005). "Why Most Published Research Findings Are False". *PLoS Medicine* **2** (8): e124. doi:10.1371/journal.pmed.0020124. PMC 1182327. PMID 16060722.

12 Further reading

- Big Data Computing and Clouds: Challenges, Solutions, and Future Directions. Marcos D. Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A. S. Netto, Rajkumar Buyya. Technical Report CLOUDS-TR-2013-1, Cloud Computing and Distributed Systems Laboratory, The University of Melbourne, 17 Dec. 2013.
- Encrypted search & cluster formation in Big Data. Gautam Siwach, Dr. A. Esmailpour. American Society for Engineering Education, Conference at the University of Bridgeport, Bridgeport, Connecticut 3–5 April 2014.
- "Big Data for Good". ODBMS.org. 5 June 2012. Retrieved 2013-11-12.
- Hilbert, Martin; López, Priscila (2011). "The World's Technological Capacity to Store, Communicate, and Compute Information". *Science* **332** (6025): 60–65. doi:10.1126/science.1200970. PMID 21310967.
- "The Rise of Industrial Big Data". GE Intelligent Platforms. Retrieved 2013-11-12.
- History of Big Data Timeline. A visual history of Big Data with links to supporting articles.

13 External links

- Media related to Big data at Wikimedia Commons
- The dictionary definition of big data at Wiktionary

14 Text and image sources, contributors, and licenses

14.1 Text

- **Big data** *Source:* http://en.wikipedia.org/wiki/Big_data?oldid=631110945 *Contributors:* William Avery, Heron, Kku, Samw, Andrewman327, Ryuch, דוד, Topbanana, Paul W, F3meyer, Sunray, Giftlite, Langec, Utcursch, Beland, Jeremykemp, David@scatter.com, Discospinster, Rich Farmbrough, Kdammers, ArnoldReinhold, Viriditas, Lenov, Gary, Pinar, Tobych, Miranche, Broeni, Tomlzz1, Axeman89, Woohookitty, Pol098, Qwertyus, Rjwilmsi, ElKevbo, Jehochman, Lumin, Tedder, DVdm, SteveLoughran, Aeusoes1, Daniel Mietchen, Dimensionsix, Katieh5584, Henryyan, McGeddon, Od Mishehu, Gilliam, Ohnoitsjamie, Chris the speller, RDBrown, Pegua, Madman2001, Krexer, Kuru, Almaz, Dl2000, The Letter J, Chris55, Yragha, Jac16888, Marc W. Abel, Cydebot, Quibik, DumbBOT, Malleus Fatuorum, Cowb0y, Lmusher, Josephmarty, Kforeman1, Rmyeid, OhanaUnited, Relyk, Magioladitis, Nyq, Tedickey, Steven Walling, Thevoid00, Casieg, McSly, Lamp90, Pchackal, Mgualtieri, VolkovBot, JohnBlackburne, Philip Trueman, Ottb19, Billinghurst, Grinq, Scottywong, Dawn Bard, Yintan, Jazzwang, Jojikiba, CutOffTies, Mkbergman, Melcombe, Siskus, PabloStraub, Dilaila, Martarius, Faalagorn, Apptrain, Morrisjd1, Grantbow, Mild Bill Hiccup, Ottawahitech, Cirt, Auntof6, Lbertolotti, Gnome de plume, Resoru, Pablomendes, SchreiberBike, MPH007, Mymallandnews, XLinkBot, Ost316, Benboy00, MystBot, P.r.newman, Addbot, Mortense, Drevicko, Thomas888b, AndrewHZ, Tothwolf, Ronhjones, Moosehadley, MrOllie, Download, Jarble, Arbitrarily0, Luckas-bot, Yobot, Fraggle81, Manivannan pk, Jean.julius, AnomieBOT, Jim1138, Babrodtk, Materialscientist, Citation bot, Xqbot, Marko Grobelnik, Bgold12, Anna Frodesiak, Tomwsulcer, Srich32977, Omnipaedista, Smallman12q, Joaquin008, FrescoBot, Jonathanchaitow, I42, AtmosNews, B3t, I dream of horses, HRoestBot, Jonesey95, Jandalhandler, Mengxr, Ethansdad, Yzerman123, Msalganik, בן גרשון, Sideways713, Stuartzs, Jfmantis, Mean as custard, RjwilmsiBot, Ripchip Bot, Mm479flarok, Winchetan, Petermcelwee, DASHBot, EmausBot, Oliverlyc, Timtempleton, Dewritech, Peaceray, Cmlloyd1969, K6ka, HiW-Bot, Richard asr, ZéroBot, Checkingfax, BobGourley, Xtzou, Chire, Kilopi, Laurawilber, Rcsprinter123, Rick jens, Palosirkka, MainFrame, ChuispastonBot, Sean Quixote, Axelode, Mhiji, Helpsome, ClueBot NG, Behrad3d, Danielg922, Pramanicks, Jj1236, Widr, WikiMSL, Lawsonstu, Fvillanustre, Helpful Pixie Bot, Lowercase sigmabot, BG19bot, And Adoil Descended, Seppemans123, Jantana, Northamerica1000, Asplanchna, MusikAnimal, AvocatoBot, Noelwclarke, Matt tubb, Jordanzhang, Bar David, InfoCmplx, Atlasowa, Fylbecatulous, Camberleybates, WH98, DigitalDev, Haroldpolo, Ryguyrg, Untioencolonia, Shirishnetke, Ampersandian, MarkTraceur, ChrisGualtieri, TheJJJunk, Khazar2, Vaibhav017, IjonTichyIjonTichy, Saturdayswiki, Seherrell, Mjvaugh2, ChazzI73, Davidogm, Mherradora, Jkofron4, Stevebillings, Indianbusiness, Toopathfind, Jeremy Kolb, Jamesx12345, OnTheNet21, BrighterTomorrow, Jacoblarsen net, Epicgenius, DavidKSchneider, Socratesplato9, Anirudhrata, Parasdoshiblog, Edwinboothnyc, JuanCarlosBrandt, Helenellis, MMeTrew, Warrenpd86, AuthorAnil, ViaJFK, Gary Simon, Bsc, FCA, FBCS, CITP, Mcioffi, Joe204, Caraconan, Evaluatorgroup, Hessmike, TJLaher123, Chengying10, IndustrialAutomationGuru, Dabramsdt, Prussonyc, Abhishek1605, Dilaila123, Willymomo, Rzicari, Mingminchi, BigDataGuru1, Sysœp, Azra2013, Dudewhereismybike, Shahbazali101, Miakeay, Accountdp, Morganmissen, JeanneHolm, Yourconnotation, JenniferAndy, Arcamacho, Bigdatavomit, Monkbot, Wikientg, Scottishweather, Textractor, Analytics ireland, Lspin01l, ForumOxford Online, Mansoor-siamak, Sightestrp, Jwdang4, Amortias, Wikiauthor22, Femiolajiga, Tttcraig, Lepro2, Mythfinder, DexterToo, Mr P. Kopee, Pablollopis, SVtechie, Smaske, Greystoke1337, Prateekkeshari, Hmrv83, KaraHayes, Iqmc, Lalith269 and Anonymous: 273

14.2 Images

- **File:2013-09-11_Bus_wrapped_with_SAP_Big_Data_parked_outside_IDF13_(9730051783).jpg** *Source:* http://upload.wikimedia.org/wikipedia/commons/8/8d/2013-09-11_Bus_wrapped_with_SAP_Big_Data_parked_outside_IDF13_%289730051783%29.jpg *License:* CC-BY-SA-2.0 *Contributors:* Bus wrapped with SAP Big Data parked outside IDF13 *Original artist:* Intel Free Press
- **File:Ambox_globe_content.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/b/bd/Ambox_globe_content.svg *License:* Public domain *Contributors:* Own work, using File:Information icon3.svg and File:Earth clip art.svg *Original artist:* penubag
- **File:Big_data_cartoon_t_gregorius.jpg** *Source:* http://upload.wikimedia.org/wikipedia/commons/b/b3/Big_data_cartoon_t_gregorius.jpg *License:* CC-BY-2.0 *Contributors:* Cartoon: Big Data *Original artist:* Thierry Gregorius
- **File:Commons-logo.svg** *Source:* http://upload.wikimedia.org/wikipedia/en/4/4a/Commons-logo.svg *License:* ? *Contributors:* ? *Original artist:* ?
- **File:Computer-aj_aj_ashton_01.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/c/c1/Computer-aj_aj_ashton_01.svg *License:* ? *Contributors:* ? *Original artist:* ?
- **File:Folder_Hexagonal_Icon.svg** *Source:* http://upload.wikimedia.org/wikipedia/en/4/48/Folder_Hexagonal_Icon.svg *License:* ? *Contributors:* ? *Original artist:* ?
- **File:Viegas-UserActivityonWikipedia.gif** *Source:* http://upload.wikimedia.org/wikipedia/commons/6/69/Viegas-UserActivityonWikipedia.gif *License:* CC-BY-2.0 *Contributors:* User activity on Wikipedia *Original artist:* Fernanda B. Viégas

14.3 Content license

- Creative Commons Attribution-Share Alike 3.0