R&D Outsourcing

25 2023

What is big data?

As the name suggests, the term `big data` simply refers to the management and analysis of large amounts of data. According to the McKinsey Institute report `Big data: The next frontier for innovation, competition and productivity` (Big data: The next frontier for innovation, competition and productivity), the term `big data` refers to data sets that are larger than typical databases (DB) for entering, storing, managing and analyzing information. And the world's data repositories are certainly growing. In mid-2011, IDC's Digital Universe Study, a report sponsored by EMC, predicted that the total global volume of data created and replicated in 2011 could be about 1.8 zettabytes (1.8 trillion gigabytes) - about 9 times more than what was created in 2006. More complex definition However, `big data` involves more than just the analysis of huge amounts of information. The problem is not that organizations create huge amounts of data, but that most of it is presented in a format that does not fit well with the traditional structured database format - it is web logs, videos, text documents, machine code, or, for example, geospatial data. . All this is stored in many different repositories, sometimes even outside the organization. As a result, corporations can have access to a huge amount of their data and not have the necessary tools to establish relationships between these data and draw meaningful conclusions from them. Add to this the fact that data is now being updated more and more often, and you get a situation in which traditional methods of information analysis cannot keep up with huge amounts of constantly updated data, which ultimately paves the way for big data technologies.

Best Definition
In essence, the concept of big data implies working with information of a huge volume and diverse composition, very often updated and located in different sources in order to increase work efficiency, create new products and increase competitiveness. The consulting firm Forrester puts it short: `Big data brings together techniques and technologies that extract meaning from data to the extreme limit of practicality`.

How big is the difference between business intelligence and big data?
Craig Bathy, Executive Director of Marketing and Chief Technology Officer of Fujitsu Australia, pointed out that business analysis is a descriptive process of analyzing the results achieved by a business in a certain period of time, while the speed of processing big data allows analysis to be predictive, able to offer business recommendations on future. Big data technologies also allow you to analyze more types of data than business intelligence tools, which makes it possible to focus not only on structured storage.

Matt Slocum of O'Reilly Radar believes that while big data and business intelligence have the same goal (answering a question), they differ in three ways.

Big data is designed to process larger amounts of information than business intelligence, and this, of course, fits the traditional definition of big data.
Big data is designed to process faster and more rapidly changing information, which means deep exploration and interactivity. In some cases, the results are generated faster than the web page loads.

Big data is designed to handle unstructured data that we are only just beginning to explore how to use it after we have been able to collect and store it, and we need algorithms and dialogue to make it easier to find the trends contained within these arrays.

According to Oracle's white paper, Oracle Information Architecture: An Architect's Guide to Big Data, we approach information differently when we work with big data than when we do business analysis.

Working with big data is not like a typical business intelligence process, where simply adding together known values yields results: for example, adding bills paid together becomes sales for a year. When working with big data, the result is obtained in the process of cleaning them through sequential modeling: first, a hypothesis is put forward, a statistical, visual or semantic model is built, on its basis the correctness of the hypothesis put forward is checked, and then the next one is put forward. This process requires the researcher to either interpret visual meanings or make interactive knowledge-based queries, or develop adaptive 'machine learning' algorithms capable of producing the desired result. Moreover, the lifetime of such anGorithma can be quite short.

Link to article:: [What is big data?](#)