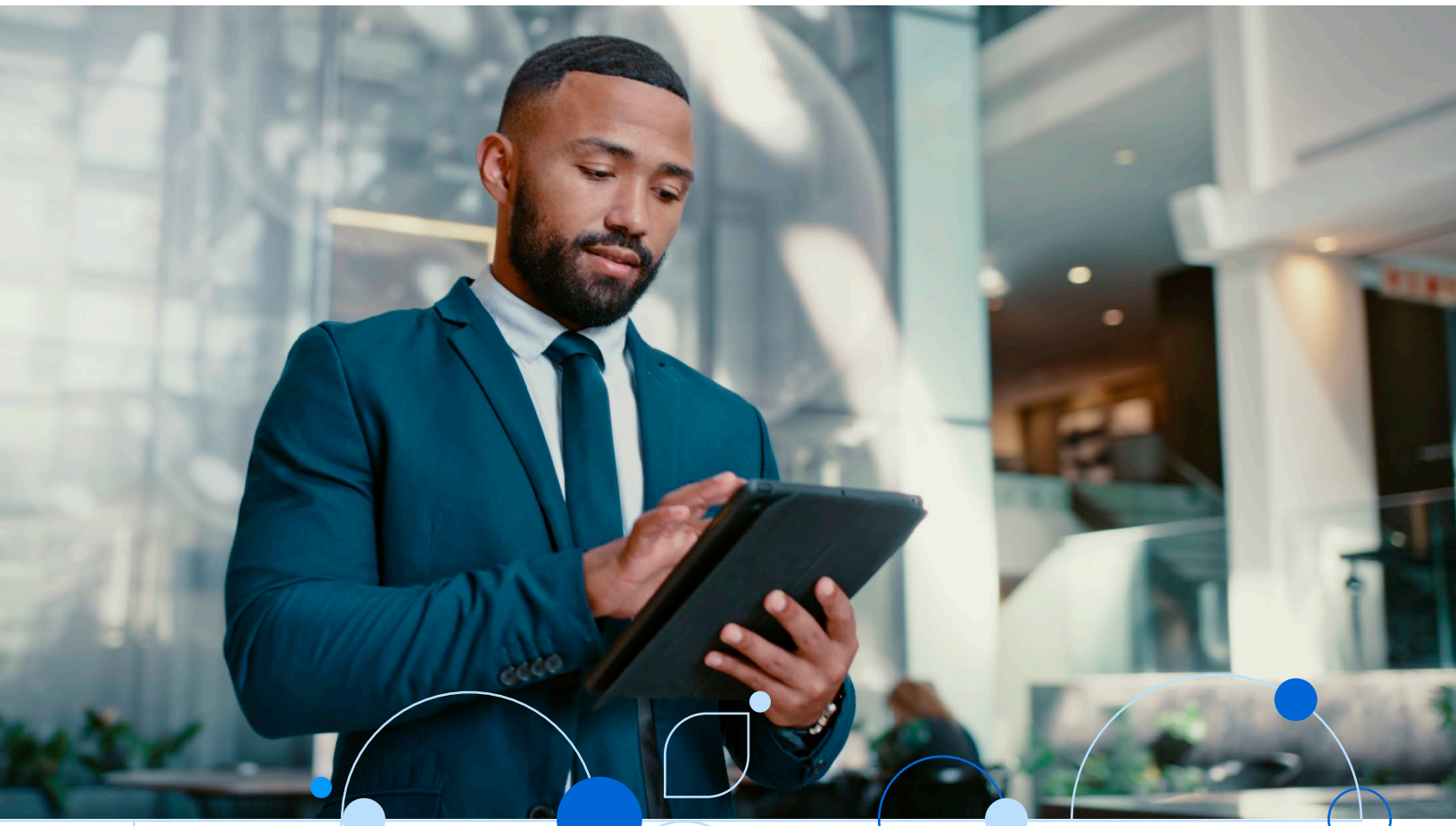
WHITE PAPER

The data management path to AI and analytics success

Maximize the business value of AI with modern data management



Contents

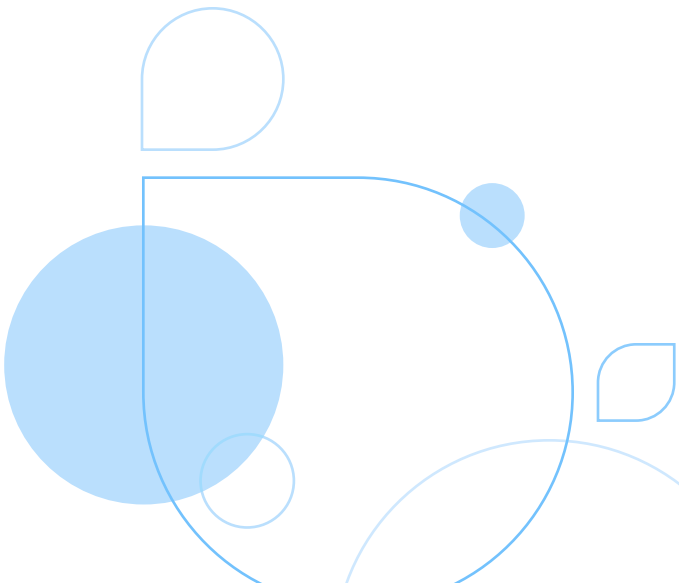
Introduction 1

Data, data everywhere, but not an insight to be found 1

Let's be honest with ourselves 1

The role of DataOps and AIOps in becoming a data-driven organization 2

Data management requirements for success in DataOps and AIOps 4



Introduction

The possibilities for using your organization's data for positive gain are enormous: better decision making, greater profitability, increased efficiency, more effective collaboration between departments and more. And data analytics and AI can help reach all those goals and more at a record pace.

But for far too many enterprises, analytics and the strategic promise of AI are just boardroom buzzwords. While the C-suite may understand the promise and possibilities of both, their organization may not have the clean, structured data that can get them there.

If analytics and AI are to power business goals, your infrastructure needs to be built on a solid data foundation.

Data, data everywhere, but not an insight to be found

Every organization today has more data than ever before – more than they know what to do with – housed in on-premises data centers, cloud-based data solutions, data lakes and more, most segregated, fractured and incomprehensible to any strategy for making gains against the data goals of the 2010s, much less the AI-driven possibilities of today.

That's where data management – the process of discovering, preparing and governing data to make it ready for analysis – brings new AI and analytics possibilities into reach.

According to a **2016 New York Times article by Steve Lohr**, data scientists at the time had to spend 50-80% of their time collecting, cleaning and preparing data. The good news today is that AI and generative AI can take over many of the mundane tasks data scientists once had to do themselves: finding patterns, generating code, etc.

According to a 2016 New York Times article by Steve Lohr, data scientists at the time had to spend 50-80% of their time collecting, cleaning and preparing data.

It's essential for organizations to monitor their data ecosystem as a living, breathing thing. That requires comprehensive data governance and data management practices that treat data appropriately, know its provenance and make sure it is validated and fit for purpose.

The old data management adage garbage in, garbage out applies now more than ever. For enterprises to succeed with the power and promise of AI, it's critical that they start with data management.

Let's be honest with ourselves

If you're spending more time asking questions about the state of your data than using it to move your organization forward, you're not ready to move forward, certainly not with AI.

Here are a few simple questions to ask yourself that a modern data management solution would make very easy to answer:

- Do you know where all your data is? In some situations, you might be given a job ticket that specifies the data sources you should use. In others, you'll need to discover what's relevant. That's where solutions such as SAS® Information Catalog – which

automatically searches all your data sources – can help. You'll have a single place to go for everything you need, helping you uncover data you didn't know existed and greatly improving your analytics.

- Are you preparing data for its intended use? Does it need to run through a data preparation process to standardize different tokens, such as name, company and email? Do some fields need to be parsed? Do different elements need to be validated? Should you worry about duplicate records? If you're sending out a billing statement or a credit card offer, you certainly need to suppress duplicates. But if you're looking for fraud, you may need to identify but not suppress duplicates.
- Once you've identified relevant data elements, are they all fit for use? Do you need to run checks to identify data that may not fit your needs? For example, if you are using a birth date as one of the qualifications for your model, you may need to exclude records outside acceptable ages.
- If your data is cleaned, standardized and fit for use, what business decisions will you be driving with it? If your data is being used for analytics or visualization, can you write it to a table, database, data mart or other storage mechanism appropriately?
- What about privacy protections? If you're keeping data for internal purposes, do you need to apply governance policies – such as GDPR, HIPAA, CCPA or others – to make sure personal information is protected?
- What metadata is collected and saved? Is your metadata discoverable so that other users can use new data tables?

The role of DataOps and AIOps in becoming a data-driven organization

If you're going to optimize data-driven decision making and deliver faster value, every business user of data – whether in marketing, accounting, operations or other departments – needs easier access to their data without intervention from data managers.

That's why you need a data operations (DataOps) program.

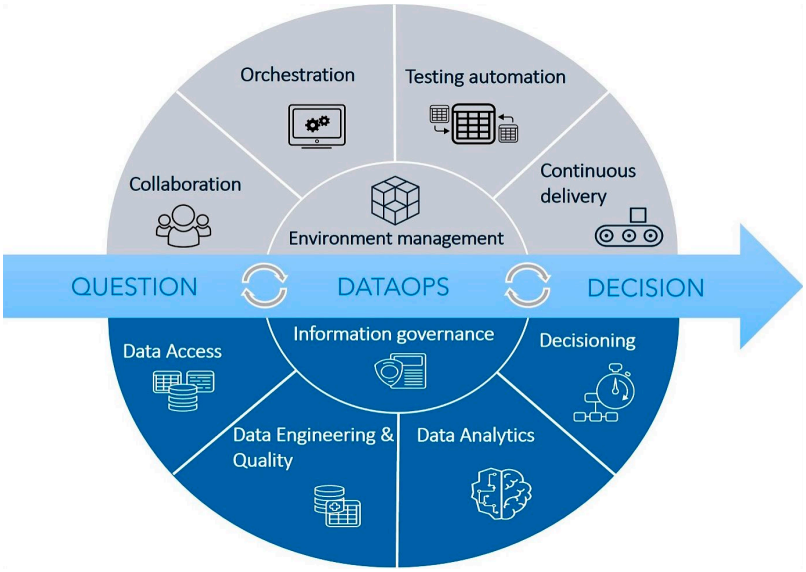
According to Gartner, **DataOps** is "a collaborative data management practice focused on improving the communication, integration and automation of data flows between data managers and users across an organization."

A key to DataOps is treating data as a product that meets specific needs. That way, enterprises can deliver data and improve productivity via end-to-end automation and process integration with minimal manual intervention.

As Gartner predicts, "By 2025, a data engineering team guided by DataOps practices and tools will be 10 times more productive than teams that do not use DataOps."

No matter the size of your enterprise, that kind of potential productivity gain is impossible to ignore.

According to Gartner, DataOps is "a collaborative data management practice focused on improving the communication, integration and automation of data flows between data managers and users across an organization."

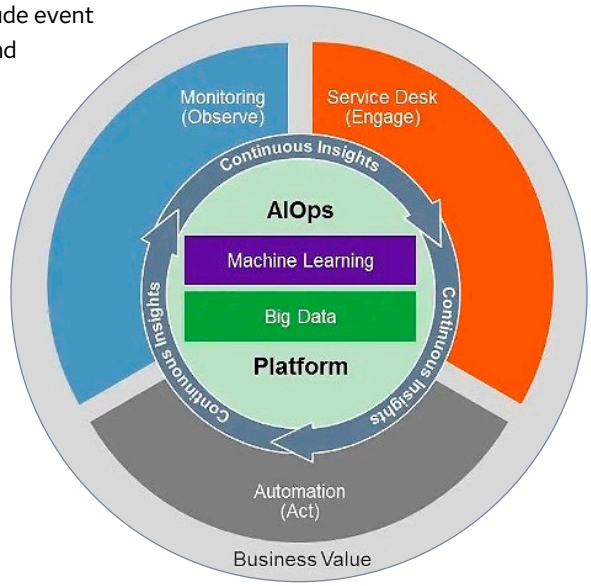


Source: SAS

AI operations (AIOps) help enterprises successfully implement AI solutions by combining big data and machine learning to automate IT operations processes. Its goal is to connect data across different environments so data scientists can put AI principles into practice.

According to Gartner, AIOps include event correlation, anomaly detection and causality determination. Among large corporations, Gartner predicts that exclusive use of AIOps and digital experience tools to monitor applications and infrastructure will rise from 5% in 2018 to 30% in the next few years.

Source: Gartner



Gartner predicts that exclusive use of AIOps and digital experience tools to monitor applications and infrastructure will rise from 5% in 2018 to 30% in the next few years.

Data management requirements for success in DataOps and AIOps

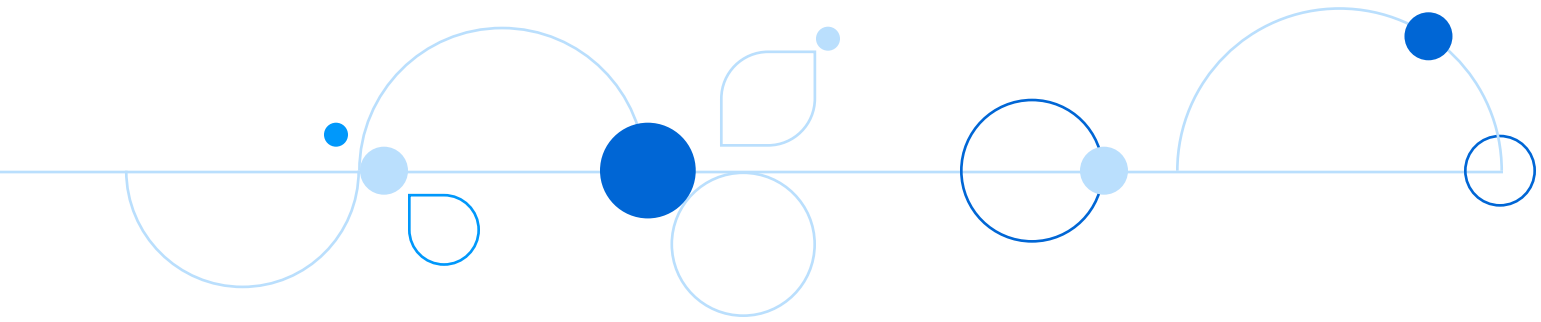
Data management is an all-encompassing term that includes several steps. Below, we've also included some important terms used in data management. Here are a few terms you'll hear on your data-driven journey to AI.

- **Data governance.** More than just a set of business rules that feed a dashboard, data governance provides:
  - **Data discovery.** Find the data you need from inside your organization.
  - **Data protection.** Where is the personally identifiable information (PII) in the organization? A sound data governance program identifies PII data and establishes policies for its sharing, use and protection.
  - **Data remediation.** Sometimes you may need to quarantine questionable data until it can be validated and then released into enterprise data warehouses, data marts or data lakes.
- **Data quality.** It's critical to parse, standardize, validate, enrich and reduce duplicate data in all your enterprise systems. It's also essential that all your quality processes follow the same core rules and processes, whether it's accounting, operations, marketing or other departments. To have real value, data must be consistent across the organization.
- **Data access.** Far too often, organizations aren't concerned about AI, mobile data or other advanced data management capabilities because they're still trying to figure out how to connect to data in a consistent, trusted fashion. Whether data is in a flat file, spreadsheet, relational database, blob storage or any other data storage type, it needs to be accessible to those who need it when they need it.
- **Real-time data capture.** Whether it's for medical devices, heavy machinery or other equipment, it's important to capture every device's data to better predict maintenance needs and schedule downtime appropriately.
- **Self-service applications.** Allow non-programmers to point and click their way to successfully implementing data management or analytics processes.

Ready to learn more?

While all this can sound very complicated, it doesn't have to be. The first step is working with the right solutions to get your data in order.

Learn more about how SAS can help at **Data Management Software | SAS.**



Learn more about how SAS can help with data management.