

Algorithm is a mathematical formula placed in software to perform analysis of a data set.

Analytics Platform is software and/or hardware that provide the tools and computational power needed to build and perform many different analytical queries.

Behavioural Analytics is the use of data about people's behaviour to understand intent and predict future actions.

Business Intelligence (BI) is a general term used for the identification, extraction, and analysis of data.

Cassandra is a popular choice of columnar database for use in big data applications.

Columnar Database (or Column-Oriented Database) is a database that stores data by column rather than by row. A key advantage of a columnar database is faster hard disk access.

Data Architecture & Design has three stages or processes. 1. Conceptual representation of business entities. 2. The logical representation of the relationships among those entities. 3. The physical construction of the system, to support the functionality.

Data-Directed Decision Making is the use of data to support making crucial decisions.

Data Mining is the process of deriving patterns or knowledge from large data sets.

Data Model / Modelling is the structure of the data for the purpose of communicating between functional and technical people to show data needed for business processes, or for communicating a plan to develop how data is stored and accessed among application development team members.

Data Profiling is the process of collecting statistics.

Data Science is a discipline that incorporates statistics, data visualisation, computer programming, data mining, machine learning, and database engineering to solve complex problems.

Data Set is a collection of data, typically in tabular form.

Data Visualisation is a visual representation of data designed for the purpose of communicating information more effectively.

Drill is an open source distributed system for performing interactive analysis on large-scale datasets.

Hadoop is a framework that allows for the distributed processing of large data sets across clusters of

computers using a simple programming model.

HBase is a distributed columnar NoSQL database.

Hive is a SQL-like query and data warehouse engine.

Machine-Learning is the use of algorithms to allow a computer to analyse data for the purpose of "learning" what action to take when a specific pattern or event occurs.

Map/reduce is a general term that refers to the process of breaking up a problem into pieces that are then distributed across multiple computers on the same network or cluster, or across a grid of disparate and possibly geographically separated systems (map), and then collecting all the results and combines them into a report (reduce). Google's branded framework to perform this function is called MapReduce.

MongoDB is an open-source NoSQL database managed by 10gen.

NoSQL is a class of database management system that does not use the relational model. NoSQL is designed to handle large data volumes that do not follow a fixed schema.

Pattern Recognition is the classification or labelling of an identified pattern in the machine learning process.

Pig is a data flow language and execution framework for parallel computation.

Predictive Analytics uses statistical functions on one or more datasets to predict trends or future events.

Predictive Modelling is the process of developing a model that will most likely predict a trend or outcome.

R is an open-source software environment used for statistical computing.

Storm is an open-source distributed computation system designed for processing multiple data streams in real time.

We have a number of roles for Hadoop Professionals or Data Scientist. We would be happy to talk to you about your needs and then introduce you to some clients, who we feel would suit you.

Just phone us on 020 8123 7769 or email us: rod@resourceondemand.com and one of our team will be happy to help.