



# BIG DATA JARGON BUSTER

Technological developments are transforming the way organisations manage and analyse their data. In recent years the volume and variety of data available for analysis has expanded exponentially with an almost unlimited wealth of data sources from call-centre recordings to customers' social media posts.

This has catalysed the emergence of increasingly powerful technologies to enable more sophisticated

data management and analytics. The software industry continues to develop a growing range of tools and applications to help extract insights from data.

As with many IT-based trends, the data-driven world comes with its own set of jargon. Here are the key terms finance professionals should get to grips with.

## Data analytics

Data analytics refers to advanced forms of analysis that can be used to explore large volumes of data and communicate insights. These can be used to identify correlations and develop algorithms to predict behaviours. Data analytics with regard to structured or enterprise data is well established. It is used by many companies and organisations to help make better business decisions, and to test and validate models or theories. There are now high expectations of data analytics with regard to new forms of unstructured data.

## Cloud computing

Cloud computing refers to the provision of various services, such as software applications, development platforms, servers, processing power and storage, via remote servers over the internet, as opposed to on a local server. Typically referred to as the 'cloud', it often entails users paying for IT services as needed, while the back-end application or infrastructure is managed by a third party vendor.

## Dashboards

A data dashboard is a user interface that organises and presents corporate information in a way that is easy to read and interpret. It can be used to aggregate a range of data and KPIs, often visually, for managers to monitor business performance.

## Data mining

A set of techniques used to sift through very large amounts of data. Data mining uses artificial intelligence techniques and advanced statistical tools (such as cluster analysis and regressions) to reveal trends, patterns and relationships.



## Data scientist

A data scientist is someone who performs statistical analysis and data mining on large volumes of data, typically to identify trends, figures and other relevant information. Makes use of advanced modelling, statistics, analytics and mathematics techniques.

## Data visualisation

Data visualisation is a general term used to describe technology that enables business managers to see trends and data patterns. These tools often go beyond the standard charts and graphs used in Excel spreadsheets, using more intuitive dials and gauges, geographic maps, time-series charts, heat maps and so on. Patterns, trends and correlations that may otherwise be missed can be spotted more easily with data visualisation software.

## Hadoop

Hadoop is a Java-based programming framework that supports the processing of large data sets in a distributed computing environment. It is available as open source software from Apache, and is commonly used to handle huge data volumes, spanning thousands of servers.

## In-memory processing

In-memory processing enables businesses to analyse large data sets significantly faster than before, by making use of random access memory to process tasks, rather than slower hard disk drives or other storage platforms.

## 'Internet of things'

The 'internet of things' describes the connecting of everyday physical objects to the internet, allowing them to provide information or alerts as a node on the network. This provides organisations with a vast new source of information on every aspect of their business, by connecting and monitoring machinery, vehicles, equipment, stock items and much more.



## MapReduce

MapReduce is a software framework that allows developers to write programmes that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers. It provides a valuable reference to help organise and locate particular types of data. Examples of metadata on a simple document would include metadata on the date the document was created, date modified and file. Metadata is also used for images, videos, spreadsheets and web pages.

## OLAP

Short for online analytical processing, a category of software tools that provides analysis of data stored in a database. OLAP tools enable users to analyse different dimensions of multidimensional data. For example, it provides time series and trend analysis views.

## Predictive analytics

Predictive analytics is the branch of data mining concerned with forecasting probabilities. It uses variables that can be measured to predict the future behaviour of a person or other entity. Predictive analytics leverages an organisation's business knowledge by applying sophisticated analysis techniques to enterprise data. In business, predictive analytics are often used to answer questions about customer behaviour and offer suggestions on how best to target resources for maximum return.

## Social media

Social media is the umbrella term for software tools and platforms that allow groups to generate content and engage in peer-to-peer conversations. Facebook and Twitter are merely the two most high profile examples of such platforms.

## Unstructured data

The term 'unstructured data' refers to any data that has no pre-defined structure, and thus cannot be easily stored within standard relational databases. Examples include email, text-based documents, images, videos and call-centre recordings.

## Further reading

For more information and resources visit [cgma.org/data](http://cgma.org/data)