Big Data

Big data is a term that describes large, hard-to-manage volumes of data – both structured and unstructured – that inundate businesses on a day-to-day basis. But it's not just the type or amount of data that's important, it's what organisations do with the data that matters. Big data can be analysed for insights that improve decisions and give confidence for making strategic business moves.

Doug Laney articulated the now-mainstream definition of big data as the three V's:

Volume. Organisations collect data from a variety of sources, including transactions, smart (IoT) devices, industrial equipment, videos, images, audio, social media and more. In the past, storing all that data would have been too costly – but cheaper storage using data lakes, Hadoop and the cloud have eased the burden.

Velocity. With the growth in the [Internet of Things](), data streams into businesses at an unprecedented speed and must be handled in a timely manner. RFID tags, sensors and smart meters are driving the need to deal with these torrents of data in near-real time.

Variety. Data comes in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, emails, videos, audios, stock ticker data and financial transactions.

Big Data

The basics of what it is and why it matters

What is big data?[1]

Big data sets are too large and complex to be processed by traditional methods.
Consider that in a single minute there are:

64,140 Instagram stories posted	**336,480** Skype calls	**567,360** Tweets sent	**5,365,260** YouTube videos watched	**5,500,560** Google searches conducted	**181,331,340** Emails sent

The 3 V's of big data: Plus 2

These are the defining properties or dimensions of big data.

Volume	Velocity	Variety	Variability	Veracity
amount of data	speed of data	types of data	variety/changes in data flows	quality of data

How do organizations optimize the value of big data?

Regardless of location, size, sources, owners or users, these steps can unleash
value from an organization's complex data landscape (data fabric).

Set a big data strategy. → Identify big data sources. → Access, manage and store data via automated data pipelines. → Analyze big data. → Make data-driven decisions.

You don't have to be big to use big data

Businesses of all sizes use big data with analytics and the cloud to be more competitive,
achieve digital transformation or dominate in a market. In turn, they can:

- Advance business intelligence.
- Minimize risk and fraud.
- Save time and cut costs.
- Boost productivity.
- Build stronger customer relationships.

Trends in big data[2]

Mobile and real-time data dominate.
By 2025, over a quarter of data will be real time in nature and IoT real-time data will account for more than 95% of it.

Artificial intelligence transforms the norm.
Insights are generated via new technologies like machine learning and natural language processing.

Security stays significant.
With increasing amounts of data being produced, protection and security of sensitive and private information is crucial.

To be most useful, big data needs big data analytics. Read our primer to learn more.

§sas

Why Is Big Data Important?

The importance of big data doesn't simply revolve around how much data you have. The value lies in how you use it. By taking data from any source and analysing it, you can find answers that 1) streamline resource management, 2) improve operational efficiencies, 3) optimise product development, 4) drive new revenue and growth opportunities and 5) enable smart decision making. When you combine big data with high-performance analytics, you can accomplish business-related tasks such as:

- Determining root causes of failures, issues and defects in near-real time.
- Spotting anomalies faster and more accurately than the human eye.
- Improving patient outcomes by rapidly converting medical image data into insights.
- Recalculating entire risk portfolios in minutes.
- Sharpening deep learning models' ability to accurately classify and react to changing variables.
- Detecting fraudulent behaviour before it affects your organisation.

How Big Data Works

Before businesses can put big data to work for them, they should consider how it flows among a multitude of locations, sources, systems, owners and users. There are five key steps to taking charge of this "big data fabric" that includes traditional, structured data along with unstructured and semistructured data:

- Set a big data strategy.
- Identify big data sources.
- Access, manage and store the data.

- Analyse the data.
- Make intelligent, data-driven decisions.

1) Set a big data strategy

At a high level, a big data strategy is a plan designed to help you oversee and improve the way you acquire, store, manage, share and use data within and outside of your organisation. A big data strategy sets the stage for business success amid an abundance of data. When developing a strategy, it's important to consider existing – and future – business and technology goals and initiatives. This calls for treating big data like any other valuable business asset rather than just a byproduct of applications.

2) Identify big data sources

- **Streaming data** comes from the Internet of Things (IoT) and other connected devices that flow into IT systems from wearables, smart cars, medical devices, industrial equipment and more. You can analyse this big data as it arrives, deciding which data to keep or not keep, and which needs further analysis.
- **Social media** data stems from interactions on Facebook, YouTube, Instagram, etc. This includes vast amounts of big data in the form of images, videos, voice, text and sound – useful for marketing, sales and support functions. This data is often in

unstructured or semistructured forms, so it poses a unique challenge for consumption and analysis.

- **Publicly available data** comes from massive amounts of open data sources like the US government's data.gov, the CIA World Factbook or the European Union Open Data Portal.
- **Other big data** may come from data lakes, cloud data sources, suppliers and customers.

3) Access, manage and store big data

- Modern computing systems provide the speed, power and flexibility needed to quickly access massive amounts and types of big data. Along with reliable access, companies also need methods for integrating the data, building data pipelines, ensuring data quality, providing data governance and storage, and preparing the data for analysis. Some big data may be stored on-site in a traditional [data warehouse](#) – but there are also flexible, low-cost options for storing and handling big data via cloud solutions, data lakes, data pipelines and Hadoop.

4) Analyse the data

- With high-performance technologies like grid computing or [in-memory analytics](#), organisations can choose to use all their big data for analyses. Another approach is to determine upfront which data is relevant before analysing it. Either way, big data analytics is how companies gain value and insights from data. Increasingly, big data feeds today's advanced analytics endeavours such as artificial intelligence (AI) and machine learning.

5) Make intelligent, data-driven decisions

- Well-managed, trusted data leads to trusted analytics and trusted decisions. To stay competitive, businesses need to seize the full value of big data and operate in a data-driven way – making decisions based on the evidence presented by big data rather than gut instinct. The benefits of being data driven are clear. Data-driven organisations perform better, are operationally more predictable and are more profitable.