How to Improve Your Big Data Management

It's not enough to just have big data – it needs to be managed effectively.

Errors stemming from bad data can cost over $15 million per year in losses.[1]

60-73% of all data within the average enterprise is never used for analytical purposes.[2]

94% of companies believe that data and analytics will be essential to their growth.[3]

39% of business domain experts aren't sure what it means to be data-driven.[4]

76% of data experts spend half of their time preparing reports for business teams.[5]

What is Big Data?

Over 2.5 quintillion bytes of data is created by users on the internet every day.[6] That's an example of big data: too much information for traditional systems to handle.

The 5 Vs of Big Data

Volume
It comes in huge quantities.

Velocity
It shows up very quickly, and gets even faster over time.

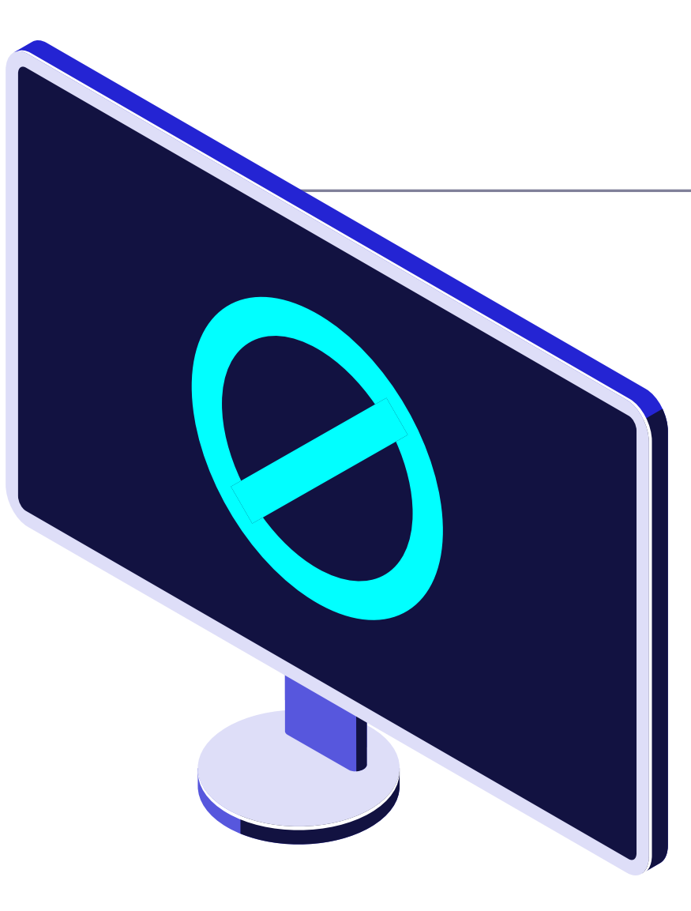
Variety
It comes in a wide variety of forms, often unstructured.

Value
It can be valuable… if you can process it.

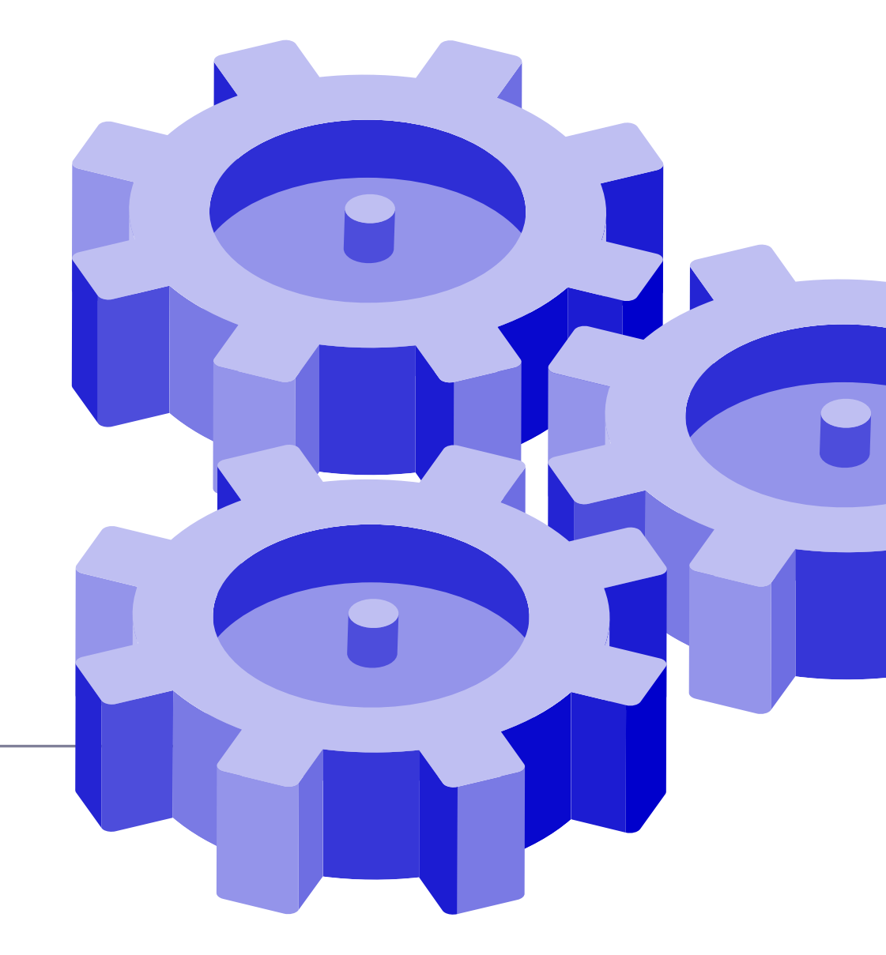
Veracity
Its value depends on whether it can be trusted.

Management Challenges of Big Data

Big data is notoriously difficult to manage without specialized tools and processes. Here's why.

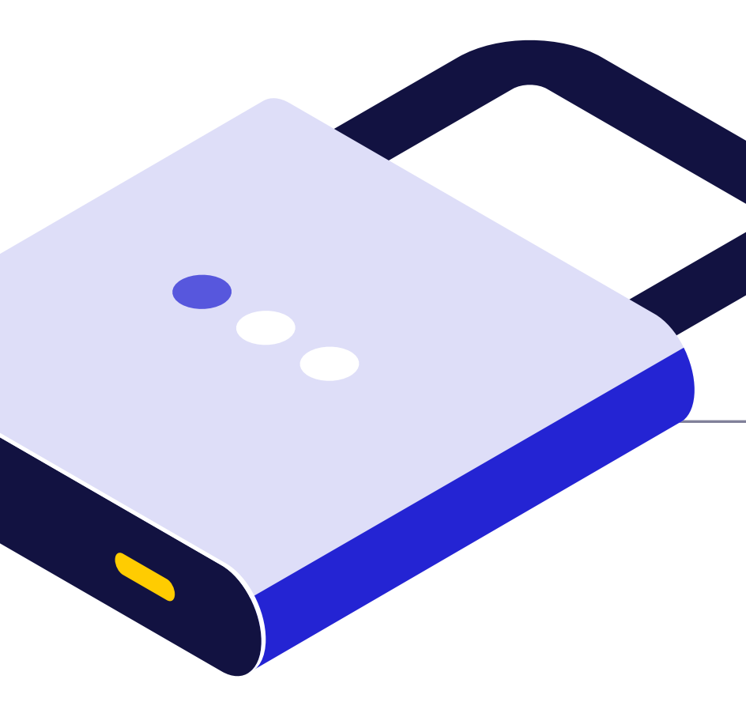


Data quality issues
Incorrect data can lead to costly misinformed decisions.

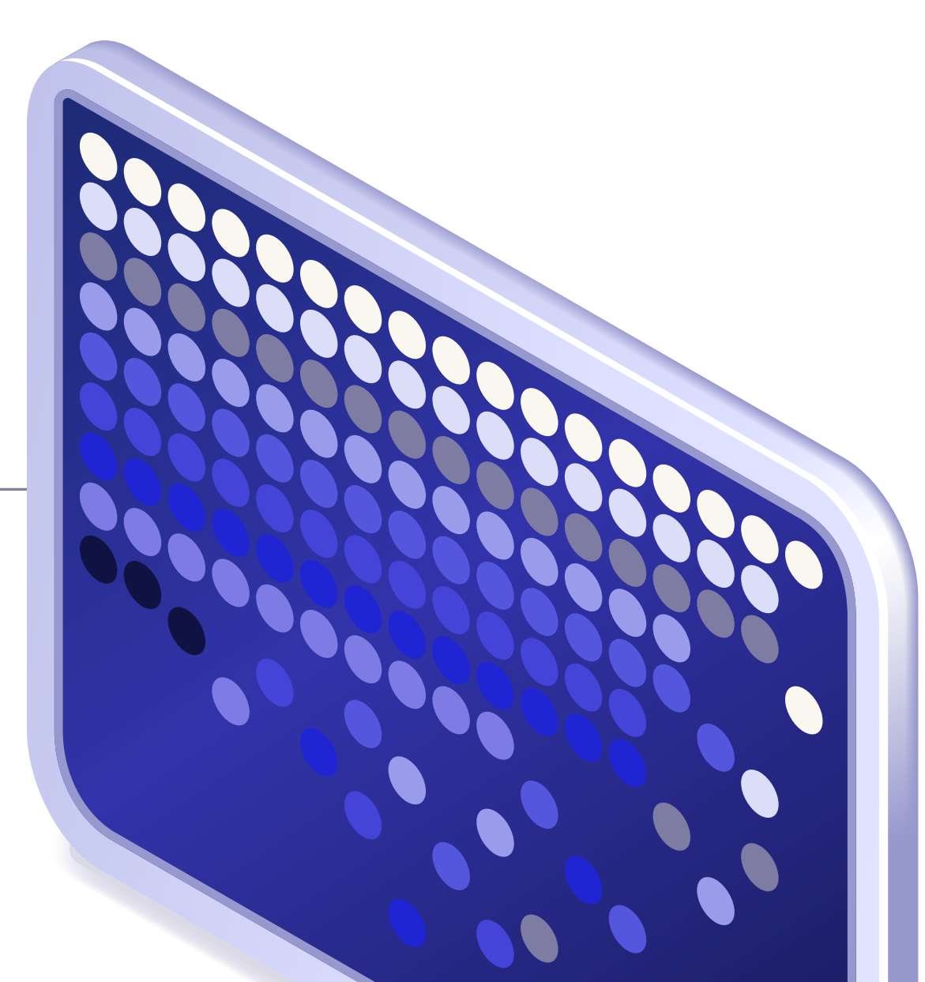


Data integration
Its volume, velocity and variety make it hard to combine into a usable form.

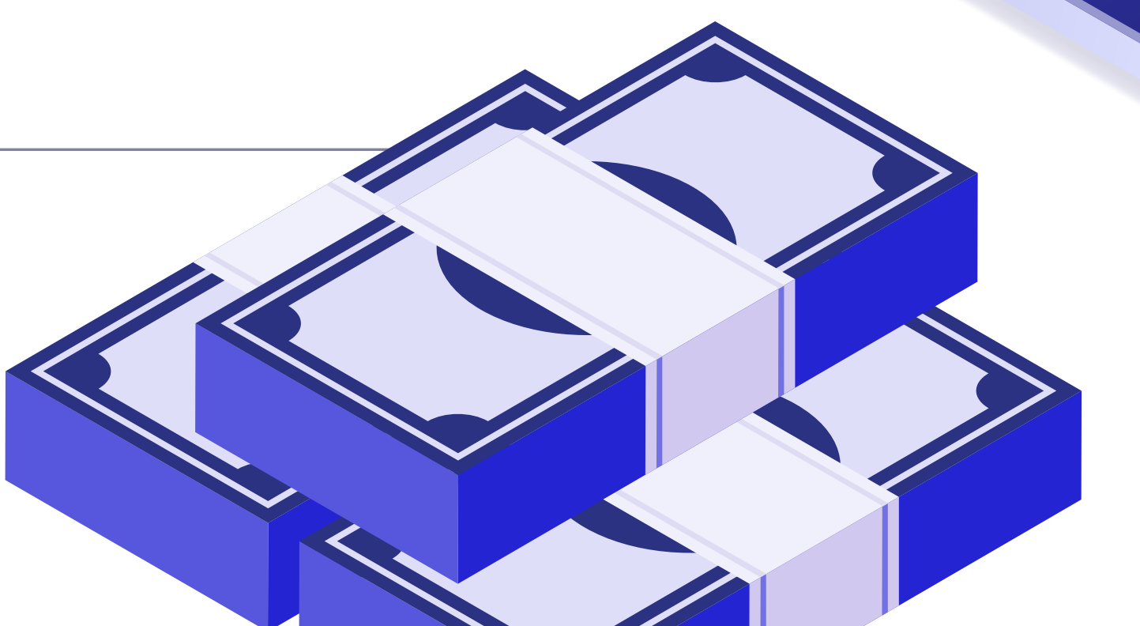
Data storage and processing
Its size and speed put strain on computing resources.



Privacy and security
Its volume and velocity make tracking who has access and how it's used difficult.



Data governance
Bulk data collection makes it very hard to weed out sensitive information to meet compliance requirements and ensure proper use.



Scalability and costs
Storing big data and scaling systems to accommodate it can be difficult and expensive.

80-90% of data we generate is unstructured, which makes it more complicated to manage.[7]

47% of enterprises cite data growth as one of their top three challenges.[8]

Data security concerns and budget constraints are the top two barriers preventing organizations from effectively using big data as an asset.[9]

How to Improve Big Data Management

Big data has led to a greater understanding of customer behavior, trends, and preferences. However, it has also created new challenges in terms of managing and protecting sensitive information.

95% of businesses cite the need to manage unstructured data as a problem for their business.[10]

Data-driven organizations with insights about customers are 23 times more likely to gain more new customers than competing organizations without those insights.[11]

They're also nine times more likely to retain the customers they gain.[12]

Companies making intensive use of customer analytics are 2.6 times more likely to have a significantly higher ROI than competitors that aren't.[13]

Here's How to Manage Big Data Better:

Understand data and business goals.
- Learn where it comes from, how it's generated and what it's used for.
- Put in place the right processes and systems to manage it effectively.
- Ensure that the data is being used to achieve the organization's objectives, not just being collected for the sake of it.

Implement data quality control processes.
- Implement data quality processes like data cleansing, profiling and validation.
- Remove inaccuracies and inconsistencies from data.
- Verify that the data is accurate and complete.

Invest in the right tools and technologies.
- Distributed processing frameworks Hadoop and Spark
- Cluster management software
- NoSQL databases
- SQL query engines
- Data lake and data warehouse platforms
- Cloud object storage services
- Stream processing engines
- Cloud-based master data management solutions

Develop a data governance framework.
- Include policies and procedures for managing big data.
- Define who is responsible for managing the data and how it should be used.
- Set up controls for data security, privacy and compliance.

Why You Should Use Master Data Management for Big Data

Using master data management to handle big data solves problems involving data quality, governance, analytics, costs and more.

Here's how:

Optimize data access
by providing a centralized view of an organization's data.

Create a single source of truth
to integrate big data and share it easily.

Identify entities in real time
and resolve data conflicts using a reference database.

Improve data quality
by creating more complete, accurate and timely customer profiles, removing duplicates and grouping similar records.

Improve governance and privacy
by dynamically masking sensitive data.

---

[1] https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/

[2] https://www.forrester.com/blogs/hadoop-is-datas-darling-for-a-reason/

[3] https://www.microstrategy.com/content/dam/website-assets/collateral/financial-documents/press-release-archive/MicroStrategy%20Global%20Analytics%20Study%20Finds%2097%20of%20Real-time%20Enterprise%20Decisions%20are%20Data-deprived.pdf

[4] https://www.sigmacomputing.com/resources/data-language-barrier/

[5] Ibid.

[6] https://www.domo.com/solution/data-never-sleeps-6

[7] https://www.cio.com/article/220347/ai-unleashes-the-power-of-unstructured-data.html

[8] https://www.sagiss.com/blog/data-storage-38-interesting-facts-and-figures

[9] https://www.capgemini.com/wp-content/uploads/2017/07/the_big_data_payoff-turning_big_data_into_business_value.pdf

[10] https://www.forbes.com/sites/rkulkarni/2019/02/07/big-data-goes-big/

[11] https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/five-facts-how-customer-analytics-boosts-corporate-performance

[12] Ibid.

[13] Ibid.