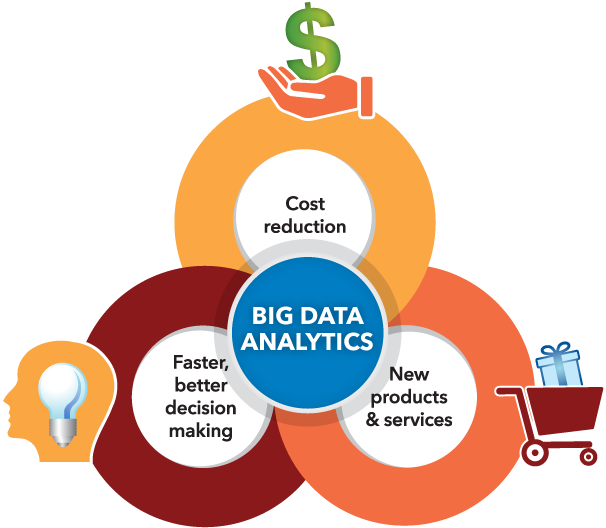
BIG DATA ANALYSIS with IBM

CLOUD DATABASES

DEVELOPMENT PART-2

BIG DATA ANALYSIS

Big data analytics describes the process of uncovering trends, patterns, and correlations in large amounts of raw data to help make data-informed decisions. These processes use familiar statistical analysis techniques—like clustering and regression—and apply them to more extensive datasets with the help of newer tools.

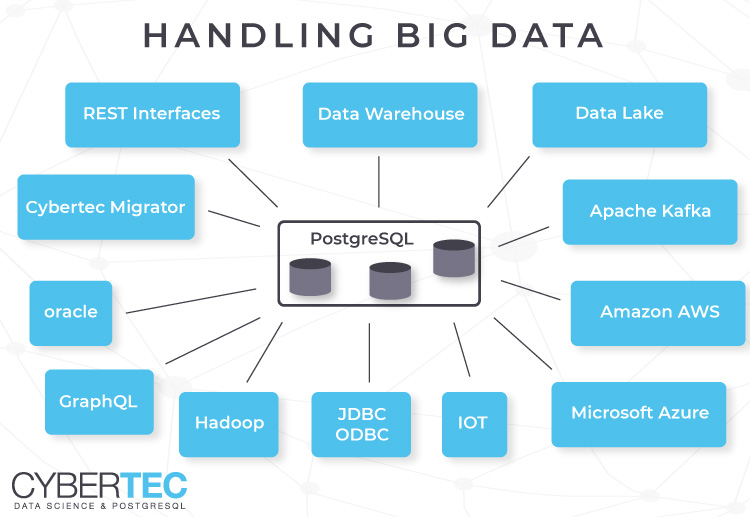


Big data analytics uses the four data analysis methods to uncover meaningful insights and derive solutions. For example, big data analytics is integral to the modern health care industry. As you can imagine, systems that must manage thousands of patient records, insurance plans, prescriptions, and vaccine information.

With big data, you can analyze and assess production, customer feedback and returns, and other factors to reduce outages and anticipate future demands. Big data can also be used to improve decision-making in line with current market demand.

Data analysis is a process for obtaining raw data, and subsequently converting it into information useful for decision-making by users. Data is collected and analyzed to answer questions, test hypotheses, or disprove theories.

 The future of big data will witness more and more predictive analytics used for decision-making.



**Big data analytics tools**

Harnessing all of that data requires tools. Thankfully, technology has advanced so that many intuitive software systems are available for data analysts to use.

* **Hadoop:** An open-source framework that stores and processes big data sets. Hadoop can handle and analyse structured and unstructured data.
* **Spark:** An open-source cluster computing framework for real-time processing and data analysis.
* **Data integration software**: Programs that allow big data to be streamlined across different platforms, such as MongoDB, Apache, Hadoop, and Amazon EMR.
* **Stream analytics tools:** Systems that filter, aggregate, and analyse data that might be stored in different platforms and formats, such as Kafka.
* **Distributed storage:** Databases that can split data across multiple servers and can identify lost or corrupt data, such as Cassandra.
* **Predictive analytics hardware and software:** Systems that process large amounts of complex data, using machine learning and algorithms to predict future outcomes, such as fraud detection, marketing, and risk assessments.
* **Data mining tools:** Programs that allow users to search within structured and unstructured big data.
* **NoSQL databases:** Non-relational data management systems ideal for dealing with raw and unstructured data.
* **Data warehouses:**Storage for large amounts of data collected from many different sources, typically using predefined schemas.

## Types of big data analytics

Four main types of big data analytics support and inform different business decisions.

### 1. Descriptive analytics

Descriptive analytics refers to data that can be easily read and interpreted. This data helps create reports and visualise information that can detail company profits and sales.

Example: During the pandemic, a leading pharmaceutical company conducted data analysis on its offices and research labs. Descriptive analytics helped them identify consolidated unutilised spaces and departments, saving the company millions of pounds.

### 2. Diagnostics analytics

Diagnostics analytics helps companies understand why a problem occurred. Big data technologies and tools allow users to mine and recover data that helps dissect an issue and prevent it from happening in the future.

Example: An online retailer’s sales have decreased even though customers continue to add items to their shopping carts. Diagnostics analytics helped to understand that the payment page was not working correctly for a few weeks.

### 3. Predictive analytics

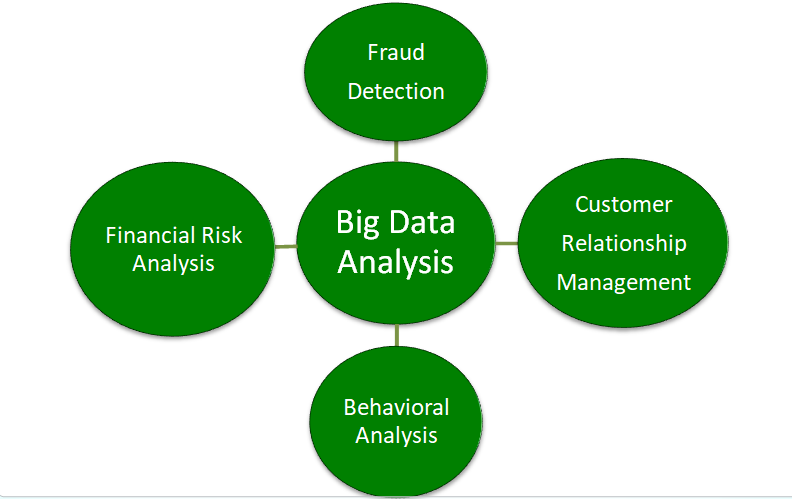
Predictive analytics looks at past and present data to make predictions. With artificial intelligence (AI), machine learning, and data mining, users can analyse the data to predict market trends.

Example: In the manufacturing sector, companies can use algorithms based on historical data to predict if or when a piece of equipment will malfunction or break down.

### 4. Prescriptive analytics

Prescriptive analytics solves a problem, relying on AI and machine learning to gather and use data for risk management.

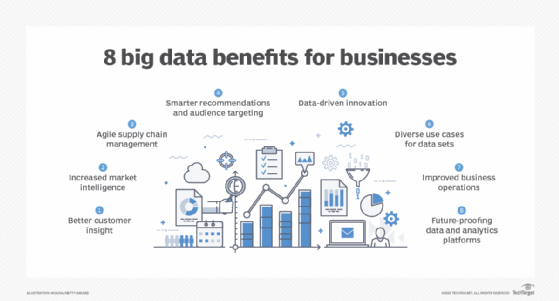
Example: Within the energy sector, utility companies, gas producers, and pipeline owners identify factors that affect the price of oil and gas to hedge risks.



### Benefits of big data analytics

Incorporating big data analytics into a business or organisation has several advantages. These include:

* **Cost reduction:**Big data can reduce costs in storing all business data in one place. Tracking analytics also helps companies find ways to work more efficiently to cut costs wherever possible.
* **Product development:** Developing and marketing new products, services, or brands is much easier when based on data collected from customers’ needs and wants. Big data analytics also helps businesses understand product viability and to keep up with trends.
* **Strategic business decisions:**The ability to constantly analyse data helps businesses make better and faster decisions, such as cost and supply chain optimisation.
* **Customer experience:** Data-driven algorithms help marketing efforts (targeted ads, for example) and increase customer satisfaction by delivering an enhanced customer experience.
* **Risk management:** Businesses can identify risks by analysing data patterns and developing solutions for managing those risks.



This is possible, because [big data systems can integrate data](https://www.techtarget.com/searchdatamanagement/feature/Establish-big-data-integration-techniques-and-best-practices) on customer trends from e-commerce sites and retail applications with supplier data, real-time pricing and even shipping and weather information to provide a level of information not seen before.

**THANK YOU**