

**What is Data Analytics?**

In this new digital world, data is being generated in an enormous amount which opens new paradigms. As we have high computing power as well as a large amount of data we can make use of this data to help us make data-driven decision making. The main benefits of data-driven decisions are that they are made up by observing past trends which have resulted in beneficial results.

In short, we can say that data analytics is the process of manipulating data to extract useful trends and hidden patterns which can help us derive valuable insights to make business predictions.

**Data analytics** is the practice of examining data to answer questions, identify trends, and extract insights. When data analytics is used in business, it’s often called [business analytics](https://online.hbs.edu/blog/post/importance-of-business-analytics).

You can use tools, frameworks, and software to analyze data, such as Microsoft Excel and Power BI, Google Charts, Data Wrapper, Infogram, Tableau, and Zoho Analytics. These can help you examine data from different angles and [create visualizations](https://online.hbs.edu/blog/post/data-visualization-tools) that illuminate the story you’re trying to tell.

Algorithms and machine learning also fall into the data analytics field and can be used to gather, sort, and analyze data at a higher volume and faster pace than humans can. Writing algorithms is a more advanced data analytics skill, but you don’t need deep knowledge of coding and statistical modeling to experience the benefits of data-driven decision-making.

|  |  |
| --- | --- |
| The name “Hadoop” was the named after Doug cutting’s son’s toy elephant. He named this project as “Hadoop” as it was easy to pronounce it. | The “MapReduce” name came into existence as per the functionality itself of mapping and reducing in key-value pairs. |

Hive, Flink, and Spark: Big Data Tools in a Nutshell

Imagine you have a massive warehouse full of stuff you need to organize and analyze. These three tools are like your trusty helpers, each with their own strengths:

Hive:

* Think of it as: A spreadsheet expert. It lets you easily query and analyze large datasets stored in Hadoop, like tables in a giant spreadsheet. You can use familiar SQL-like commands to filter, summarize, and join data.

Flink:

* Think of it as: A speedy delivery driver. It excels at processing data streams in real-time, like a constant flow of packages you need to sort and distribute as they arrive. It's great for things like fraud detection or analyzing sensor data.

Spark:

* Think of it as: A multi-talented assistant. It can handle various tasks, from simple data analysis to complex machine learning algorithms. It's like having a Swiss Army knife for big data, working both with batch processing (like organizing a warehouse) and real-time analysis (like managing incoming deliveries).

Here's a quick table summarizing their key differences:

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Hive | Flink | Spark |
| Main strength | Analyzing large stored datasets | Real-time data processing | Versatile: batch & real-time, various tasks |
| Data type | Tables | Streams | Both tables & streams |
| Learning curve | Easy (familiar SQL) | Moderate | Moderate |
| Typical use cases | Data warehousing, reporting | Fraud detection, sensor analysis | Machine learning, complex analytics |

Remember, these are just simplified analogies. Each tool has its own nuances and complexities. But hopefully, this gives you a basic understanding of how they can help you navigate the world of big data!