What is Big Data?

Enter

ORACLE


ORACLE

Table of Contents

ORACLE

Definition of Big Data

What exactly is big data?

To put it simply: big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can't manage them. But these massive volumes of data can be used to address business problems you wouldn't have been able to tackle before.

To really understand big data, it's helpful to have some historical background. Here's Gartner's definition, circa 2001 (which is still the go-to definition):

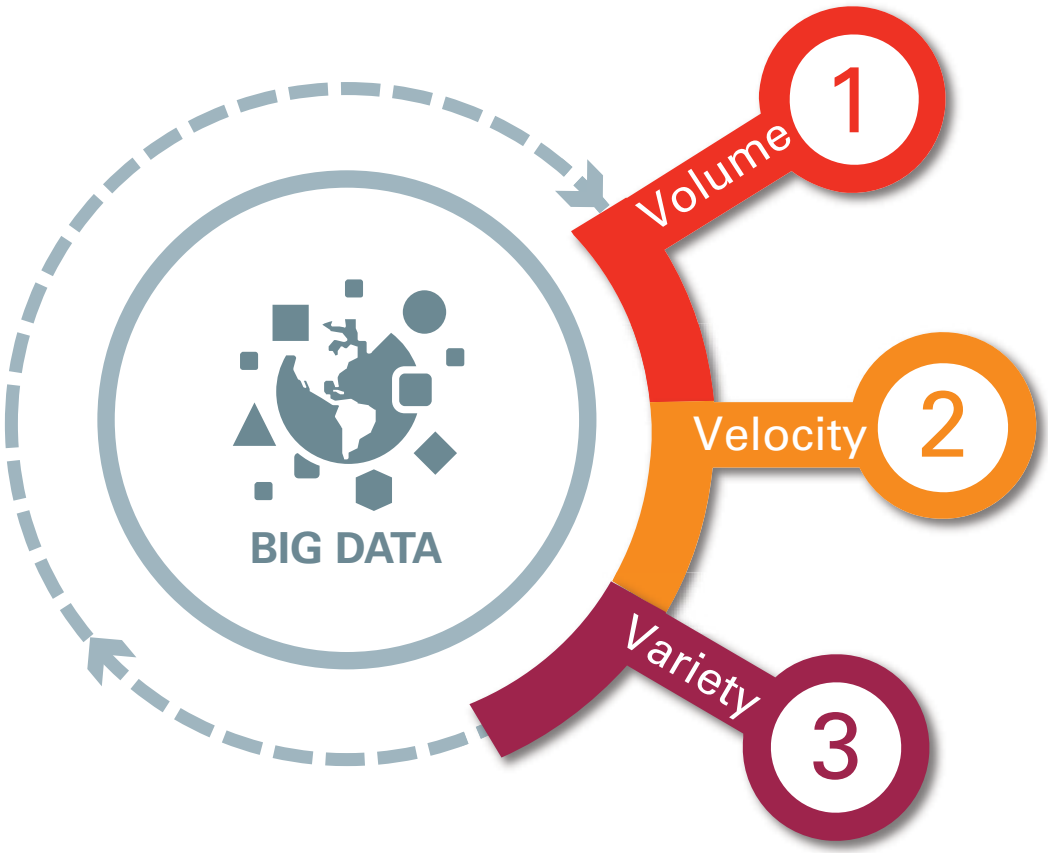
"Big data is data that contains greater variety arriving in increasing volumes and with ever higher velocity. This is known as the three Vs."

THE THREE VS OF BIG DATA

- **Volume.** The amount of data matters. With big data, you'll have to process high volumes of low-density, unstructured data. This can be data of unknown value, such as Twitter data feeds, clickstreams on a webpage or a mobile app, or sensor-enabled equipment. For some organizations, this might be tens of terabytes of data. For others, it may be hundreds of petabytes.

Velocity. Velocity is the fast rate at which data is received and (perhaps) acted on. Normally, the highest velocity of data streams directly into memory versus being written to disk. Some internet-enabled smart products operate in real time or near real time and will require real-time evaluation and action.

- **Variety.** In today's big data world, data comes in new unstructured data types. Unstructured and semi-structured data types, such as text, audio, and video require addition preprocessing to derive meaning and support metadata.



THE VALUE—AND TRUTH—OF BIG DATA

Since 2001, two more Vs have become apparent: **value** and **veracity**. Data has intrinsic value. But it's of no use until that value is discovered.

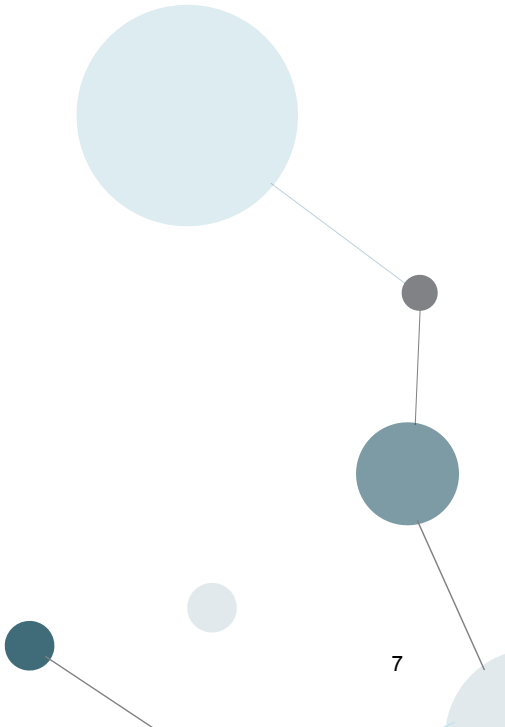
Equally important: How truthful is your data—and how much can you rely on it?

Today, big data has become capital. Think of some of the world's biggest tech companies. A large part of the value they offer comes from their data, which they're constantly analyzing to produce more efficiency and develop new products.

Recent technological breakthroughs have exponentially reduced the cost of data storage and compute, making it easier and less expensive to store more data than ever before. With an increased volume of big data now cheaper and more accessible, you can make more accurate and precise business decisions.

Finding value in big data isn't only about analyzing it (which is a whole other benefit). It's an entire discovery process that requires insightful analysts, business users, and executives who ask the right questions, recognize patterns, make informed assumptions, and predict behavior.

But how did we get here?



The History of Big Data

Around 2005, people began to realize just how much data users generated through Facebook, YouTube, and other online services. Hadoop (an open-source framework created specifically to store and analyze big data sets) was developed that same year. NoSQL also began to gain popularity during this time.

The development of open-source frameworks, such as Hadoop (and more recently, Spark) was essential for the growth of big data because they make big data easier to work with and cheaper to store. In the years since then, the volume of big data has skyrocketed. Users are still generating huge amounts of data—but it's not just humans.

With the advent of Internet of Things (IoT) , more objects and devices are connected to the internet, gathering data on customer usage patterns and product performance. The emergence of machine learning has produced still more data.

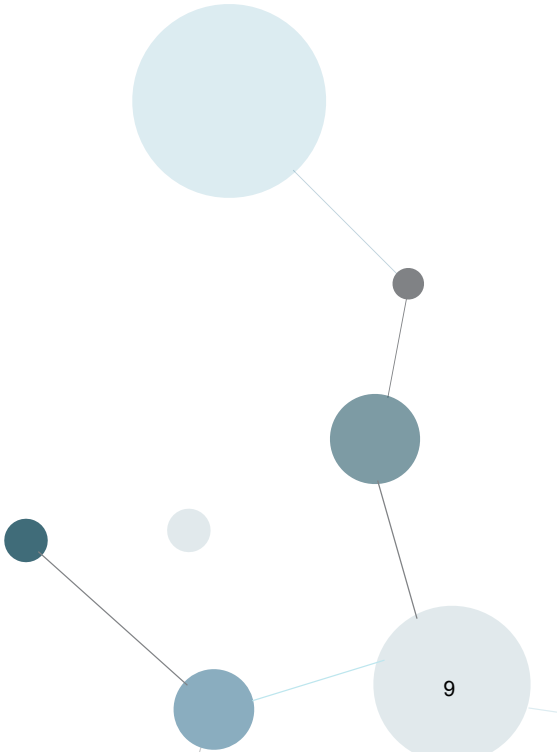
While big data has come far, its popularity is only just beginning. Cloud computing has expanded big data possibilities even further.

The cloud offers a truly elastic scalability, where developers can just spin up ad hoc clusters to test around a subset of data. It's an exciting time to see what's going to happen next.

BENEFITS OF BIG DATA AND DATA ANALYTICS

THE VALUE OF BIG DATA COMES IS TWOFOLD:

1. Big data makes it possible for you to gain more complete answers because you have more information
2. More complete answers means more confidence in the data–which means a completely different approach to tackling problems



Big Data Use Cases

This new knowledge enables you to handle use cases that you haven't been able to fully delve into before. Here are just a few. (More use cases are on our solutions page):

PRODUCT DEVELOPMENT

Companies like Netflix and Procter & Gamble use big data to anticipate customer demand. By classifying key attributes of past and current products or services, and then modeling the relationship between those attributes and the commercial success of the offerings, they build predictive models for new products and services. In addition, P&G uses data and analytics from focus groups, social media, test markets, and early store rollouts to plan, produce, and launch new products.

PREDICTIVE MAINTENANCE

Factors that can predict mechanical failures may be deeply buried in structured data, such as the equipment year, make, and model, as well as in unstructured data that covers millions of log entries, senor data, error messages, and engine temperature., By analyzing these indications of potential issues before the problems happen, organizations can deploy maintenance more cost effectively and maximize parts and equipment uptime.

CUSTOMER EXPERIENCE

The race for customers is on. A clearer view of customer experience is more possible now than ever before. Big data enables you to gather data from social media, web visits, call logs, and other data sources to improve the interaction experience and maximize the value delivered. Start delivering personalized offers, reduce customer churn, and handle issues proactively.

FRAUD AND COMPLIANCE

When it comes to security, it's not just a few rogue hackers; you're up against entire expert teams. Security landscapes and compliance requirements are constantly evolving. Big data helps you identify patterns in data that indicate fraud and aggregate large volumes of information to make regulatory reporting much faster.

MACHINE LEARNING

Machine learning is the hottest of hot topics right now. And data—specifically big data—is one of the reasons why. It's only recently that we've been able to teach machines instead of program them. And the availability of big data to train machine-learning models makes that happen.

OPERATIONAL EFFICIENCY

Operational efficiency may not always make the news, but it's an area in which big data is having the most impact. 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.

DRIVE INNOVATION

Big data can help you innovate by studying interdependencies between humans, institutions, entities, and process and then determining new ways to use those insights. Use data insights to improve decisions about financial and planning considerations. Examine trends and what customers want to deliver new products and services. Implement dynamic pricing. There are endless possibilities.

BIG DATA CHALLENGES

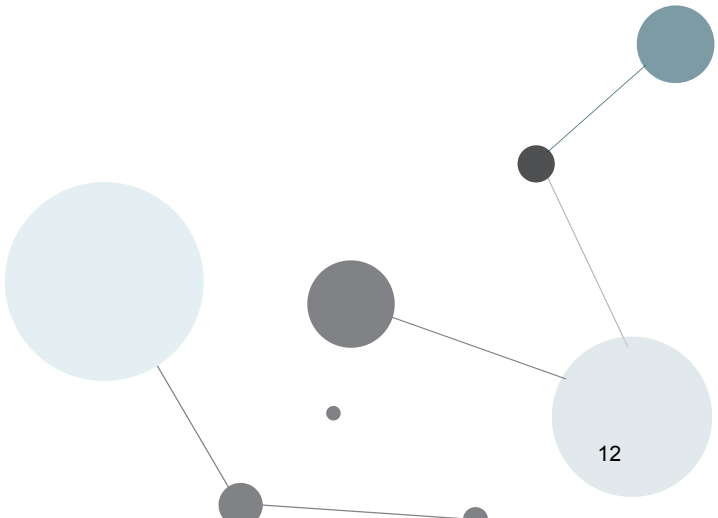
While big data holds a lot of promise, it is not without its challenges.

First, big data is… big. Although new technologies have been developed to store data, data volumes are doubling in size around every two years. Organizations still struggle to keep pace with their data and find ways to effectively store it.

But it's not enough to just store the data. Data must be used to be valuable, and that depends on curation. Clean data, or data that's relevant to the client and organized in a way that enables meaningful analysis requires a lot of work.

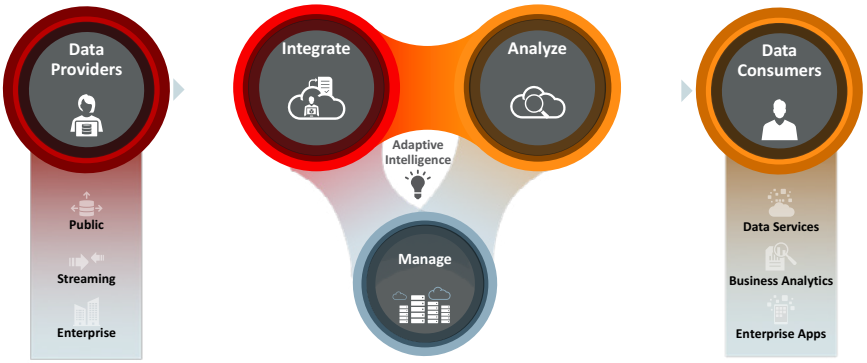
Data scientists spend 50 to 80 percent of their time curating and preparing data before it can actually be used.

Finally, big data technology is changing at a fast pace. A few years ago, Apache Hadoop was the popular technology used to handle big data. That is, until Apache Spark was introduced in 2014. Today, a combination of the two frameworks appears to be the best approach. Keeping up with big data technology is an ongoing challenge.



How Big Data Works

Oracle Cloud for Big Data Analytics



Big data gives you new insights that open up new opportunities and business models. Getting started involves three key actions:

INTEGRATE

Big data brings together data from many disparate sources and applications. Traditional data integration mechanisms, such as ETL (extract, transform, and load) generally aren't up to the task. It requires new strategies and technologies to analyze big data sets at terabyte, or even petabyte, scale. At the same time, big data has the same requirements for quality, governance, and confidence as traditional data sources. During integration, you need to bring in the data, process it, and make sure it's formatted and available in a form that your business analysts can get started with.

MANAGE

Big data requires storage. Your storage solution can be in the cloud, on-premises, or both. You can store your data in any form you want and bring your desired processing requirements and necessary process engines to those data sets on an on-demand basis. Many people choose their storage solution according to where their data is currently residing. The cloud is gradually gaining popularity because it supports your current compute requirements and enables you to spin up resources as needed.

ANALYZE

Your investment in big data pays off when you analyze and act on your data. Get new clarity with a visual analysis of your varied data sets. Explore the data further to make new discoveries. Share your findings with others. Build data models with machine learning and artificial intelligence. Put your data to work.

To help you on your big data journey, we've put together some key best practices for you to keep in mind. Here are our guidelines for building a successful big data foundation.

Big Data Best Practices

#1: ALIGN BIG DATA WITH SPECIFIC BUSINESS GOALS

More extensive data sets enable you to make new discoveries. To that end, it is important to base new investments in skills, organization, or infrastructure with a strong business-driven context to guarantee ongoing project investments and funding. To determine if you are on the right track, ask how big data supports and enables your top business and IT priorities. Examples include understanding how to filter web logs to understand ecommerce behavior, deriving sentiment from social media and customer support interactions, and understanding statistical correlation methods and their relevance for customer, product, manufacturing, and engineering data.

#2: EASE SKILLS SHORTAGE WITH STANDARDS AND GOVERNANCE

One of the biggest obstacles to big data is a skills shortage. You can mitigate this risk by ensuring that big data technologies, considerations, and decisions are added to your IT governance program.

Standardizing your approach will allow you to manage costs and leverage resources. Organizations implementing big data solutions and strategies should assess their skill requirements early and often and should proactively identify any potential skill gaps. These can be addressed by training/cross-training existing resources, hiring new resources, and leveraging consulting firms.

#3: OPTIMIZE KNOWLEDGE TRANSFER WITH A CENTER OF EXCELLENCE

Use a Center of Excellence approach to share knowledge, control oversight, and manage project communications. Whether big data is a new or an expanding investment, the soft and hard costs can be shared across the enterprise. Leveraging this approach can help increase big data capabilities and overall information architecture maturity in a more structured and systematic way.

#4: TOP PAYOFF IS ALIGNING UNSTRUCTURED WITH STRUCTURED DATA

It is certainly valuable to analyze big data on its own. But you can bring even greater business insights by connecting and integrating low density big data with the structured data you are already using today.

Whether you are capturing customer, product, equipment, or environmental big data, the goal is to add more relevant data points to your core master and analytical summaries, leading to better conclusions. For example, there is a difference in distinguishing all customer sentiment from that of only your best customers. Which is why many see big data as an integral extension of their existing business intelligence capabilities, data warehousing platform, and information architecture.

Keep in mind that the big data analytical processes and models can be both human- and machine-based. Big data analytical capabilities include statistics, spatial analysis, semantics, interactive discovery, and visualization. Using analytical models, you can correlate different types and sources of data to make associations and meaningful discoveries.

#5: PLAN YOUR DISCOVERY LAB FOR PERFORMANCE

Discovering meaning in your data is not always straightforward. Sometimes we don't even know what we're looking for. That's expected. Management and IT needs to support this "lack of direction" or "lack of clear requirement."

At the same time, it's important for analysts and data scientists to work closely with the business to understand key business knowledge gaps and requirements. To accommodate the interactive exploration of data and the experimentation of statistical algorithms, you need high-performance work areas. Be sure that sandbox environments have the power they need—and are properly governed.

#6: ALIGN WITH THE CLOUD OPERATING MODEL

Big data processes and users require access to broad array of resources for both iterative experimentation and running production jobs. A big data solution includes all data realms including transactions, master data, reference data, and summarized data. Analytical sandboxes should be created on demand. Resource management is critical to ensure control of the entire data flow including pre- and post-processing, integration, in-database summarization, and analytical modeling. A well-planned private and public cloud provisioning and security strategy plays an integral role in supporting these changing requirements.

Put Big Data To Work

Clearly, big data has tremendous potential. Putting a big data initiative to work at your organization can help you better understand your customers, make more accurate decisions, and create new growth opportunities. Contact us to learn more.

See how Oracle can help your big data journey.

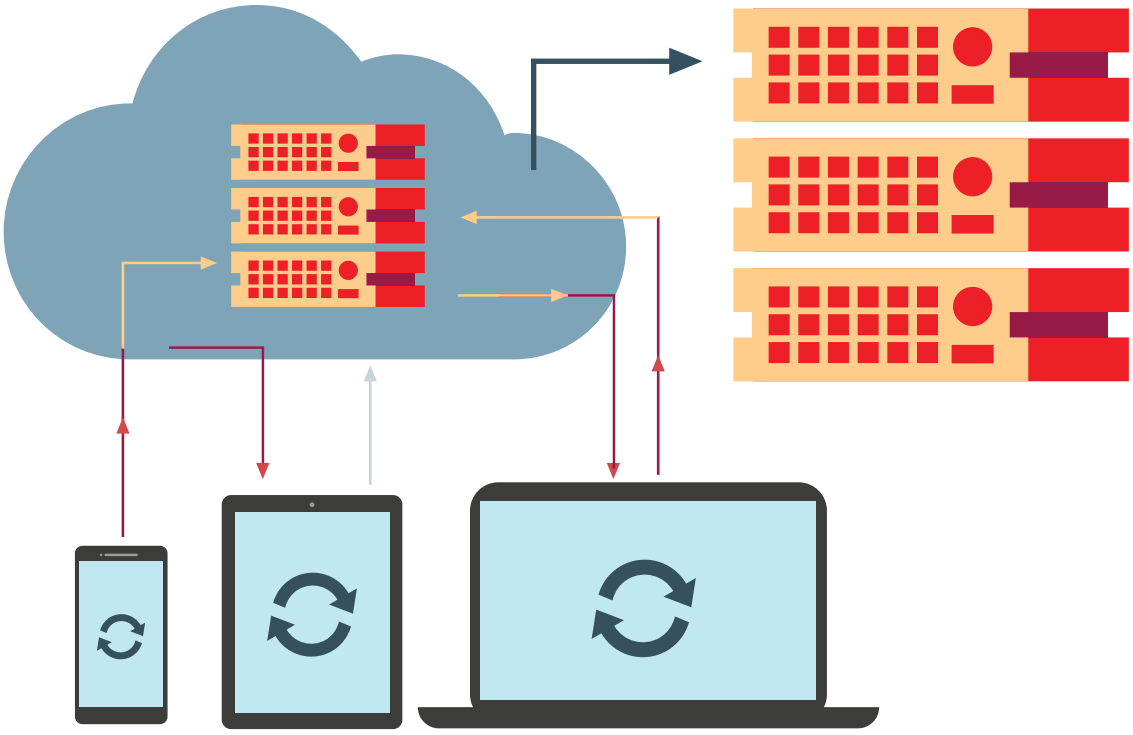
Start your free trial today.

Contact us URL:

https://www.oracle.com/marketingcloud/contact-sales.html

Free Trial URL:

https://go.oracle.com/LP=50758/?



Oracle Corporation

WORLDWIDE HEADQUARTERS

500 Oracle Parkway
Redwood Shores
CA 94065
USA

WORLDWIDE INQUIRIES

Phone: +1.650.506.7000
+1.800.ORACLE1

Fax: +1.650.506.7200

oracle.com

Integrated Cloud Applications & Platform Services

Oracle is committed to developing practices and products that help protect the environment

CONNECT WITH US

ORACLE