

# Data Analysis with DevOps: DataOps

---

## Abstract

Understanding data can be tricky and sometimes overwhelming. Using automation helps a data analysis team improve their methods more consistently. Specifically, automating software tasks is the best way to manage the many repetitions and repetitive jobs that come with thorough data analysis. When starting a project that involves software, automation, and working together, DevOps is the best way to go. DevOps improves all parts of the data analysis process, allowing teams to automatically handle software-related tasks in data analysis and work well with everyone involved in the project. The combination of DevOps, Data Analytics, and Data Engineering introduced the new DataOps which is used for data processing and analytics. DataOps improves communication, collaboration, and faster data pipeline.

**Keywords:** DevOps; Data Analysis; DataOps; CI/CD Pipeline.

## Authors

### **Amita Pandit**

Rabindranath Tagore University, Computer Science & Engineering  
Bhopal, India  
Amita14.pandit@gmail.com

### **Rakesh Kumar**

Rabindranath Tagore University, Computer Science & Engineering  
Bhopal, India  
rakeshmittan@gmail.com

### **Priti Maheswary**

Rabindranath Tagore University, Computer Science & Engineering  
Bhopal, India  
pritimaheswary@gmail.com

## **I. Introduction**

The merging of development and operations, or "DevOps," increases an organization's capacity to deploy applications and improve products more quickly than those utilizing the conventional software development method.

Cloud-centricity is evolving in DevOps automation. The majority of public and private cloud computing companies support DevOps on their platforms consistently, including tools for continuous integration and continuous development. The transition from waterfall to DevOps in modern software development has been a long one. Continuous Integration, Continuous Testing, Continuous Delivery, and Continuous Deployment are all components of this journey. IT organizations began to recognize how much more modern methods are superior to conventional methods in every way. DevOps is a recent idea that has become highly popular in the software business, but it still has certain difficulties when it comes to continuous delivery across numerous environments.

Continuous integration, a key DevOps technique in software development, involves developers constantly integrating their updated code into a common repository, which is followed by automated builds and testing. Continuous testing is the process of executing automated test cases throughout the pipeline of software development to get feedback or a reaction on the business risks. Continuous Delivery is a method of developing software over time. Using this method, the team produces software in tiny batches to guarantee its dependability for release at any moment. With the use of automated technologies, it aims to develop, test, and release software with enhancements to the program code and user environments. Continuous Deployment is a software release strategy used in DevOps to ensure that updated code is accurate and reliable for future usage [2][3].

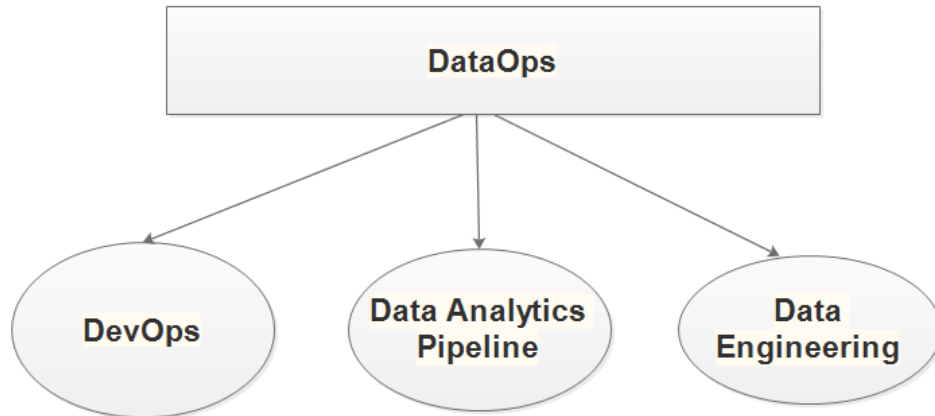
The data retrieval phase of the data analysis process is significantly enhanced by DevOps and automation. The analytic team will have a continually expanding data set after the data retrieval procedure is automated, which will result in better findings.

DataOps is a cooperative data management practice that aims to enhance data flows between data managers and data consumers across an organization by integrating and automating data flows, increasing communication between them, and enhancing integration. DataOps aims to accelerate value delivery by establishing predictable delivery and change management of data, data models, and related artifacts. DataOps employs technology to automate the design, deployment, and administration of data distribution with the proper levels of control, and it leverages metadata to increase the usability and value of data in a dynamic context.

## **II. DataOps Terminology**

### **1. DataOps**

The goal of the relatively new technique known as DataOps is to create high-quality data products as quickly as possible by combining data engineering, data analytics, and DevOps. It uses Agile methodologies to speed up business outcomes and covers the complete data analytics lifecycle, from data extraction through visualization and reporting [20].



**Figure 1:**

The success of DevOps, which enables businesses to shorten software release cycles from months to hours and minutes, served as inspiration for DataOps. DataOps is the combination of DevOps, Data Analytics pipeline automation, and Data Engineering which is shown in Figure 1. It enhances the data quality and increases value generation in data-intensive initiatives to increase confidence in analytics.

## 2. DataOps Process

**Figure 2 describes the detailed process of DataOps. Data Analytics provided the following key stages:**

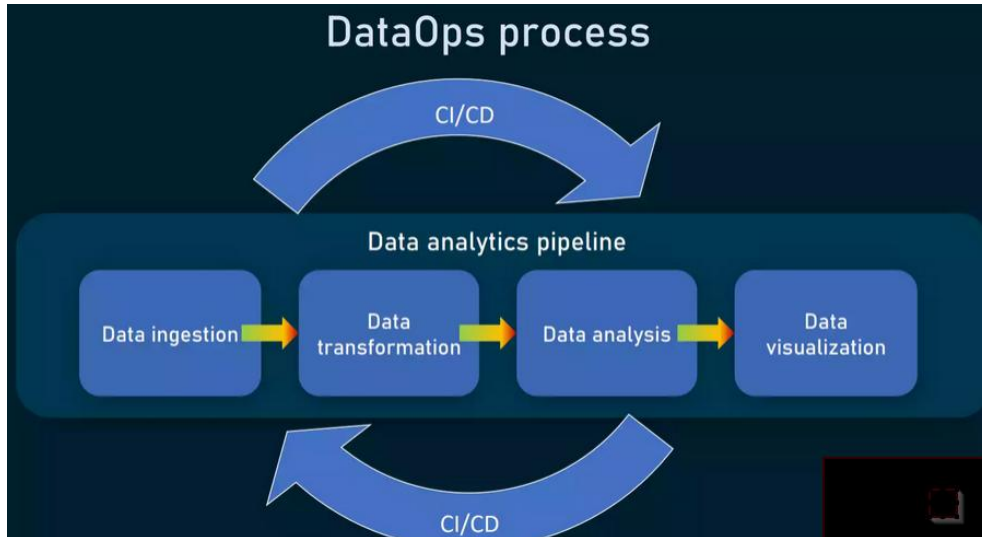
Although data analytics pipelines can consist of hundreds of operations, they may be divided into four main phases [20].

**Data Ingestion.** Data is examined, verified, and put into a downstream system after being retrieved from numerous sources.

**Data Transformation:** Data is supplemented and cleaned. The initial data models are created to satisfy business requirements.

**Data Analysis:** Data teams may realize at this point that further data collection is necessary in order to reach reliable results: Otherwise, they provide insights utilizing various data analysis methods.

**Data Visualization:** Reports or interactive dashboards are used to convey data insights.



**Figure 2:** DataOps Process

**CI/CD (DevOps) for Data Operation**

DevOps is the combination of the Development phase and Operation phase. Table 1 describes the different phases [19].

**Table 1: DevOps Phases**

DevOps Development Phases		
Sl. no	Phase	Feature
1	Continuous Development (CD)	This approach encompasses the initial stages of the DevOps lifecycle, encompassing planning and coding. Techniques such as version control may be employed. It encompasses the pre-release activities of software development, including planning, coding, building, and testing.
2	Continuous Integration (CI)	This approach to software engineering involves systematically integrating its various elements to construct software. Its purpose is to prevent disruptions in builds or other problems resulting from changes made to the source code.
3	Continuous Testing (CT)	Within this phase of the DevOps lifecycle, automated and well-structured code testing takes place as the application code is developed or altered. These tests can be generated manually or through continuous automation tools.
4	Continuous Deployment (CD)	This stage of the DevOps lifecycle revolves around uninterrupted deployment to guarantee

		that any alterations to the code won't adversely affect the performance of a live website.
<b>DevOps Operation Phases</b>		
<b>Sl. no</b>	<b>Phase</b>	<b>Feature</b>
1	Continuous Feedback	Ongoing Feedback in the Operational Stage: Similar to continuous feedback, this phase in the DevOps process involves the automated transmission of data concerning user issues and system performance. It offers customers the opportunity to voice their opinions and share their experiences.
2	Continuous Monitoring	Persistent Monitoring in the Operational Stage: During this stage, developers collect data, closely monitor each function, and identify issues such as low memory or a disrupted server connection. For instance, if users encounter difficulties accessing their accounts post-login, your software may be encountering problems.
3	Continuous Operation	Uninterrupted Operation in the Operational Stage: This is the final, simplest, and briefest step within the DevOps framework. It also incorporates automating the application's delivery and any associated updates, streamlining cycle times, and allowing developers and providers more space to focus on innovation.

### 3. Benefits of DataOps

- Decrease errors
- Enhance teamwork among teams
- Raise the general efficacy of data management tactics.
- Enhancing data manager-to-data manager communication, integration, and automation of data flows
- Data consumers across an organization.

## References

- [1] Maximiliano Agustin Mascheroni, Emanuel Irrazabal& Gustavo Rossi 2021 “Continuous Testing Improvement Model” DOI: 10.1109/AST52587.2021.00020
- [2] MojtabaShahina, Muhammad Ali Babara& Liming Zhub “Continuous Integration, Delivery and Deployment: A Systematic Review on Approaches, Tools, Challenges and Practices” IEEE Access PP(99) DOI: 10.1109/ACCESS.2017.2685629
- [3] Maximiliano A. Mascheroni& Emanuel Irrazábal, “Continuous Testing and Solutions for Testing Problems in Continuous Delivery: A Systematic Literature Review” Vol. 22, No. 3 pp. 1009–1038 2018 doi: 10.13053/CyS-22-3-2794
- [4] Aayush Agarwal,Subhash Gupta &Tanupriya Choudhury “Continuous and Integrated Software Development using DevOps” 2018 DOI: 10.1109/ICACCE.2018.8458052
- [5] Roberto Pietrantuono,RobertoPietrantuono,AntoniaBertolino,Guglielmo De Angelis & Stefano Russo “Towards Continuous Software Reliability Testing in DevOps” 2019 DOI: 10.1109/AST.2019.00009
- [6] SriniketanMysari&VaibhavBejgam “Continuous Integration And Continuous Deployment PipelineAutomationUsing Jenkins Ansible” 2020 DOI: 10.1109/ic-ETITE47903.2020.239
- [7] Ahmad Alnafessah, AlimUIGias&Runan Wang “Quality-Aware DevOps Research: Where Do We Stand” IEEE Access PP(99):1-1 Follow Journal 2021 DOI: 10.1109/ACCESS.2021.3064867
- [8] Louis S Wheatcraft& Layne Lewis “Concept Maturity Level (Article)” September INCOSE International Symposium 28(1) 2016 DOI: 10.1002/j.2334-5837.2018.00570.x
- [9] Ravi TejaYarlagadda “DevOps and Its Practices” 2021 SSRN Electronic Journal 9(3):2320-2882
- [10] Ramtin Jabbari &Nauman Bin Ali, Kai Petersen Towards a Benefits Dependency Network for DevOps based on a Systematic Literature Review (JOURNAL OF SOFTWARE: EVOLUTION AND PROCESS) Journal of Software: Evolution and Process 30(06):e1957 DOI: 10.1002/smr.1957
- [11] MayankGokarna& Raju Singh “DevOps: A Historical Review and Future Works” 2021 DOI: 10.1109/ICCCIS51004.2021.9397235
- [12] MayankGokarna “DevOps phases across Software Development Lifecycle” 2018 DOI: 10.36227/techrxiv.13207796.v2
- [13] Ravi Teja& Sameer Mohammad “Understanding DevOps & bridging the gap from continuous integration to continuous delivery” 2018 SSRN Electronic Journal 5(2):1420-1424
- [14] Roshan Rajapakse, Mansooreh Zahedi, M. Ali Babar &Haifeng Shen “Challenges and solutions when adopting DevSecOps: A systematic review” Information and Software Technology 141(6):106700 2021 DOI: 10.1016/j.infsof.2021.106700
- [15] MojtabaShahin& Muhammad Ali Babar “On the Role of Software Architecture in DevOps Transformation: An Industrial Case Study” International Conference on Software and Systems Process (ICSSP 2020)
- [16] MojtabaShahin, Muhammad Ali Babar, Mansooreh Zahedi & Liming Zhu “Beyond Continuous Delivery: An Empirical Investigation of Continuous Deployment Challenges 2017 “DOI: 10.1109/ESEM.2017.18
- [17] Ashutosh Chaudhary RishabhSethiaShubham Kant Sinha Mary Gabriel “Cloud DevOps CI – CD Pipeline”
- [18] Omer Uludag, Sascha Nagele, Matheus Hauder& Florian Matthes, "A Tool Supporting Architecture Principles and Guidelines in Large-Scale Agile Development”.
- [19] Ana Filipa, Nogueira; Jose, C.B. Ribeiro; Mário A. Zenha-Rela& Antoine Craske Improving La Redoute's CI/CD Pipeline and DevOps Processes by Applying Machine Learning Techniques.
- [20] <https://www.altexsoft.com/blog/dataops-essentials/>