

# BIG DATA GLOSSARY

## ADVANCED RESEARCH COMPUTING

High performance computing and storage needs that are too complex to be handled by a standard desktop workstation, specifically in support of research. Includes cloud environments, massive-scale infrastructure and large computational power.<sup>1,2</sup>

## ALGORITHMS

Formal specifications used in software to process and analyze datasets. Typically made up of a series of calculation steps, an algorithm will describe how to process data with the goal of solving a specific problem.<sup>3,4</sup>

## ANALYTICS

The process of using statistics, algorithms and software to transform data into useful information and to draw conclusions.<sup>3,4</sup>

## BEHAVIOUR(AL) ANALYTICS

A type of business analytics that examines consumer or user behaviour data to understand how and why individuals behave the way they do, with the goal of making more accurate predictions about future behaviours.<sup>5,6</sup>

## BIG DATA

The term is often used to refer to the massive amounts of structured and unstructured data generated around the world that is too large, complex or varied for traditional processing software. There are numerous processes and issues related to big data, including its potential to be mined or analyzed for valuable information.<sup>7,8</sup>

## CLOUD COMPUTING

Data and computing resources available and accessed on-demand over the internet instead of using the resources of a local computer or device.<sup>9,10</sup>

## DATA MINING

An analytical process where large datasets are explored or “mined” in search of meaningful patterns, relationships or insights. The process can include machine learning, statistics or artificial intelligence.<sup>3,4</sup>

## DATA VISUALIZATION

The use of visual representation of data (including infographics, charts, maps) with goals that include sensemaking, effective communication of information, and improved pattern detection.<sup>11, 12</sup>

## DATA AGGREGATION

The process of compiling information from different data sources in preparation for data processing or statistical analysis.<sup>4</sup>

## DATABASE

A large, organized collection of data that allows easy access, management, updating and analysis of information.<sup>3</sup>

## DESCRIPTIVE ANALYTICS

An initial stage of data processing that involves creating a summary of historical data with the goal of producing useful information or, answering the question, “What happened?”<sup>13,14</sup>

## DISTRIBUTED FILE SYSTEM

A mechanism that stores files on servers and allows clients, with permission, to store and process files as if they were stored on their own computer.<sup>15,16</sup>

## IOT (INTERNET OF THINGS)

A system of “connected” or “smart” objects, including smartphones, wearables and smart-appliances, that collect and exchange information without necessarily requiring human interaction. The IoT is a huge generator of data.<sup>17,18</sup>

## NETWORK

A system where multiple computing devices are connected to each other and exchange information and resources through a data link. For example, the Internet.<sup>19,20</sup>

## NETWORK ANALYSIS

The mapping and measuring of relationships and flows between people, groups, organizations, computers, and other connected information/knowledge entities.<sup>21</sup>

## PREDICTIVE ANALYTICS

Related to descriptive and prescriptive analytics, a process for analyzing current data to make predictions about future events. Draws on techniques from data mining, modeling, machine learning and statistics.<sup>3</sup>

## PRESCRIPTIVE ANALYTICS

Related to descriptive and predictive analytics, a form of business analytics with the goal of using data to determine the best course of action for a specific scenario.<sup>3</sup>

## SECURE COMPUTING

Related to “computer security” or “IT security,” this term is often used when discussing privacy and protection of data, devices and computing systems from harm, such as theft, damage or malpractice. It includes protection of hardware, software, and system information, as well as disruption of the services a system may provide.<sup>22,4</sup>

## STRUCTURED DATA

Any data that has been organized into structured fields, such as a database or spreadsheet, so that it can be easily processed or analyzed. Examples include dates, words, numbers.<sup>24,25</sup>

## UNSTRUCTURED DATA

Information that either does not have a pre-defined data model or is not organized in a pre-defined manner. It is thus not stored in a database in structured fields. Examples include text, images, audio, video.<sup>26,27</sup>

# REFERENCES

- [1] "Westgrid: What We Do." [Online]. Available: [https://www.westgrid.ca/about\\_westgrid/what\\_we\\_do](https://www.westgrid.ca/about_westgrid/what_we_do).
- [2] "Compute Canada. Update January 11, 2017." [Online]. Available: <https://www.computeCanada.ca/wp-content/uploads/2015/02/Update-January-11-2017.pdf>
- [3] "A beginner's guide to big data terminology." [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology>.
- [4] "Mini-glossary: Big data terms you should know." [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know>.
- [5] "Behavioral analytics, definition by Techopedia." [Online]. Available: <https://www.techopedia.com/definition/30308/behavioral-analytics>.
- [6] "Detecting advanced threats with user behavior analytics." [Online]. Available: <http://www.networkworld.com/article/2904356/security0/detecting-advanced-threats-with-user-behavior-analytics.html>.
- [7] "Big data definition by Tech Target." [Online]. Available: <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>.
- [8] "What Is Big Data?" [Online]. Available: <https://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/#4fb6509b5c85>.
- [9] "Who Coined 'Cloud Computing'?" [Online]. Available: <https://www.technologyreview.com/s/425970/who-coined-cloud-computing>.
- [10] "Cloud computing: a simple introduction." [Online]. Available: <http://www.explainthatstuff.com/cloud-computing-introduction.html>.
- [11] "Data visualization definition by Tech Target." [Online]. Available: <http://searchbusinessanalytics.techtarget.com/definition/data-visualization>.
- [12] "Data visualization for human perception." [Online]. Available: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception>.
- [13] "Descriptive analytics definition by Tech Target." [Online]. Available: <http://whatis.techtarget.com/definition/descriptive-analytics>.
- [14] "Four Types of Big Data Analytics and Examples of Their Use." [Online]. Available: <http://www.ingrammicroadvisor.com/data-center/four-types-of-big-data-analytics-and-examples-of-their-use>.
- [15] "Distributed file system definition by Tech Target." [Online]. Available: <http://searchwindowserver.techtarget.com/definition/distributed-file-system-DFS>.
- [16] "Distributed file system definition by Techopedia." [Online]. Available: <https://www.techopedia.com/definition/1825/distributed-file-system-dfs>.
- [17] "A Simple Explanation Of 'The Internet Of Things.'" [Online]. Available: <https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#34e4f89b1d09>.
- [18] "Internet of things definition by Tech Target." [Online]. Available: <http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT>.
- [19] "Computer network definition by Techopedia." [Online]. Available: <https://www.techopedia.com/definition/25597/computer-network>.
- [20] "What is computer networking?" [Online]. Available: <https://www.lifewire.com/what-is-computer-networking-816249>.
- [21] "Social Network Analysis: An Introduction by Orgnet,LLC." [Online]. Available: <http://www.orgnet.com/sna.html>.
- [22] "The myth of secure computing." [Online]. Available: <https://hbr.org/2003/06/the-myth-of-secure-computing>.
- [23] "Computer security definition by PC Magazine." [Online]. Available: <http://www.pcmag.com/encyclopedia/term/40169/computer-security>.
- [24] "Introduction to Structured Data, by Google." [Online]. Available: <https://developers.google.com/search/docs/guides/intro-structured-data>.
- [25] "Structured data definition by Webopedia." [Online]. Available: [http://www.webopedia.com/TERM/S/structured\\_data.html](http://www.webopedia.com/TERM/S/structured_data.html).
- [26] "Unstructured data definition by Techopedia." [Online]. Available: <https://www.techopedia.com/definition/13865/unstructured-data>.
- [27] "Solving the Unstructured Data Challenge." [Online]. Available: <http://www.cio.com/article/2941015/big-data/solving-the-unstructured-data-challenge.html>.