*BIG DATA*

Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information.

Because big data takes too much time and costs too much money to load into a **traditional** relational database for analysis, new approaches to storing and analyzing data have emerged that rely less on data schema and data quality. Instead, raw data with **extended metadata** is aggregated in a **data** **lake** and machine learning and artificial intelligence programs use complex algorithms to look for repeatable patterns.

Big data analytics is often associated with **cloud computing** because the analysis of large data

sets in real-time requires a platform like Hadoop to store large data sets across a

distributed cluster and MapReduce to coordinate, combine and process data from multiple sources.

Although the demand for big data analytics is high, there is currently a shortage of data scientists and other analysts who have experience working with big data in a distributed, **open source**

environment. In the enterprise, vendors have responded to this shortage by creating Hadoop

appliances to help companies take advantage of the semi-structured and unstructured data they

own.

Big data can be contrasted with **small data**, another evolving term that's often used to describe

data whose volume and format can be easily used for **self-service analytics**. A commonly quoted

axiom is that "big data is for machines; small data is for people.”

Большие данные - это развивающийся термин, который описывает любое объемное количество структурированных, полуструктурированных и неструктурированных данных, которые могут быть заминированы для информации.  
Поскольку большие данные занимают слишком много времени и стоят слишком много денег для загрузки в традиционную реляционную базу данных для анализа, появились новые подходы к хранению и анализу данных, которые меньше зависят от схемы данных и качества данных. Вместо этого необработанные данные с расширенными метаданными агрегируются в окне данных, а программы машинного обучения и искусственного интеллекта используют сложные алгоритмы для поиска повторяемых шаблонов.  
Большая аналитика данных часто связана с облачными вычислениями, поскольку анализ больших данных  
в режиме реального времени требуется такая платформа, как Hadoop, для хранения больших наборов данных через  
распределенный кластер и MapReduce для координации, объединения и обработки данных из нескольких источников.  
Хотя спрос на большую аналитику данных высок, в настоящее время наблюдается нехватка ученых-исследователей данных и других аналитиков, которые имеют опыт работы с большими данными в распределенном, с открытым исходным кодом  
Окружающая среда. На предприятии поставщики ответили на этот недостаток, создав Hadoop  
приборы, чтобы помочь компаниям использовать полуструктурированные и неструктурированные данные, которые они  
своя.  
Большим данным можно сопоставить небольшие данные, еще один развивающийся термин, который часто используется для описания  
данные, объем и формат которых могут быть легко использованы для анализа самообслуживания. Обычно цитируется  
аксиома заключается в том, что «большие данные для машин, небольшие данные для людей».