DevOps and DataOps (Development, Operations) ..

Abstract

In this report I will talk about a set of comprehensive and multi-sided terms in addition to an introduction that contains some comprehensive definitions, and although there is no precise definition of DVBs, I mean that their idea is based on cooperation between both software developers and IT experts, to complete the development process. Software in addition to I offered tools for both of them and in the end the difference between them.

Introduction

It is an amazing time to be working with data. Exponential growth in data collection, advances in machine learning (ML) and artificial intelligence (AI) algorithms, explosion in software libraries for working with much bigger quantities of data than was possible even a decade ago, and advances in big data technologies for storing and processing data have ushered in a transformative period for business, science, and government. Data science aims to aid in better decision-making, leading to beneficial actions than we otherwise could achieve by extracting knowledge from data. Data science does this by applying the scientific method, algorithms, and processes to data in various forms. Data science cannot exist on its own and is part of an ecosystem of skills that includes data engineering and the broader field of data analytics. Although there is hype associated with any technological change, and data science is no exception, many industries and fields are still at the beginning of the data-driven digital transformation. Over the next decade, machine learning, deep learning, and other data science techniques will transform every aspect of our lives from personalized healthcare to financial management to how we interface with machines, whether self driving cars or virtual assistants. Just as we are at the beginning of data-driven transformation, we are also only at the start of the journey to understand the best processes required to deliver our desired outcomes from raw data. Modern data science is still in the comparable transition phase between bespoke hand-crafted production and mechanized automation that manufacturing was confronting during the early 19th century.{1}

DevOps (Development, Operations)

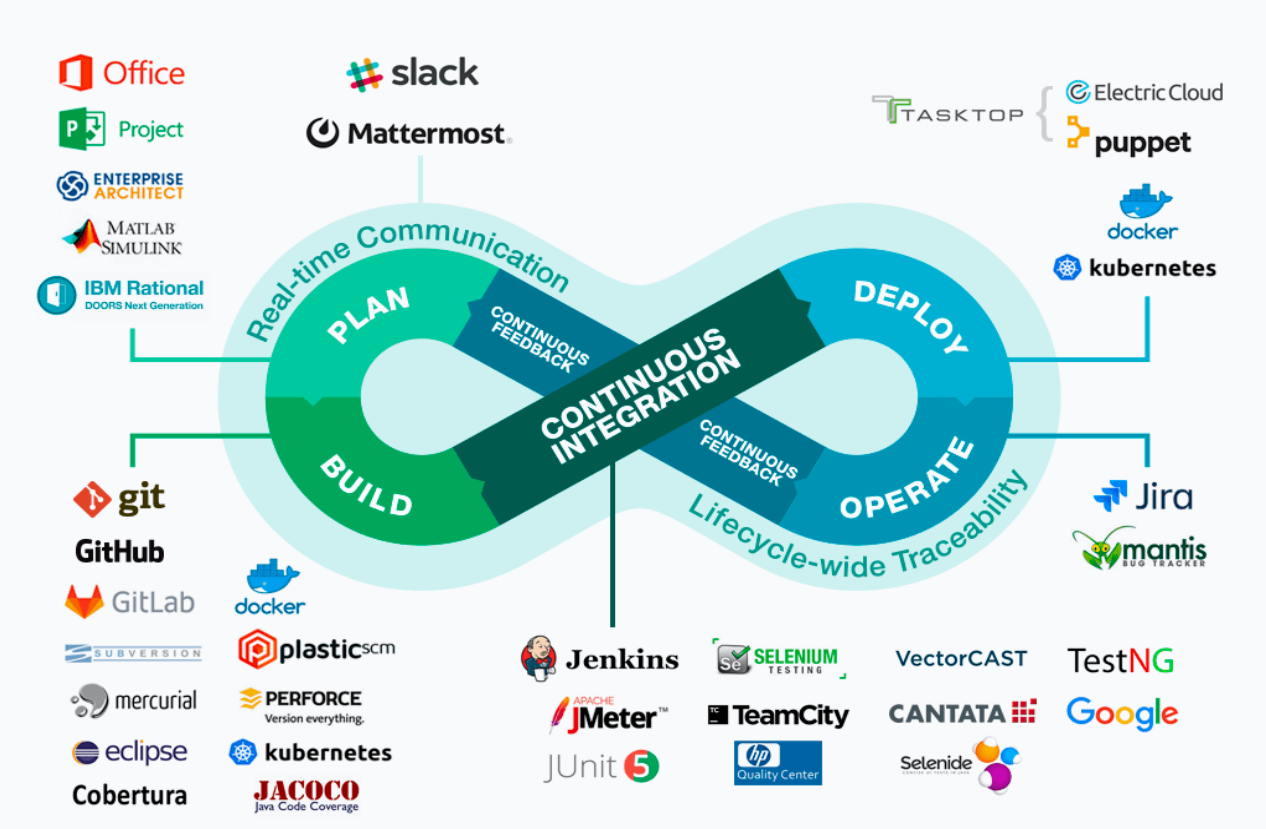


Figure 1: devops-it-operations

DevOps (Development, Operations) is a set of practices that combines software development (Dev) and IT operations (Ops). It aims to shorten the systems development life cycle and provide continuous delivery with high software quality.

DevOps is complementary with Agile software development; several DevOps aspects came from the Agile methodology {2}

Definition

Other than it being a cross-functional combination of the terms and concepts for "development" and "operations," academics and practitioners have not developed a unique definition for the term "DevOps".

From an academic perspective, Len Bass, Ingo Weber, and Liming Zhu—three computer science researchers from the CSIRO and the Software Engineering Institute—suggested defining DevOps as "a set of practices intended to reduce the time between committing a change to a system and the change being placed into normal production, while ensuring high quality".

The term DevOps, however, has been used in multiple contexts . {3}

History

Many of the ideas fundamental to DevOps practices are inspired by, or mirror, practices such as Lean and Deming's Plan-Do-Check-Act cycle, through to The Toyota Way and the Agile approach of breaking down components and batch sizes. In 1993 the Telecommunications Information Networking Architecture Consortium (TINA-C) defined a Model of a Service Lifecycle that combined software development with (telecom) service operations. Some say that DevOps emerged in part as a reaction to the "top-down" proscriptive approach of ITIL in the 1990s. DevOps, as a "bottom-up" approach, gained traction and persisted because it was created by software engineers for software engineers, and is a flexible practice rather than a rigid framework.

In 2009, the first conference named devopsdays was held in Ghent, Belgium. The conference was founded by Belgian consultant, project manager and agile practitioner Patrick Debois. [who?] The conference has now spread to other countries.

In 2012, the State of DevOps report was conceived and launched by Alanna Brown at Puppet. As of 2014, the annual State of DevOps report was published by Nicole Forsgren, Gene Kim, Jez Humble and others. In 2014, they found that DevOps adoption was accelerating. Also in 2014, Lisa Crispin and Janet Gregory wrote the book More Agile Testing, containing a chapter on testing and DevOps.{4}

The best DevOps tools :

1.**Gradle**

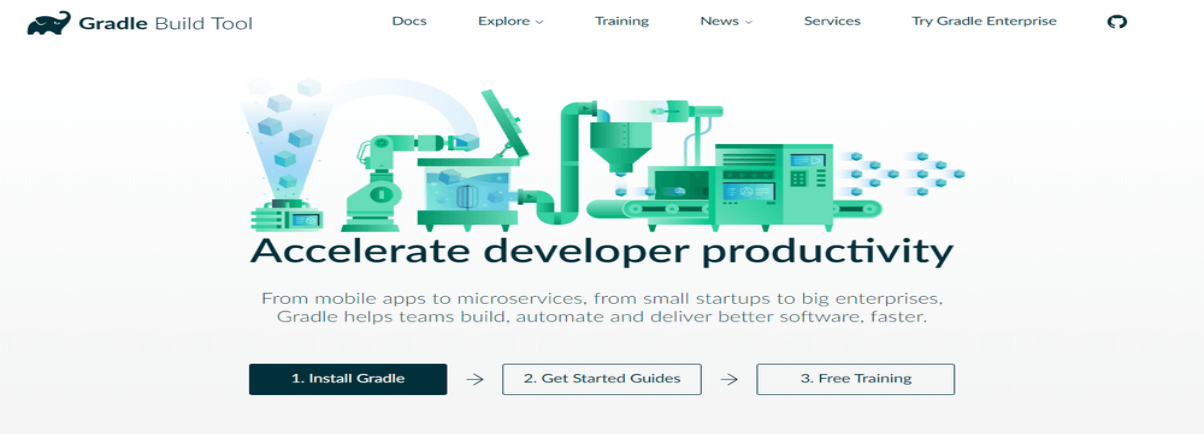
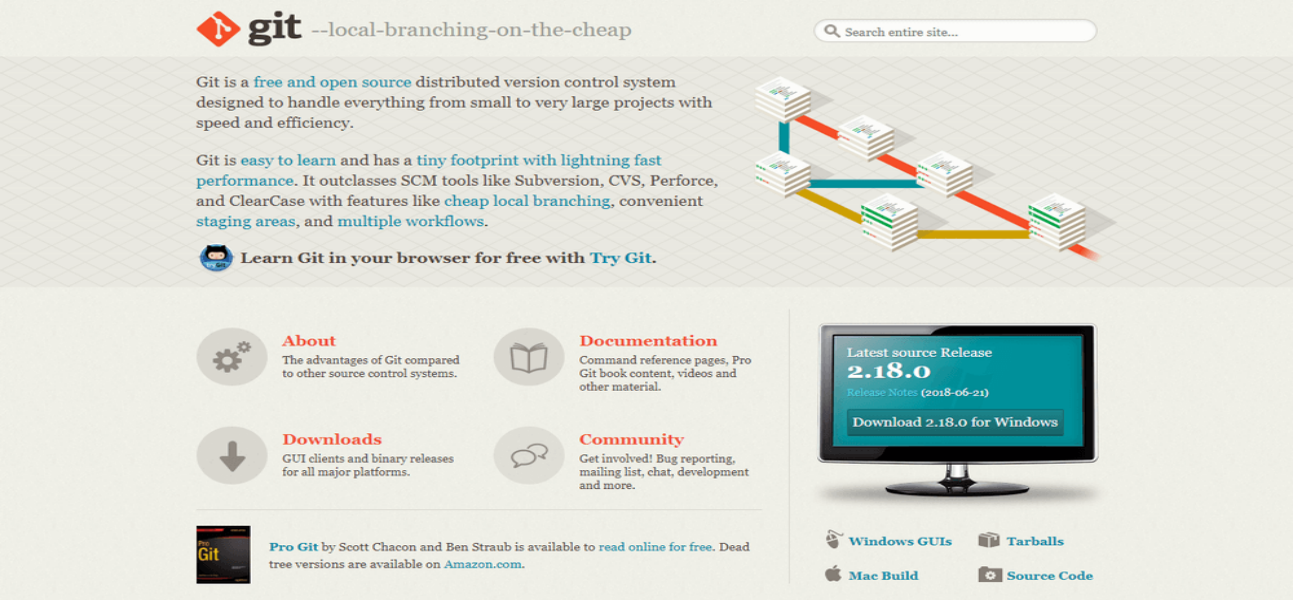


Figure {2}:

Your DevOps tool stack will need a reliable build tool. Apache Ant and Maven dominated the automated build tools market for many years, but Gradle showed up on the scene in 2009, and its popularity has steadily grown since then. Gradle is an incredibly versatile tool which allows you to write your code in Java, C++, Python, or other languages. Gradle is also supported by popular IDEs such as Netbeans , Eclipse, and IntelliJ IDEA. If that doesn't convince you, it might help to know that Google also chose it as the official build tool for Android Studio.

While Maven and Ant use XML for configuration, Gradle introduces a Groovy-based DSL for describing builds. In 2016, the Gradle team also released a Kotlin-based DSL, so now you can write your build scripts in Kotlin as well. This means that Gradle does have some learning curves, so it can help a lot if you have used Groovy, Kotlin or another JVM language before. Besides, Gradle uses Maven's repository format, so dependency management will be familiar if you have prior experience with Maven. You can also import your Ant builds into Gradle.

The best thing about Gradle is incremental builds, as they save a nice amount of compile time. According to Gradle's performance measurements, it's up to 100 times faster than Maven. This is in part because of incrementality, but also due to Gradle's build cache and daemon. The build cache reuses task outputs, while the Gradle Daemon keeps build information hot in memory in-between builds. All in all, Gradle allows faster shipping and comes with a lot of configuration possibilities.{5}



2.Git

Git is one of the most popular DevOps tools, widely used across the software industry. It's a distributed SCM (source code management) tool, loved by remote teams and open source contributors. Git allows you to track the progress of your development work. You can save different versions of your source code and return to a previous version when necessary. It's also great for experimenting, as you can create separate branches and merge new features only when they're ready to go.

To integrate Git with your DevOps workflow, you also need to host repositories where your team members can push their work. Currently, the two best online Git repo hosting services are GitHub and Bitbucket. GitHub is more well-known, but Bitbucket comes with free unlimited private repos for small teams (up to five team members). With GitHub, you get access only to public repos for free—which is still a great solution for many projects.

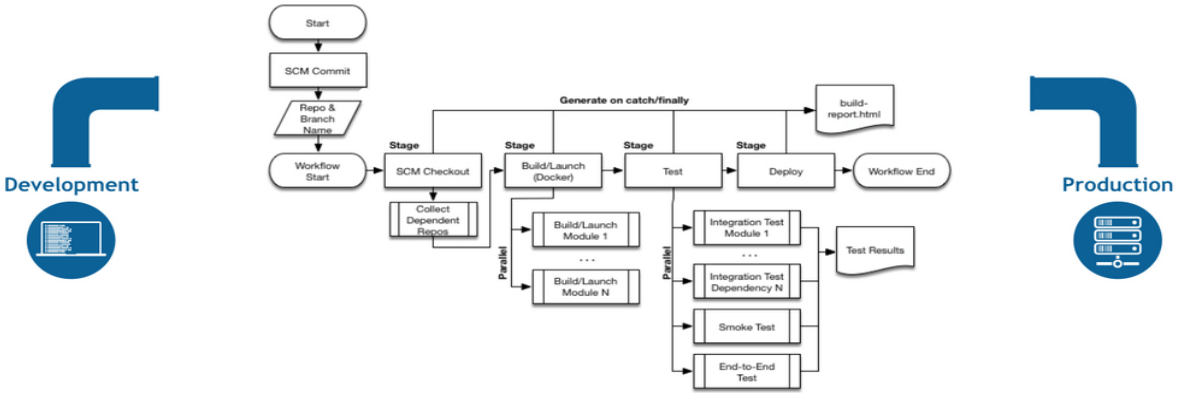
Both GitHub and Bitbucket have fantastic integrations. For example, you can integrate them with Slack, so everyone on your team gets notified whenever someone makes a new commit.{6}

3. Jenkins

Figure {4}:

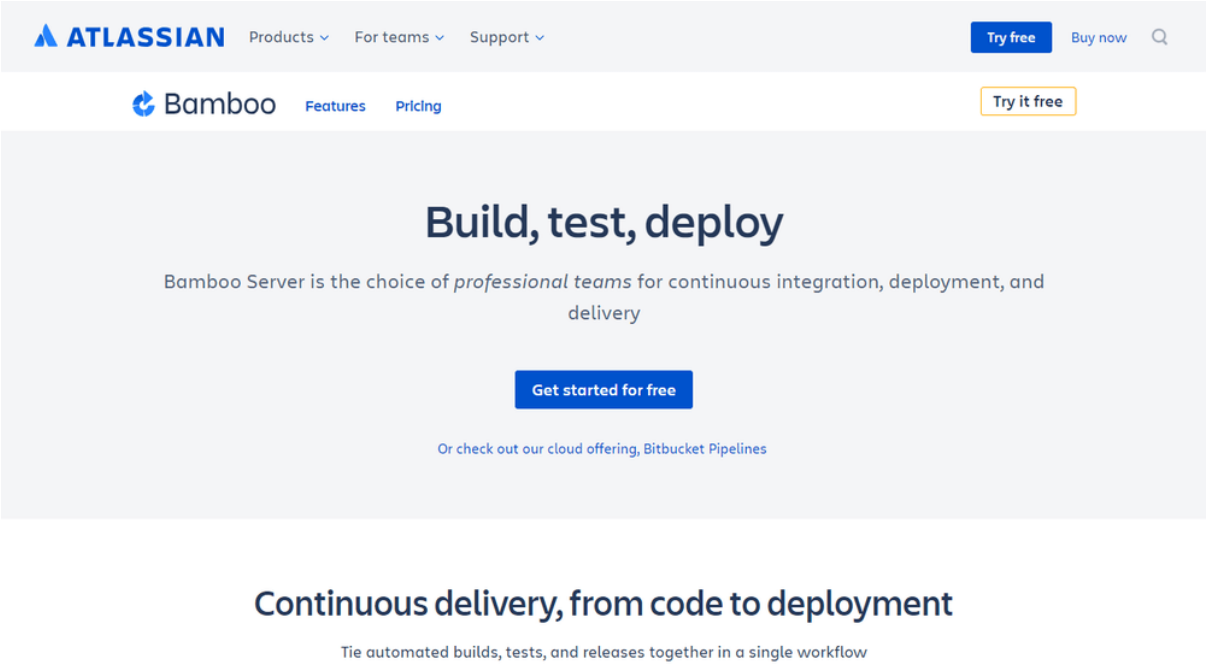
Jenkins is the go-to DevOps automation tool for many software development teams. It's an open source CI/CD server that allows you to automate the different stages of your delivery pipeline. The main reason for Jenkins' popularity is its huge plugin ecosystem. Currently, it offers more than 1,000 plugins, so it integrates with almost all DevOps tools, from Docker to Puppet.

With Jenkins, you can set up and customize your CI/CD pipeline according to your own needs. I found the following example in the Jenkins Docs. And, this is just one of the possibilities. Nice, isn't it?



It's easy to get started with Jenkins, as it runs out-of-the-box on Windows, Mac OS X, and Linux. You can also easily install it with Docker. You can set up and configure your Jenkins server through a web interface. If you are a first-time user, you can choose to install it with frequently used plugins. However, you can create your own custom config as well With Jenkins, you can iterate and deploy new code as quickly as possible. It also allows you to measure the success of each stepof your pipeline. I've heard people complaining about Jenkins' "ugly" and non-intuitive UI. However, I could still find everything I wanted without any problem.{7}

4. Bamboo



Bamboo is Atlassian's CI/CD server solution that has many similar features to Jenkins. Both are popular DevOps tools that allow you to automate your delivery pipeline, from builds to deployment. However, while Jenkins is open source, Bamboo comes with a price tag. So, here's the eternal question: is it worth choosing proprietary software if there's a free alternative? It depends on your budget and goals.

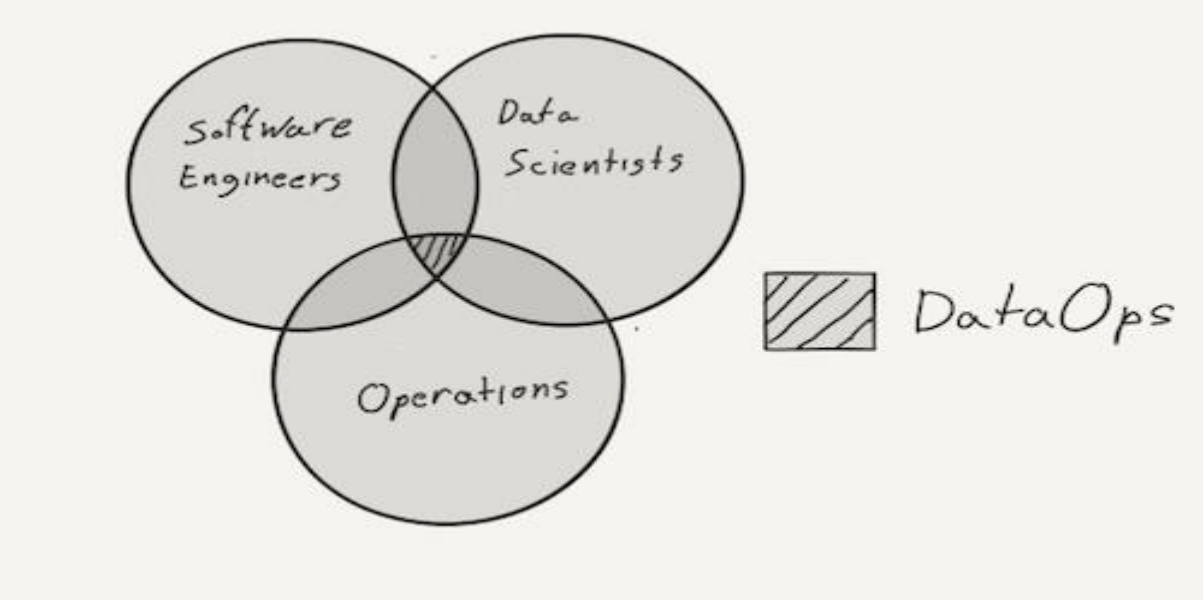
Bamboo has many pre-built functionalities that you have to set up manually in Jenkins. This is also the reason why Bamboo has fewer plugins (around 100 compared to Jenkins' 1000+). In fact, you don't need that many plugins with Bamboo, as it does many things out-of-the-box.

Bamboo seamlessly integrates with other Atlassian products such as Jira and Bitbucket. You also have access to built-in Git and Mercurial branching workflows and test environments. All in all, Bamboo can save you a lot of configuration time. It also comes with a more intuitive UI with tooltips, auto-completion, and other handy features.{8}

Which DevOps tools are right for your team?

Finding the best DevOps tools takes some testing and experimentation. It usually takes more time to set up and configure open-source tools. Most commercial DevOps tools come with free trials that allow you to test and evaluate them at no cost. It all boils down to your needs and goals.{9}

DataOps :(Data operations)



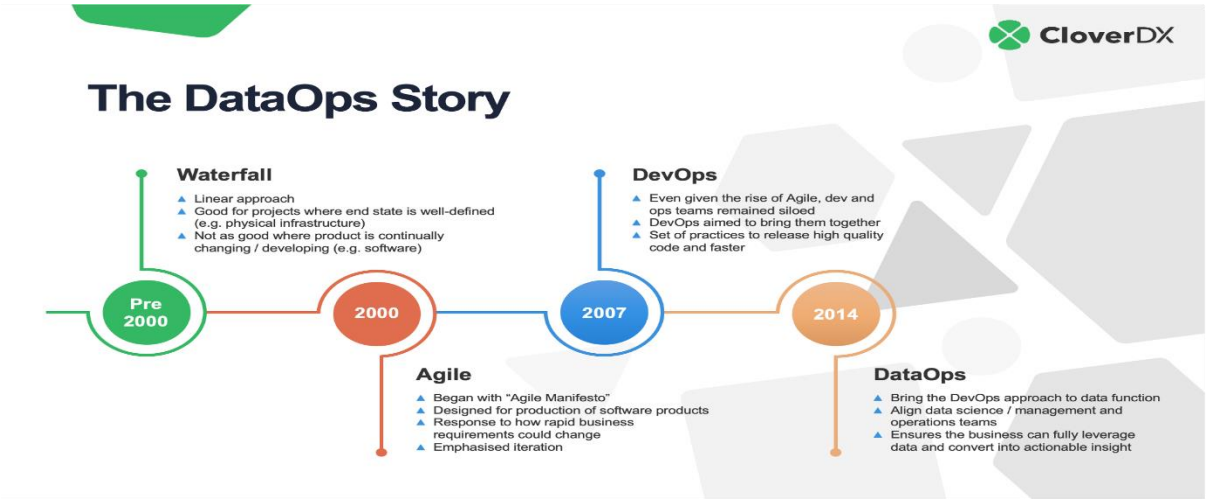
DataOps is an automated, process-oriented methodology, used by analytic and data teams, to improve the quality and reduce the cycle time of data analytics. While DataOps began as a set of best practices, it has now matured to become a new and independent approach to data analytics. DataOps applies to the entire data lifecycle from data preparation to reporting, and recognizes the interconnected nature of the data analytics team and information technology operations.

DataOps incorporates the Agile methodology to shorten the cycle time of analytics development in alignment with business goals.

DevOps focuses on continuous delivery by leveraging on-demand IT resources and by automating test and deployment of software. This merging of software *development* and IT *operations* has improved velocity, quality, predictability and scale of software engineering and deployment. Borrowing methods from DevOps, DataOps seeks to bring these same improvements to data analytics.

DataOps utilizes statistical process control (SPC) to monitor and control the data analytics pipeline. With SPC in place, the data flowing through an operational system is constantly monitored and verified to be working. If an anomaly occurs, the data analytics team can be notified through an automated alert.

DataOps is not tied to a particular technology, architecture, tool, language or framework. Tools that support DataOps promote collaboration, orchestration, quality, security, access and ease of use. {10}



What is DataOps?

DataOps is the orchestration of people, processes, and technology to deliver trusted, business-ready data to data citizens, operations, and applications throughout the data lifecycle. With properly governed data, businesses can comply with complex regulations, data privacy and ensure artificial intelligence (AI) model accuracy by monitoring data quality.

Organizations can deploy market-leading DataOps capabilities with IBM Cloud Pak® for Data, a containerized cloud-native platform built on Red Hat® OpenShift®. Solutions include multicloud data integration, automated data cataloging and master data management that increase efficiency, data quality, findability and governing rules in order to provide a self-service data pipeline.{11}

Compare the Top DataOps Tools in 2021

DataOps, short for Data Operations, is a newer methodology comprised of technical practices, patterns, and workflows designed to streamline rapid innovation, quality, collaboration, and more. DataOps incorporates elements of DevOps, agile development, and big data. DataOps tools are useful when looking to implement DataOps practices within your organization. Here's a list of the best

DataOps tools:

1-K2View

K2View provides an operational data fabric dedicated to making every customer experience personalized and profitable. The K2View platform continually ingests all customer data from all systems, enriches it with real-time insights, and transforms it into a patented Micro-Database™ - one for every customer. To maximize performance, scale, and security, every micro-DB is compressed and individually encrypted. It is then delivered in milliseconds to fuel quick, effective, and pleasing customer interactions. Global 2000 companies – including AT&T, Vodafone, Sky, and Hertz – deploy K2View in weeks to deliver outstanding multi-channel customer service, minimize churn, achieve hyper-segmentation, and assure data compliance.

2-Tengu

Tengu enables companies to become data-driven, and boost their business by: · making this data most useful and accessible at the right moment, · increasing the efficiency of the data scientists and engineers in executing their tasks to fasten up the data-to-insights cycle, and · helping them understand and manage the complexity of building and operating a data-driven company. TENGU is a DataOps platform for data-driven companies, that enables them to improve the efficiency of data scientists, analysts and other profiles inside the company. It enables them to focus on business intelligence instead of data operations.

3-Superb AI

Superb AI provides a new generation machine learning data platform to AI teams so that they can build better AI in less time. The Superb AI Suite is an enterprise SaaS platform built to help ML engineers, product teams, researchers and data annotators create efficient training data workflows, saving time and money. Majority of ML teams spend more than 50% of their time managing training datasets Superb AI can help. On average, our customers have reduced the time it takes to start training models by 80%. Fully managed workforce, powerful labeling tools, training data quality control, pre-trained model predictions, advanced auto-labeling, filter and search your datasets, data source integration, robust developer tools, ML workflow integrations, and much more. Training data management just got easier with Superb AI. Superb AI offers enterprise-level features for every layer in an ML organization

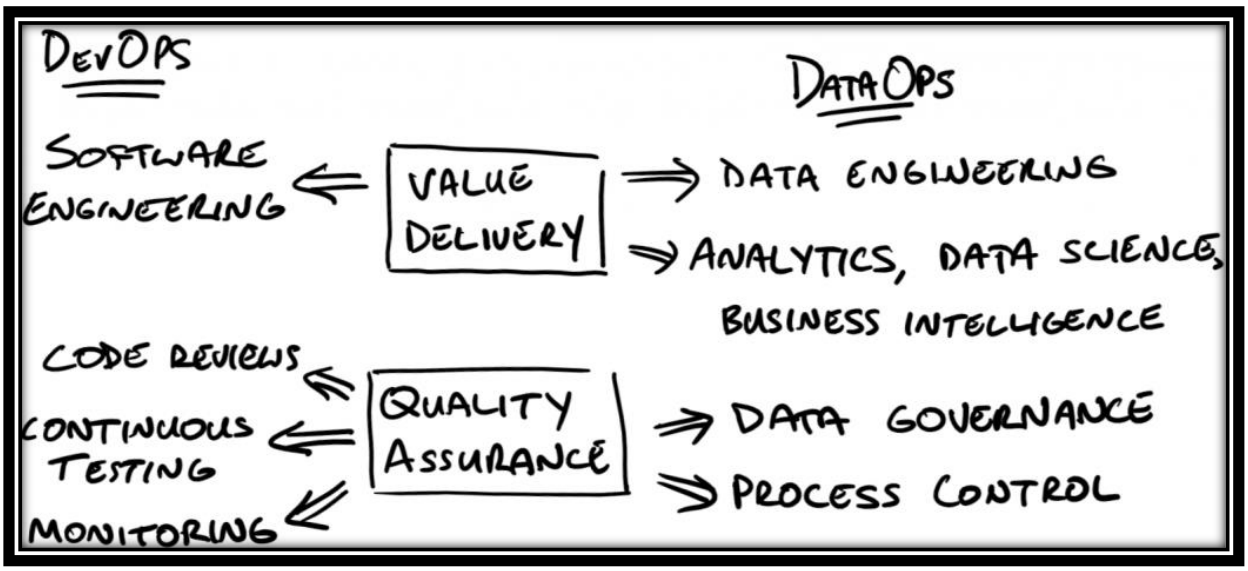
5-Zaloni

End-to-end DataOps built on an agile platform that improves and safeguards your data assets. Arena is the premier augmented data management platform. Our active data catalog enables self-service data enrichment and consumption to quickly control complex data environments. Customizable workflows that increase the accuracy and reliability of every data set. Use machine-learning to identify and align master data assets for better data decisioning. Complete lineage with detailed visualizations alongside masking and tokenization for superior security. We make data management easy. Arena catalogs your data, wherever it is and our extensible connections enable analytics to happen across your preferred tools. Conquer data sprawl challenges: Our software drives business and analytics success while providing the controls and extensibility needed across today’s decentralized, multi-cloud data complexity.

4-StreamSets

DataOps Platform. Your data integration engine for flowing data from myriad batch and streaming sources to your modern analytics platforms. * Collaborative, visual pipeline design * Deploy and scale on-edge, on-prem or in the cloud * Map and monitor dataflows for end-to-end visibility * Enforce Data SLAs for availability, quality and privacy Go further, faster. Replace specialized coding skills will visual pipeline design, test and deployment. Get projects up and running in a fraction of the time. Don't let brittle pipelines and lost data cripple your applications. Handle unexpected changes automatically. Get a live map with metrics, alerting and drill down. The platform decouples data integration from the underlying data systems. Add and upgrade new sources and analytics platforms without downtime. Be sure your applications are getting the right data in time to perform as you designed them. Only StreamSets provides the end-to-end reliability, visibility & control.**{12}**

The difference between DevOps and DataOps :



DataOps aims to improve the quality and reduce the cycle time of Data and Analytics initiatives. The difference between DataOps and DevOps is in the unique nature of developing with data and delivering data to users.

Having seen some of the ways that Data and Analytics differs from software engineering (the Data and Analytics pipeline, the stateful Data Store, and Process Control) we can also show the different ways in which DevOps and DataOps deliver value and assure quality{13}

References

{1} https://link.springer.com/content/pdf/bfm%3A978-1-4842-5104-1%2F1.pdf

{2}{3}{4} https://en.wikipedia.org/wiki/DevOps

Figure {1}:https://intland.com/wp-content/uploads/2019/07/devops-infinity-1-1.png

{5}{6}{7}{8} https://raygun.com/blog/best-devops-tools/

Figure {2}: https://raygun.com/blog/images/best-devops-tools/git.png

{9}https://raygun.com/blog/best-devops-tools/

{10}https://en.wikipedia.org/wiki/DataOps

{11}https://www.ibm.com/analytics/dataops

{12}https://sourceforge.net/software/dataops/

{13}https://altis.com.au/the-difference-between-devops-and-dataops/