

# BIG DATA AND ITS BENEFITS FOR ORGANIZATIONS

# CONTENTS

PREFACE	3
WHAT IS BIG DATA?	3
CHARACTERISTICS OF BIG DATA	4
TYPES OF BIG DATA	5
COLLECTING BIG DATA	6
ORGANIZATION USES OF BIG DATA	7
CHALLENGES OF ANALYZING BIG DATA	8
CONCERNS OVER BIG DATA COLLECTION	9

# PREFACE

The term big data is familiar to anyone with even a passing interest in the world of information technology (IT). Big data is not a new concept. However, recent technological advancements and cultural shifts have contributed to its rising level of prominence in the world of organizations and society. This paper will explore the characteristics of big data, how organizations are using it, and potential concerns about its use.

# WHAT IS BIG DATA?

There are different claims regarding the coinage of the term, but the concept of big data has existed since the mid-20th Century. This activity was before introducing digital storage when the information was contained in physical volumes and stored in libraries. Librarians and custodians of scientific information identified a trend in the quantity of data being generated. The amount of data would double in less than 20 years and eventually threaten to overwhelm the facilities in which it was stored.

Subsequent advances in technology led to the ability to store digitally vast quantities of data in a reduced physical footprint. These advances also led to new methods of rapidly generating more information. These techniques have led to a tremendous increase in the rate of data creation. It is estimated that by 2025 the data stores of the world will contain an incredible 175 zettabytes of data. A zettabyte is a trillion gigabytes, and a gigabyte can hold over 650,000 pages of text or upwards of 15,000 images. It is an almost unimaginable amount of data.

The rise of the Internet of Things (IoT) and the consumer Internet have been significant contributors to the increased rates of information creation. Most of the data stores of the world have been generated in the last few years in terms of sheer volume. This trend shows no signs of slowing down as the percentage of the population with access to the Internet grows, and the use of smart technology increases.

The vast repositories of information resulting from this escalating production make up the entity known as big data. It refers to massive amounts of structured and unstructured data that traditional databases and software approaches cannot process efficiently.

# CHARACTERISTICS OF BIG DATA

Big data is often characterized in the IT community by its ability to be defined using the three Vs first proposed by information analyst Doug Laney in 2001. The three Vs classify big data as data that contains greater variety arriving in increasing volumes and with ever-higher velocity.

- **Volume** - The amount of data is the primary factor describing big data. High volumes of data of potentially unknown value need to be processed when big data is involved. It may require organizations to handle hundreds of terabytes or petabytes of information in various formats. This volume of data overwhelms the traditional means of storage and processing.
- **Velocity** - The speed at which data is generated, received, and possibly acted upon is another characteristic of big data. High-velocity data is collected directly into memory rather than being written to disk. Smart devices often demand real-time data analysis and responses.
- **Variety** - The third principal characteristic of big data is the variety in the available types of information. Structured data types that fit nicely into relational databases have been joined by many streams of data that are unstructured or semi-structured. Big data makes use of information that demands preprocessing to get any meaning or value.

In keeping with the alliterative nature of these characteristics, data professionals have suggested additional Vs to describe big data further.

- **Value** - Effective use of big data has the potential to provide substantial value to an organization. Identifying patterns and trends can cause financial gains and a competitive edge.
- **Veracity** - The quality and trustworthiness of data are critical as it becomes used more extensively for automated decision-making. Ensuring the correctness of its information is a challenge faced by enterprises making use of big data.
- **Variability** - The variable nature of unstructured data requires new methods to be developed that can decipher context and meaning through better natural language processing.
- **Visualization** - It is essential to make the results of big data processing understandable to stakeholders and decision-makers. Visualization is a critical technology in producing value from big data.



# TYPES OF BIG DATA

There are three forms of big data.

**Structured** - Information in a fixed format that enables it to be easily accessed, processed, and stored is structured data. Traditional computer technology, such as relational databases, can process structured data.

**Unstructured** - Data that has no defined form or structure is referred to as unstructured data. It cannot be readily processed or analyzed. Unstructured data may contain text, images, and many other types of items.

**Semi-structured** - Data that is semi-structured exhibits some organizational properties that allow it to be analyzed but cannot be strictly formatted for a relational database. Email is an example of semi-structured data.



# COLLECTING BIG DATA

The characteristics of the variety, volume, and velocity are apparent when considering the way big data is collected. Big data is generated from multiple sources and is gathered at varying intervals and timeframes. Sometimes, deliberate human interaction with software tools or physical collection methods such as customer surveys provides the information. Smart devices and sensors may directly contribute to big data stores or be used for monitoring users to produce a digital representation of their activities and behavior.

Here are some methods most commonly used to collect big data.

**Transactional data** - Point of sale (POS) software combined with a customer relationship management (CRM) tool can create a pool of transactional data. That pool of data can be mined and analyzed. Customer information that can be gathered at each transaction includes:

- What was purchased;
- How much was bought and what promotions influenced their decision;
- When and where the purchase was completed;
- The payment method used.

Over time, profiles can be built up that allow targeted marketing based on purchasing history. And the effectiveness of promotional materials can be evaluated so they can be fine-tuned going forward.

**In-store traffic monitoring** - Tracking customers as they move through a store is possible with motion-sensitive sensors. This data enables merchants to determine which departments or displays are attracting the most attention and can help them adapt their offerings to satisfy customer interest. Electronic monitoring of customers is an example of the IoT helping to gather data on human activities.

**Online marketing analytics** - Analytical engines like Google Analytics provide a wealth of information concerning how customers interact with an online presence of an organization. These tools can help guide marketing campaigns and suggest modifications to webpages to make them more attractive to visitors. Details such as which pages generate the most activity in clicks or interactions can be useful in developing a website tailored to customer demand.

**Social media** - Billions of people use social media networks to interact in a wide variety of ways. Social media analytics can uncover behavioral and demographic information about your current and potential customers. Platform tools such as those on Facebook enable targeted marketing based on the analysis of the unstructured data flowing through social media feeds.

**Customer reward programs** - Organizations that offer discounts or reward programs to repeat shoppers do so primarily to create another data stream on their customers. The reward card that customers swipe at the grocery store gives the customer some immediate savings. And it provides the merchant with information on purchasing habits that can be used for purposes such as creating marketing materials or deciding regarding inventory.

**Satellite imagery** - The inclusion of global positioning systems (GPS) in smartphones and mobile devices enables information to be gathered concerning physical location and movement of customers.

**Employer databases** - Big data can profile employees and evaluate their performance. This information can include identifying which programs they used most often, time of day when activity peaks, and when devices are powered on and off.

**Gameplay** - Game developers can get data that helps them create more engaging products for their customers. Such information includes the time spent with a game, levels that cause difficulty, and the rate of in-app purchases. That information provides game manufacturers with valuable data they can analyze to improve their offering.

## ORGANIZATION USES OF BIG DATA

Big data is currently used in many ways by organizations across all market sectors. Analytics is a critical technique in the processing of big data and helps to provide its benefits. Here are some ways organizations are deriving value from the use of big data.

**Personalizing the customer experience** - Gathering information on specific customers through multiple data streams enables the creation of personalized marketing campaigns. It allows the tailoring of online interaction to conform to individual tastes and preferences.

**Product development** - Big data analytics can anticipate customer demand and proactively bring new offerings to the market. Predictive models take into account past activity, and product attributes to develop new products and services that appeal to the customer base.

**Preventative maintenance** - Maintenance can be performed more effectively by analyzing the structured and unstructured data that is generated regarding equipment and machinery of organizations. Structured data (such as make, model number, and installation date) supplemented by unstructured data (such as problem logs and error reports) allows informed decisions. Such decisions result in maximizing uptime by proactively addressing problems.

**Machine learning** - Enormous quantities of data are essential to successful machine learning (ML) initiatives. Big data provides the raw materials that enable artificial intelligence and ML techniques to be used. These techniques create autonomous machines and robots that promise to change the way organizations and industry operate in the future.



**Innovation** - Big data offers unlimited channels for innovative decision-making throughout an organization. Better conclusions regarding finances, planning, and product focus can be obtained by analyzing the information in big data stores of an enterprise.

**Operational efficiency** - Collecting big data on its internal processes enables an organization to use analytics to increase its operational efficiency. Trends that are not apparent with traditional analysis can be uncovered and result in financial savings and increased productivity.

## CHALLENGES OF ANALYZING BIG DATA

Big data poses challenges to organizations attempting to use it constructively. The significant obstacles are encountered in storing and processing the information effectively. The issues of the volume, velocity, and variety in which big data is generated are reasons that it is hard to use effectively.

Mining the large datasets that are associated with big data requires scalable solutions that can handle the varied nature of the information they are expected to process.

Analytical tools need to provide the visualizations to use the insights uncovered in big data stores productively. They need to offer high performance and enterprise-grade security to ensure that data stays safe and out of the hands of unauthorized entities.

The immense amounts of information comprising big data make it challenging to store using traditional methods. It is often not practical for organizations to address this issue with on-premises data centers. The scalability of cloud storage resources is an attractive alternative and are one way that an enterprise can supplement their storage capacity efficiently.

Devices with embedded intelligence to assist in data mining and analytics promise to help tame some problems of viably storing big data. Edge computing that combines some features of data collection, storage, and processing is becoming more prevalent. It presents another technique with which to address the volume and complexity of big data.



# CONCERNS OVER BIG DATA COLLECTION

Not everyone is enamored with the way companies collect and use big data. The vast amounts of personal information held in big data repositories present an inviting target for hackers intent on compromising it. This fact has resulted in a pushback by consumers who wish to exert control over how their private information is collected and used.

Initiatives like the General Data Protection Regulation (GDPR) of the European Union and similar legislation in other jurisdictions are attempting to address the concerns of citizens over controlling personal data. These regulatory standards enable individuals to opt-out of corporate data collection procedures or have their information removed from enterprise databases. The regulations also put the burden of protecting personally identifiable information (PII) squarely on the organizations collecting and using it.

The responsibility for protecting enterprise data demands a heightened emphasis on security at every stage of the information life cycle. Encryption is one technique that organizations need to implement to secure their data assets. Security concerns also mandate a focus on maintaining compliance with regulatory standards. Failure to show compliance in the wake of a data breach can cause severe financial penalties and negatively impact consumer confidence.

Big data is here to stay, and organizations that use it wisely can attain many tangible competitive advantages over their rivals. It provided benefits that outweigh the additional security and compliance issues that surround its use, collection, and storage. As the volume and rate of data generation continue to increase, new methodologies will be developed that enable organizations to use it more productively. Ignoring the value hidden in this bonanza of information is a risky proposition that should be avoided in the modern world.



# IDERA'S SOLUTION

## Complete SQL Server Management with SQL MANAGEMENT SUITE

SQL Management Suite is a bundle of five essential products for complete SQL Server management. It covers performance, compliance, security, backup, and index fragmentation.

The suite of products includes SQL Diagnostic Manager Pro (with SQL Workload Analysis and SQL Query Tuner), SQL Compliance Manager, SQL Secure, SQL Safe Backup, and SQL Defrag Manager.

Start **FREE** trial

