ما معنى داتا أوبس؟

داتا أوبس (DataOps): اختصار لمصطلح "عمليات البيانات" (Data Operations)، وهي منهجية آلية رشيقة، تستخدمها فرق تحليل البيانات ومعالجتها، لتصميم بنية البيانات الموزعة وصيانتها، وذلك بهدف تحسين الجودة، وخلق قيمة من البيانات الضخمة.

قُدّم مفهوم داتا أوبس لأول مرة بواسطة المحرر "ليني لايبمان" (Lenny Liebmann)، في منشور على مدونة "آي بي إم" (IBM) بعنوان "ثلاثة أسباب تجعل داتا أوبس ضرورية لنجاح البيانات الضخمة" في عام 2014.

آلية عمل داتا أوبس

تستخدم داتا أوبس التحكم الإحصائي في العملية، لمراقبة مسار تدفق تحليلات البيانات، والتحكم فيه، إذ تتم مراقبة البيانات المتدفقة من خلال نظام التشغيل باستمرار، والتحقق من عملها، لتدارك أي مشكلة تطرأ أثناء العملية، وإخطار فريق تحليلات البيانات من خلال تنبيه آلي. كما تدمج داتا أوبس منهجية "أجايل" (Agile)، لتقصير وقت دورة تطوير التحليلات بما يتماشى مع أهداف العمل.

أهمية داتا أوبس

تساهم داتا أوبس في تبسيط عملية تصميم التطبيقات، وتطويرها، وصيانتها، استناداً إلى تحليلات البيانات، كما تهدف إلى تحسين طريقة إدارة البيانات، وإنشاء المنتجات، وتنسيق هذه التحسينات مع أهداف العمل.

مزايا داتا أوبس

تمتاز داتا أوبس بالعديد من الخصائص، فهي تعدّ ابتكاراً تجريبياً سريعاً، يقدم رؤى جديدة للعملاء عن المنتجات والتطبيقات، وتمتاز بجودة بيانات عالية، ومعدلات خطأ منخفضة جداً، كما تتيح التعاون بين مصفوفات معقدة من الأشخاص والتكنولوجيا والبيئات، وقياس شفافية نتائج تحليلات البيانات بوضوح.

# DataOps (data operations)

* [**Jack Vaughan,**](https://www.techtarget.com/contributor/Jack-Vaughan)Senior News Writer

[DataOps](https://www.techtarget.com/searchdatamanagement/news/252494497/Quest-Software-adds-data-governance-and-DataOps-with-Erwin) (data operations) is an Agile approach to designing, implementing and maintaining a distributed data architecture that will support a wide range of open source tools and frameworks in production. The goal of DataOps is to [create business value](https://www.techtarget.com/searchdatamanagement/feature/Using-DataOps-to-create-business-value-from-big-data) from big data.

Inspired by the [DevOps](https://www.techtarget.com/searchitoperations/definition/DevOps) movement, the DataOps strategy strives to speed the production of applications running on big data processing frameworks. Additionally, DataOps seeks to break down silos across [IT operations](https://www.techtarget.com/searchitoperations/definition/IT-operations) and software development teams, encouraging [line-of-business](https://www.techtarget.com/searchcio/definition/LOB) stakeholders to also work with data engineers, [data scientists](https://www.techtarget.com/searchenterpriseai/definition/data-scientist) and analysts. This helps to ensure that the organization’s data can be used in the most flexible, effective manner possible to achieve positive business outcomes.

Since it incorporates so many elements from the [data lifecycle](https://www.techtarget.com/searchstorage/definition/data-life-cycle-management), DataOps spans a number of information technology disciplines, including data development, [data transformation](https://www.techtarget.com/searchdatamanagement/definition/data-transformation), data extraction, data quality, data governance, data access control, data center capacity planning and system operations. DataOps teams are often managed by an organization’s chief data scientist or chief analytics officer and supported by employees like [data engineers](https://www.techtarget.com/searchdatamanagement/definition/data-engineer) or data analysts.

As with DevOps, there are no “DataOps” specific software tools; there are only frameworks and related toolsets that support a DataOps approach to collaboration and increased agility. Such tools include [ETL](https://www.techtarget.com/searchdatamanagement/definition/Extract-Load-Transform-ELT)/[ELT](https://www.techtarget.com/searchdatamanagement/definition/Extract-Load-Transform-ELT) tools, [data curation](https://whatis.techtarget.com/definition/data-curation) and cataloging tools, log analyzers and systems monitors. Tools that support microservices architectures, as well as open source software that lets applications blend structured and unstructured data, are also associated with the DataOps movement. Such software can include MapReduce, HDFS, Kafka, Hive and Spark.

### **How DataOps works**

The goal of DataOps is to combine DevOps and Agile methodologies to manage data in alignment with business goals. For example, if the goal is to raise lead conversion rate, DataOps would position data to make recommendations for marketing products better, thus converting more leads. Agile processes are used for [data governance](https://www.techtarget.com/searchdatamanagement/definition/data-governance) and analytics development while DevOps processes are used for optimizing code, product builds and delivery.

Building new code is only one part of DataOps as streamlining and improving the [data warehouse](https://www.techtarget.com/searchdatamanagement/definition/data-warehouse) is equally as important. Similar to the process of [lean manufacturing](https://www.techtarget.com/searcherp/definition/lean-production), DataOps uses statistical process control (SPC) to monitor and verify the data analytics pipeline consistently. SPC makes sure that statistics remain within feasible ranges, advances data processing efficiency and raises data quality. If an anomaly or error occurs, SPC helps to alert data analysts immediately for a response.

### **How to implement DataOps**

As the volume of data is estimated to continue to grow exponentially, implementing a DataOps strategy has become crucial. The first step to DataOps involves cleaning [raw data](https://www.techtarget.com/searchdatamanagement/definition/raw-data) and developing an infrastructure that makes it readily available for use, typically in a self-service model. Once data is made accessible, software, platforms and tools should be developed or deployed that orchestrate data and integrate with current systems. These components will then continuously process new data, monitor performance and produce real-time insights.

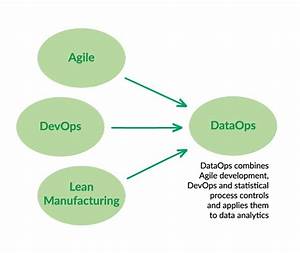
A few best practices associated with implementing a DataOps strategy include:

* Establish progress benchmarks and performance measurements at every stage of the data lifecycle.
* Define semantic rules for data and metadata early on.
* Incorporate [feedback loops](https://www.techtarget.com/searchitchannel/definition/feedback-loop) to validate the data.
* Use [data science](https://www.techtarget.com/searchenterpriseai/definition/data-science) tools and [business intelligence](https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-BI) data platforms to automate as much of the process as possible.
* Optimize processes for dealing with [bottlenecks](https://www.techtarget.com/searchnetworking/definition/bottleneck) and [data silos](https://www.techtarget.com/searchdatamanagement/definition/data-silo); this typically involves software automation of some sort.
* Design for growth, evolution and scalability.
* Use disposable environments that mimic the real production environment for experimentation.
* Create a DataOps team with a variety of technical skills and backgrounds.
* Treat DataOps like lean manufacturing by focusing on continuous improvements to efficiency.

### **Benefits of DataOps**

Transitioning to a DataOps strategy can bring an organization the following benefits:

* Provides real-time data insights.
* Reduces cycle time of data science applications.
* Enables better communication and collaboration between teams and team members.
* Increases transparency by using data analytics to predict all possible scenarios.
* Processes are built to be reproducible and reuse code whenever possible.
* Ensures higher [data quality](https://www.techtarget.com/searchdatamanagement/definition/data-quality).
* Creates a unified, interoperable data hub.



# **What is DataOps?**

Most businesses collect data but are unable to use it to generate business value or deliver insights in a timely fashion. Data volume and data types continue to grow, as do the different types of data citizens—ranging from business users to data scientists. As a result, data management and delivery often become critical bottlenecks. Enter [DataOps](https://www.ibm.com/analytics/dataops).

DataOps (data operations) refers to practices that bring speed and agility to end-to-end data pipelines process, from collection to delivery. The term DataOps and related concepts are at early stages of awareness and adoption, so many working definitions exist today. Research leaders, like Gartner and MIT, have focused their definitions around improving communication between data stakeholders and implementing automation within data flows and lifecycles to enhance delivery practices. Others are simply describing it as “[DevOps for data](https://www.ibm.com/analytics/dataops).”

IBM defines [DataOps](https://www.ibm.com/analytics/dataops) as the orchestration of people, process, and technology to deliver trusted, [high-quality data](https://www.ibm.com/analytics/data-quality) to data citizens fast. The practice is focused on enabling collaboration across an organization to drive agility, speed, and new data initiatives at scale. Using the power of [automation](https://www.ibm.com/analytics/dataops), DataOps is designed to solve challenges associated with inefficiencies in accessing, preparing, integrating and making data available.

It is equally important to know what it is not. DataOps is not: a product; a single event or step; a specific team or person. As a rule of thumb, DataOps methodology or practices you implement should consider interaction between these aspects:

### **People and Process**

DataOps supports highly productive teams with automation technology to deliver huge efficiency gains in project outputs and time. However, to experience the benefits, the internal culture needs to evolve to truly be data-driven. With more business segments requiring and wanting to manage data to drive contextual insights, the time is right to 1) increase the quality and speed of data flowing to the organization and 2) get commitment from leadership to support and sustain a data-driven vision across the business.

This type of transformational change begins by understanding the true goals of the business. How does data inform the decisions and services impacting customers? How can data help maintain a competitive advantage in the market? What are the revenue priorities that data can help us solve?

DataOps leaders will need to align business goals to any pilot project deliverables to demonstrate the linkage between executive stakeholders and the ability to exhibit quick, tangible results. They will also need to define the roles all data citizens play to drive the culture and DataOps practice forward. Each organization has unique needs where stakeholders in IT, data science, the lines of business, and everyone in between need to add value to drive success. What roles each play for your business requires deep collaboration across all functions and commitment to sustainability of a practice.

### **Technology**

Tooling is necessary to support any practice that relies on automation. At the core of DataOps is your organization’s [information architecture](https://www.ibm.com/analytics/dataops). Do you know your data? Do you trust your data? Are you able to quickly detect errors? Can you make changes incrementally without “breaking” your entire data pipeline? If you are unsure how to answer these questions, the first step is to take inventory of your [data governance](https://www.ibm.com/analytics/data-governance) and[data integration](https://www.ibm.com/analytics/data-integration)tools and practices.

As you consider tooling to support a DataOps practice within your businesses, think about how automation in the five critical areas below can transform your data pipeline:

1. Data curation services
2. [Metadata management](https://www.ibm.com/analytics/data-governance)
3. Data governance
4. [Master data management](https://www.ibm.com/analytics/master-data-management)
5. Self-service interaction

Implementing tools can be a tangible way to show progress in adoption of DataOps, but doing so requires a holistic vision. Companies that focus on one element at the expense of others are unlikely to realize the benefits from implementing DataOps practices. The technology conversation and implementation should not live siloed from the ongoing planning regarding people and process. The tooling lives to support and sustain the culture.

### **Continue learning about the IBM DataOps Program**

The shift to adopt DataOps is real. According to a recent survey, [73 percent of companies plan to invest in DataOps.](https://www.nexla.com/data-operations-survey-2018/) IBM is here to help you on your path to a [DataOps practice](https://www.ibm.com/analytics/dataops) with a prescriptive methodology, leading technology, and the [IBM DataOps Center of Excellence](mailto:dataops@us.ibm.com?subject=I%27m%20interested%20in%20an%20IBM%20DataOps%20Garage%20Workshop), where experts work with you to customize an approach based on your business goals and identify the right pilot projects to drive value for your executive team.

Accelerate your DataOps learning and dive into the methodology by reading the whitepaper [Implementing DataOps to deliver a business-ready data pipeline](https://www.ibm.com/account/reg/signup?formid=urx-40583).