[**Table of Contents**](#Table-of-Contents)

[**Term Categories**](#Categories)

# Glossary

### ADVANCED RESEARCH COMPUTING

Revolutionizing how we conduct research, advanced research computing combines high performance computing and large [storage](#Storage-and-Physical-Data-Sources) that are too complex to be handled by a standard desktop workstation, utilizing [cloud environments](#CLOUD-COMPUTING), massive-scale infrastructure and large computational power. [[1]](#endnote-2),[[2]](#endnote-3)

### ALGORITHM

A procedure to address a specific problem by a sequence of instructions to process input [data](#DATA) resulting in some output. For instance, data analysis algorithms, such as [classification](#CLASSIFICATION) or [clustering](#CLUSTERING) provide labels and predictions for further exploration and analysis. [[3]](#endnote-4),[[4]](#endnote-5)

### ANALYTICS

The process of using statistics [algorithms](#ALGORITHM) and software to [transform data](#DATA-REPRESENTATION-AND-TRANSFORMATION) into useful [information](#INFORMATION) and to draw conclusions towards effective decision making.[[5]](#endnote-6),[[6]](#endnote-7)

### BEHAVIOUR(AL) ANALYTICS

A type of business [analytics](#ANALYTICS) that examines consumer or user behavioural data to understand how and why individuals act the way they do, with the goal of making more accurate predictions about future behaviours.[[7]](#endnote-8),[[8]](#endnote-9)

### BIG DATA

Refers to the massive amounts of [data](#DATA) generated around the world that is too large, complex or varied for traditional processing software. Its potential to be analyzed for valuable information is enabled by technology such as [advanced research computing](#ADVANCED-RESEARCH-COMPUTING). [[9]](#endnote-10),[[10]](#endnote-11)

### CAUSAL INFERENCE

Determine whether observations made in one variable are the reason for an effect observed in another variable, possibly occurring at a later time. [Correlation association](#CORRELATION-MINING) and [ability to predict](#PREDICTIVE-ANALYTICS) often result from causal relationships, but do not necessarily indicate the cause.

### CLASSIFICATION

Classification in [machine learning](#MACHINE-LEARNING) can identify the class that previously unseen, unlabeled examples belong to, for instance by naming an object shown in a photograph. The [construction](#MODEL-FITTING) of a classifier requires training [data](#DATA) that provides labels for each example data point (or photo). The requirement of given labels for training is the main difference between classification and [clustering](#CLUSTERING).

### CLOUD COMPUTING

Access to data, applications, and other computing resources made available to many users on-demand over the [Internet](#NETWORK). User responsibilities to own local [storage](#DISTRIBUTED-FILE-SYSTEM) or maintain system integrity are delegated to remote third-party providers running data centers with advantages including [scalable computing power](#ADVANCED-RESEARCH-COMPUTING) and reliability.[[11]](#endnote-12),[[12]](#endnote-13)

### CLUSTER

A compute cluster is a [network](#NETWORK) of computers (or compute nodes) that work on tasks together in parallel. The concept is different from [clustering](#CLUSTERING).

### CLUSTERING

Clustering is a task in [machine learning](#MACHINE-LEARNING) that groups examples or points to optimize some measure of distance among or similarity within groups of points, or so-called clusters, to uncover information about hidden structure in the data. As opposed to [classification](#CLASSIFICATION) clustering does not require labelled training data and is a type of unsupervised learning.

### COMPUTING WITH DATA

Learning from data using computing tools and programming languages, such as Python or R with large ecosystems of libraries providing a convenient level of abstraction and statistical methods that can be organized into data processing pipelines.[[13]](#endnote-14)

### CORRELATION MINING

Correlation is a statistical measure that describes the association between two or more quantitative variables. It can help to predict one quantity from another. Although correlation can indicate the presence of [causal relationship](#CAUSAL-INFERENCE) it does, however, not sufficiently imply causation. [[14]](#endnote-15)

### DATA

Collection of examples, observations, measurements, facts, points, or other items of information that can be represented in [structured](#STRUCTURED-DATA) or [unstructured](#UNSTRUCTURED-DATA) form. Examples include: tables or [relational data](#TABULAR-SCHEMA); spatial and temporal data from [sensors](#IOT-INTERNET-OF-THINGS) or event logs; graphs, such as social networks or web-based data; and other complex data such as text or images.

### DATABASE

Organized collection of [data](#DATA) that allows easy access, management, updating and analysis of large amounts of data. Commonly used databases are MySQL, PostgreSQL, as well as various [NoSQL](#NO-SQL-DATABASE) options.[[15]](#endnote-16)

### DATA EXPLORATION AND PREPARATION

Exploratory data analysis (EDA) is a formative step in the [creation of models](#MODEL-FITTING). Views of the data are used to learn about patterns or relationships among variables. This includes data cleaning (careful removal of erroneous or unreliable data points) and manipulation for further analysis.

### DATA INTEGRATION

The process of combining information from different data sources in preparation for data processing. [[16]](#endnote-17)

### DATA MINING

An analytical process where large datasets are explored or “mined” in search of meaningful patterns, relationships or insights. The process can include statistics, [machine learning](#MACHINE-LEARNING) or other forms of artificial intelligence.[[17]](#endnote-18),[[18]](#endnote-19)

### DATA MODELING

Construction of a [computational representation](#ALGORITHM) via [machine learning](#MACHINE-LEARNING) that imitates the behaviour of some real-world process captured in [data](#DATA). Two main approaches are [generative modeling](#GENERATIVE-MODELING) and [predictive modeling](#PREDICTIVE-ANALYTICS) producing new possible data examples or determining hidden attributes of given examples, respectively.

### DATA REPRESENTATION AND TRANSFORMATION

Data representation refers to the form in which [data](#DATA) is stored, processed, and transmitted retaining the information content and [context within the original data](#METADATA) to the greatest degree possible. Choices of data representation can be influenced by constraints in hardware, software or other resources needed for processing and analysis. Data transformation converts data into different representations that may be more insightful, for instance converting [unstructured data](#UNSTRUCTURED-DATA), such as text, numbers, photo, or music, into [forms](#STRUCTURED-DATA) that can be analysed further. [[19]](#endnote-20)

### DATA SCIENCE

Combination of activities to extract actionable knowledge from [data](#DATA) involving programming skills, math and statistical analysis, and expertise in a specific sector or domain. Often performed in stages: 1.[Data Exploration and Preparation](#DATA-EXPLORATION-AND-PREPARATION) 2.[Data Representation and Transformation](#DATA-REPRESENTATION-AND-TRANSFORMATION) 3.[Computing with Data](#COMPUTING-WITH-DATA) 4.[Data Modeling](#DATA-MODELING) 5.[Data Visualization and Presentation](#VISUALIZATION). [[20]](#endnote-21),[[21]](#endnote-22)

### 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?”[[22]](#endnote-23),[[23]](#endnote-24)

### 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.[[24]](#endnote-25),[[25]](#endnote-26)

### FEATURE

[Data](#DATA) to represent a real-world object in abstract form to capture its key characteristics that are of relevance for a particular [machine learning](#MACHINE-LEARNING) problem or other types of [analysis](#ANALYTICS).

### FEATURE ENGINEERING

[Feature](#FEATURE) engineering is a human-driven process of developing [prediction models](#PREDICTIVE-ANALYTICS) and is an important part of [data modeling](#DATA-MODELING). It finds the most informative and independent features among different choices of features by studying their impact on [goodness](#GOODNESS-OF-FIT) of [model fit](#MODEL-FITTING).

### FLEXIBLE SCHEMA

Unlike [SQL](#SQL) databases, where a [tabular schema](#TABULAR-SCHEMA) is a set of tables, a flexible schema by default, does not require data structure to have the same [schema](STRUCTURED-DATA). That is, the documents in a single collection do not need to have the same set of fields and the data type for a field. They can be different from one another. MongoDB is an example of database management systems that supports flexible schema.[[26]](#endnote-27)

### GENERATIVE MODELING

[Fitting](#MODEL-FITTING) a [model](#DATA-MODELING) that can generate synthetic data beyond [given observations](#DATA). For example, a language model trained on a corpus of text can be used to suggest possible words to modify or continue a given phrase. [[27]](#endnote-28)

### GOODNESS OF FIT

Summarizes how well the values observed in the data agree with those values expected by the model.

### INFORMATION

Information is meaning encoded in [data](#DATA) that answers questions to better understand a concept by interpreting data within the context of its problem setting or domain. See also data analysis [algorithms](#ALGORITHM) and information [visualization](#VISUALIZATION). [[28]](#endnote-29)

### INSIGHT

Insight is actionable [information](#INFORMATION) gained by interpreting [data analysis](#ANALYTICS) results. It can help people to make more informed decisions.[[29]](#endnote-30)

### IOT (INTERNET OF THINGS)

A system of [connected](#NETWORK), “smart” objects, including smartphones, wearables and smart-appliances, that collect and exchange information without requiring human interaction. The IoT is a huge generator of data.[[30]](#endnote-31),[[31]](#endnote-32)

### KNOWLEDGE DISCOVERY

Knowledge discovery (KDD) aims to extract [insight](#INSIGHT) from data in [databases](#DATABASE). It involves a [number of steps](#DATA-SCIENCE) including the evaluation and possibly [interpretation](#VISUALIZATION) of [patterns](#DATA-MINING) to decide on what qualifies as knowledge.[[32]](#endnote-33)

### MACHINE LEARNING

Use of [algorithmic](#ALGORITHM) [models](#DATA-MODELING) that are able to perform a certain task without specific instructions, but instead are fitted via training to imitate input/output behaviour or other patterns present in example [data](#DATA). Numerous applications include [prediction](#PREDICTIVE-ANALYTICS) and other forms of data [analytics](#ANALYTICS).

### METADATA

Also called “data about data,” metadata acts as a [structured description](#STRUCTURED-DATA) and context of a data object such as a document, image, or a file to help organize, find and understand the data. [[33]](#endnote-34)

### MODEL FITTING

Model fitting [optimizes](#GOODNESS-OF-FIT) how well a [machine learning model](#MACHINE-LEARNING) generalizes beyond training data ensuring that its predictions also work well for previously unseen test data.

### 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.[[34]](#endnote-35),[[35]](#endnote-36)

### NETWORK ANALYSIS

The mapping and measuring of relationships and flows between people, groups, organizations, computers, and other connected entities. For example, identification of relevant web search results by ranking a page of matching content based on how many other pages refer to it. [[36]](#endnote-37)

### NO-SQL DATABASE

NoSQL databases offer more [flexible](#FLEXIBLE-SCHEMA) ways of data access than generic [SQL](#SQL) databases to [optimize other requirements](#ADVANCED-RESEARCH-COMPUTING), for instance, to give faster access to data organized contiguously in [columns](#TABULAR-SCHEMA), collections of documents, graph structures of items with connections, or key-value pairs.

### PREDICTIVE ANALYTICS

A process for analyzing current data to determine probable observations for future or previously unseen events or [data](#DATA) points, for example, forecasting of weather conditions or election outcomes based on polls. Related to [descriptive](#DESCRIPTIVE-ANALYTICS) and [prescriptive analytics](#PRESCRIPTIVE-ANALYTICS), it draws on techniques from [data mining](#DATA-MINING), modeling, [machine learning](#MACHINE-LEARNING) and statistics.[[37]](#endnote-38),[[38]](#endnote-39)

### PRESCRIPTIVE ANALYTICS

Related to descriptive and [predictive analytics](#PREDICTIVE-ANALYTICS), a form of [analytics](#ANALYTICS) with the goal of using data to determine the best course of action for a specific scenario.[[39]](#endnote-40)

### REGRESSION

Regression analysis is a form of [predictive](#PREDICTIVE-ANALYTICS) [modelling](#DATA-MODELING) for numerical or continuous variables, for example, estimating the value of a property based on similar properties in the neighbourhood. [Classification](#CLASSIFICATION) is a similar analysis technique, but instead predicts a class name or [label](#FEATURE).

### RELATIONSHIP MINING

Relationship mining examines associations between two or more variables in a dataset, for example, by [correlation mining](#CORRELATION-MINING) and [causal mining](#CAUSAL-INFERENCE).

### 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 hacking, damage or malpractice. It includes protection of hardware, software, and system information, as well as disruption of the services a system provides.[[40]](#endnote-41),[[41]](#endnote-42)

### SQL

SQL is a language that works with structured, [tabular data](#TABULAR-SCHEMA) allowing to query and construct such data. It forms a standard for many [database](#DATABASE) systems. [28]

### STRUCTURED DATA

[Data](#DATA) that is organized into clearly defined fields associated with variables or attributes, such as dates, words, or numbers that are recorded for each observation or item. Items are often represented as rows of [tabular spreadsheets](#TABULAR-SCHEMA) that can be stored in a [database](#DATABASE) for easy processing and analysis.[[42]](#endnote-43),[[43]](#endnote-44)

### TABULAR SCHEMA

Many interesting data sources in computing can be organized as tabular items arranged in a rectangular grid of horizontal rows and vertical columns.[[44]](#endnote-45) A cell is formed by the intersection of a column and row and can contain a value.[[45]](#endnote-46) Often all cells in a column correspond to one variable and have the same type, such as number, string, date etc. and a row corresponds to an item, data point, or observation.

### TEXT MINING

Text mining is a process of analyzing [text](#UNSTRUCTURED-DATA) to capture key concepts, themes, relationships and trends.[[46]](#endnote-47)

### UNSTRUCTURED DATA

Data that carries [information](#INFORMATION) without a pre-defined data model or that is not organized in a pre-defined manner. It is thus not stored in a [database](#DATABASE) in structured fields. Examples include [text](#TEXT-MINING), images, audio, or video.[[47]](#endnote-48),[[48]](#endnote-49)

### VISUALIZATION

The use of visual representation of data including plots, charts, maps, and infographics with goals that include sensemaking, effective communication of information, and improved pattern detection with a human-in-the-loop. For example, a treadmill screen at the gym showing a profile of calories burnt with respect to distance covered, or flight operation using a radar screen displaying sensor data in real-time. [[49]](#endnote-50),[[50]](#endnote-51)

### VISUALIZATION DASHBOARD

Multiple interactive graphical views of [data](#DATA) that would be too complex if presented in a [single visualization](#VISUALIZATION), for instance, to support complex narratives and [insights](#INSIGHT) by showing key performance indicators (KPIs) for monitoring of a business or other processes.

# Categories

The categories offer an alternative grouping that helps to find related terms.

### Data Inquiry Disciplines

[ADVANCED RESEARCH COMPUTING](#ADVANCED-RESEARCH-COMPUTING)  
[ANALYTICS](#ANALYTICS)  
[BIG DATA](#BIG-DATA)  
[CLOUD COMPUTING](#CLOUD-COMPUTING)  
[COMPUTING WITH DATA](#COMPUTING-WITH-DATA)  
[KNOWLEDGE DISCOVERY](#KNOWLEDGE-DISCOVERY) [DATA MINING](#DATA-MINING)  
[DATA SCIENCE](#DATA-SCIENCE)  
[MACHINE LEARNING](#MACHINE-LEARNING)  
[SECURE COMPUTING](#SECURE-COMPUTING)

### Data Analysis Techniques

[DATA MODELING](#DATA-MODELING)  
[GENERATIVE MODELING](#GENERATIVE-MODELING)  
[PREDICTIVE ANALYTICS](#PREDICTIVE-ANALYTICS)  
[REGRESSION](#REGRESSION)  
[FEATURE ENGINEERING](#FEATURE-ENGINEERING)  
[CLASSIFICATION](#CLASSIFICATION)  
[CLUSTERING](#CLUSTERING)  
[DESCRIPTIVE ANALYTICS](#DESCRIPTIVE-ANALYTICS)  
[PRESCRIPTIVE ANALYTICS](#PRESCRIPTIVE-ANALYTICS)  
[BEHAVIOUR(AL) ANALYTICS](#BEHAVIOURAL-ANALYTICS)  
[RELATIONSHIP MINING](#RELATIONSHIP-MINING)  
[CAUSAL INFERENCE](#CAUSAL-INFERENCE)  
[CORRELATION MINING](#CORRELATION-MINING)  
[VISUALIZATION](#VISUALIZATION)  
[VISUALIZATION DASHBOARD](#VISUALIZATION-DASHBOARD)  
[NETWORK](#NETWORK)  
[NETWORK ANALYSIS](#NETWORK-ANALYSIS)  
[TEXT MINING](#TEXT-MINING)

### Data Processing

[ALGORITHM](#ALGORITHM)  
[DATA AGGREGATION](#DATA-INTEGRATION)  
[DATA EXPLORATION AND PREPARATION](#DATA-EXPLORATION-AND-PREPARATION)  
[DATA REPRESENTATION AND TRANSFORMATION](#DATA-REPRESENTATION-AND-TRANSFORMATION)  
[SQL](#SQL)  
[GOODNESS OF FIT](#GOODNESS-OF-FIT)

### Storage and Physical Data Sources

[DISTRIBUTED FILE SYSTEM](#DISTRIBUTED-FILE-SYSTEM)  
[IOT (INTERNET OF THINGS)](#IOT-INTERNET-OF-THINGS)  
[DATABASE](#DATABASE)  
[NoSQL](#NO-SQL-DATABASE)

### Types of Data

[STRUCTURED DATA](#STRUCTURED-DATA)  
[TABULAR SCHEMA](#TABULAR-SCHEMA)  
[UNSTRUCTURED DATA](#UNSTRUCTURED-DATA)  
[FLEXIBLE SCHEMA](#FLEXIBLE-SCHEMA)  
[METADATA](#METADATA)

# Table of Contents

**Table of Contents**

[Glossary 1](#__RefHeading___Toc3611_1691662456)

[ADVANCED RESEARCH COMPUTING 1](#__RefHeading___Toc3613_1691662456)

[ALGORITHM 1](#__RefHeading___Toc3615_1691662456)

[ANALYTICS 1](#__RefHeading___Toc3617_1691662456)

[BEHAVIOUR(AL) ANALYTICS 1](#__RefHeading___Toc3619_1691662456)

[BIG DATA 1](#__RefHeading___Toc3621_1691662456)

[CAUSAL INFERENCE 1](#__RefHeading___Toc3623_1691662456)

[CLASSIFICATION 2](#__RefHeading___Toc3625_1691662456)

[CLOUD COMPUTING 2](#__RefHeading___Toc3627_1691662456)

[CLUSTER 2](#__RefHeading___Toc3629_1691662456)

[CLUSTERING 2](#__RefHeading___Toc3631_1691662456)

[COMPUTING WITH DATA 2](#__RefHeading___Toc3633_1691662456)

[CORRELATION MINING 2](#__RefHeading___Toc3635_1691662456)

[DATA 2](#__RefHeading___Toc3637_1691662456)

[DATABASE 3](#__RefHeading___Toc3639_1691662456)

[DATA EXPLORATION AND PREPARATION 3](#__RefHeading___Toc3641_1691662456)

[DATA INTEGRATION 3](#__RefHeading___Toc3643_1691662456)

[DATA MINING 3](#__RefHeading___Toc3645_1691662456)

[DATA MODELING 3](#__RefHeading___Toc3647_1691662456)

[DATA REPRESENTATION AND TRANSFORMATION 3](#__RefHeading___Toc3649_1691662456)

[DATA SCIENCE 3](#__RefHeading___Toc3651_1691662456)

[DESCRIPTIVE ANALYTICS 4](#__RefHeading___Toc3653_1691662456)

[DISTRIBUTED FILE SYSTEM 4](#__RefHeading___Toc3655_1691662456)

[FEATURE 4](#__RefHeading___Toc3657_1691662456)

[FEATURE ENGINEERING 4](#__RefHeading___Toc3659_1691662456)

[FLEXIBLE SCHEMA 4](#__RefHeading___Toc3661_1691662456)

[GENERATIVE MODELING 4](#__RefHeading___Toc3663_1691662456)

[GOODNESS OF FIT 4](#__RefHeading___Toc3665_1691662456)

[INFORMATION 4](#__RefHeading___Toc3667_1691662456)

[INSIGHT 5](#__RefHeading___Toc3669_1691662456)

[IOT (INTERNET OF THINGS) 5](#__RefHeading___Toc3671_1691662456)

[KNOWLEDGE DISCOVERY 5](#__RefHeading___Toc3673_1691662456)

[MACHINE LEARNING 5](#__RefHeading___Toc3675_1691662456)

[METADATA 5](#__RefHeading___Toc3677_1691662456)

[MODEL FITTING 5](#__RefHeading___Toc3679_1691662456)

[NETWORK 5](#__RefHeading___Toc3681_1691662456)

[NETWORK ANALYSIS 5](#__RefHeading___Toc3683_1691662456)

[NO-SQL DATABASE 6](#__RefHeading___Toc3685_1691662456)

[PREDICTIVE ANALYTICS 6](#__RefHeading___Toc3687_1691662456)

[PRESCRIPTIVE ANALYTICS 6](#__RefHeading___Toc3689_1691662456)

[REGRESSION 6](#__RefHeading___Toc3691_1691662456)

[RELATIONSHIP MINING 6](#__RefHeading___Toc3693_1691662456)

[SECURE COMPUTING 6](#__RefHeading___Toc3695_1691662456)

[SQL 6](#__RefHeading___Toc3697_1691662456)

[STRUCTURED DATA 6](#__RefHeading___Toc3699_1691662456)

[TABULAR SCHEMA 7](#__RefHeading___Toc3701_1691662456)

[TEXT MINING 7](#__RefHeading___Toc3703_1691662456)

[UNSTRUCTURED DATA 7](#__RefHeading___Toc3705_1691662456)

[VISUALIZATION 7](#__RefHeading___Toc3707_1691662456)

[VISUALIZATION DASHBOARD 7](#__RefHeading___Toc3709_1691662456)

[Categories 7](#__RefHeading___Toc3711_1691662456)

[Data Inquiry Disciplines 7](#__RefHeading___Toc3713_1691662456)

[Data Analysis Techniques 8](#__RefHeading___Toc3715_1691662456)

[Data Processing 8](#__RefHeading___Toc3717_1691662456)

[Storage and Physical Data Sources 8](#__RefHeading___Toc3719_1691662456)

[Types of Data 8](#__RefHeading___Toc3721_1691662456)

[Table of Contents 9](#__RefHeading___Toc3723_1691662456)

[References 11](#__RefHeading___Toc3725_1691662456)

# References

1. “Westgrid: What We Do.” [Online]. Available: <https://www.westgrid.ca/about_westgrid/what_we_do> [↑](#endnote-ref-2)
2. “Compute Canada. Update January 11, 2017.” [Online]. Available: <https://www.computecanada.ca/wp-content/uploads/2015/02/Update-January-11-2017.pdf> [↑](#endnote-ref-3)
3. “A beginner’s guide to big data terminology.” [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-4)
4. “Mini-glossary: Big data terms you should know.” [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-5)
5. “A beginner’s guide to big data terminology.” [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-6)
6. “Mini-glossary: Big data terms you should know.” [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-7)
7. “Behavioral analytics, definition by Techopedia.” [Online]. Available: <https://www.techopedia.com/definition/30308/behavioral-analytics> [↑](#endnote-ref-8)
8. “Detecting advanced threats with user behavior analytics.” [Online]. Available: <http://www.networkworld.com/article/2904356/security0/detecting-advanced-threats-with-user-behavior-analytics.html> [↑](#endnote-ref-9)
9. “Big data definition by Tech Target.” [Online]. Available: <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data> [↑](#endnote-ref-10)
10. “What Is Big Data?” [Online]. Available: [https://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/#4fb6509b5c85](https://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/" \l "4fb6509b5c85) [↑](#endnote-ref-11)
11. “Who Coined ‘Cloud Computing’?” [Online]. Available: <https://www.technologyreview.com/s/425970/who-coined-cloud-computing> [↑](#endnote-ref-12)
12. “Cloud computing: a simple introduction.” [Online]. Available: <http://www.explainthatstuff.com/cloud-computing-introduction.html> [↑](#endnote-ref-13)
13. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-14)
14. “Introduction to Correlation”. Available: https://www.datascience.com/blog/introduction-to-correlation-learn-data-science-tutorials [↑](#endnote-ref-15)
15. “A beginner’s guide to big data terminology.” [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-16)
16. “Mini-glossary: Big data terms you should know.” [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-17)
17. “A beginner’s guide to big data terminology.” [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-18)
18. “Mini-glossary: Big data terms you should know.” [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-19)
19. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-20)
20. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-21)
21. “The Data Science Venn Diagram”. Available: <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram> [↑](#endnote-ref-22)
22. “Descriptive analytics definition by Tech Target.” [Online]. Available: <http://whatis.techtarget.com/definition/descriptive-analytics> [↑](#endnote-ref-23)
23. “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> [↑](#endnote-ref-24)
24. “Distributed file system definition by Tech Target.” [Online]. Available: <http://searchwindowsserver.techtarget.com/definition/distributed-file-system-DFS> [↑](#endnote-ref-25)
25. “Distributed file system definition by Techopedia.” [Online]. Available: <https://www.techopedia.com/definition/1825/distributed-file-system-dfs> [↑](#endnote-ref-26)
26. “Data Modeling Introduction”. Available: <https://docs.mongodb.com/manual/core/data-modeling-introduction/> [↑](#endnote-ref-27)
27. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-28)
28. “Data vs. Information vs. Insight”. Available: https://online.ben.edu/programs/mba/resources/data-vs-information-vs-insight [↑](#endnote-ref-29)
29. “Data vs. Information vs. Insight”. Available: https://online.ben.edu/programs/mba/resources/data-vs-information-vs-insight [↑](#endnote-ref-30)
30. “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](https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/" \l "34e4f89b1d09) [↑](#endnote-ref-31)
31. “Internet of things definition by Tech Target.” [Online]. Available: <http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT> [↑](#endnote-ref-32)
32. “Overview of the KDD Process”. Available: http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1\_kdd.html [↑](#endnote-ref-33)
33. Greenberg, J. (2003). Metadata and the world wide web. *Encyclopedia of library and information science*, *3*, 1876-1888. [↑](#endnote-ref-34)
34. “Computer network definition by Techopedia.” [Online]. Available: <https://www.techopedia.com/definition/25597/computer-network> [↑](#endnote-ref-35)
35. “What is computer networking?” [Online]. Available: <https://www.lifewire.com/what-is-computer-networking-816249) [www.lifewire.com/what-is-computer-networking-816249.](https://www.lifewire.com/what-is-computer-networking-816249> [↑](#endnote-ref-36)
36. “Social Network Analysis: An Introduction by Orgnet,LLC.” [Online]. Available: <http://www.orgnet.com/sna.html> [↑](#endnote-ref-37)
37. “A beginner’s guide to big data terminology.” [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-38)
38. Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. [↑](#endnote-ref-39)
39. “A beginner’s guide to big data terminology.” [Online]. Available: <http://dataconomy.com/2016/05/a-beginners-guide-to-big-data-terminology> [↑](#endnote-ref-40)
40. “The myth of secure computing.” [Online]. Available: <https://hbr.org/2003/06/the-myth-of-secure-computing) [hbr.org/2003/06/the-myth-of-secure-computing.](https://hbr.org/2003/06/the-myth-of-secure-computing> [↑](#endnote-ref-41)
41. “Mini-glossary: Big data terms you should know.” [Online]. Available: <http://www.techrepublic.com/article/mini-glossary-big-data-terms-you-should-know> [↑](#endnote-ref-42)
42. “Introduction to Structured Data, by Google.” [Online]. Available: <https://developers.google.com/search/docs/guides/intro-structured-data> [↑](#endnote-ref-43)
43. “Structured data definition by Webopedia.” [Online]. Available: <http://www.webopedia.com/TERM/S/structured_data.html> [↑](#endnote-ref-44)
44. “Introduction to Tabular Data” Available: <https://papl.cs.brown.edu/2016/intro-tabular-data.html> [↑](#endnote-ref-45)
45. “Tabular Database”. Available: <https://www.techopedia.com/definition/26181/tabular-database> [↑](#endnote-ref-46)
46. “About Text Mining”. Available: https://www.ibm.com/support/knowledgecenter/en/SS3RA7\_15.0.0/com.ibm.spss.ta.help/tm\_intro\_tm\_defined.htm [↑](#endnote-ref-47)
47. “Unstructured data definition by Techopedia.” [Online]. Available: <https://www.techopedia.com/definition/13865/unstructured-data> [↑](#endnote-ref-48)
48. “Solving the Unstructured Data Challenge.” [Online]. Available: <http://www.cio.com/article/2941015/big-data/solving-the-unstructured-data-challenge.html> [↑](#endnote-ref-49)
49. “Data visualization definition by Tech Target.” [Online]. Available: <http://searchbusinessanalytics.techtarget.com/definition/data-visualization> [↑](#endnote-ref-50)
50. “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> [↑](#endnote-ref-51)