Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Good practices for the adoption of DataOps in the software industry

To cite this article: Manuel Rodriguez *et al* 2020 *J. Phys.: Conf. Ser.* **1694** 012032

Good practices for the adoption of DataOps in the software industry

Manuel Rodriguez[1], Luiz Jonatã Pires de Araújo and Manuel Mazzara

Innopolis University, Innopolis 420500, Russia

E-mail: [1]`m.rodriguez.osuna@innopolis.university`

October 2020

Abstract. The increasing adoption of DevOps, the growing availability of data concerning data development processes gives rise to the need for a systematic process for collecting, processing and using data into companies. Enterprises are making significant investments in data science applications while still struggling to realize the value of this effort. Data science is emerging as a fast-growing practice within enterprises. Several tools and platforms are being continuously introduced that support data science models while managing large data sets used to train data science models. Such a scenario lead to the emergence of DataOps. This paper summarises some of the good practices in the DataOps from the literature, offering guidelines intended to approach an organizational shift towards better data-driven decision making. This study presents a picture of the definition, the steps for adopting and challenges of the adoption of DataOps.

1. Introduction

In recent years, the amount of data produced has dramatically increased while the use of such data for generating value and insights by companies has been limited. The effort required to set up and maintain a complex data source frequently scales up, preventing its use without sophisticated software tools contributing to the current situation where enterprises are assumed to analyze less than one-sixth of their potentially relevant data [1]. In order to take a more accurate decision to increase the success of a business endeavour, a data-driven process which covers from data collection to analysis and decision is necessary. DataOps emerged to attend such requirements.

However, while DevOps focuses on a more agile software development process by automating the process which connects from development to deployment, DataOps focuses on the lifecycle of the data and its utilization to generate for the company. Moreover, DataOps can be understood as the function within an organization that controls the data journey from source to value [2]. The main focus of DataOps is to improve practices for data management and processes to increase the speed and the accuracy of analytics [3]. DataOps takes its cue from DevOps which is "an organizational approach that stresses empathy and cross-functional collaboration within and between teams" to increase the quality of the software and accelerate the delivery of changes and fixes [4].

Several companies have implemented DataOps to employ their data to drive functions to increase productivity [5]. Some of the AI-driven data applications that benefit from this process include internet of things IoT, complex financial analysis models, and the autonomous vehicle operations which require petabytes of data to be managed. DataOps was introduced to address two main issues related to the management and utilization of data collected during the software process:

- *The need for more agile processes for data.* Business run at an incredibly fast pace. Therefore the data needs to move at the same pace; otherwise, it is ignored from the decision-making process.
- *Data became more mainstream over the years.* There is a growth in data sources thanks to the advancements in the collection from all the applications such as sensors, e-commerce, Internet of Things (IoT), and social media platforms.

The majority of the literature available has approached DataOps in an isolated manner. For example, there have been publications of the definitions, while other authors approach the challenges of the implementation of DataOps in industries. While the paper is not intended to provide a step-by-step guide for the implementation of DataOps. It contributes with general guidelines which are flexible for several software companies.

This paper illustrates some of the good practices of DataOps regarding collaboration between different groups within the organization, the automation and re-utilization of sub processes. It serves as a starting point for software companies interested in adopting DataOps practices. The steps presented here can work as guidelines to support data-driven decision making in the context of DataOps.

Section 2 of the paper presents the basics about DataOps including the methodology of development and the actors involved in the collaboration process. Section 3 goes into more depth on the components and sub-components of DataOps. Furthermore, Section 4 describes the steps towards the implementation of the discussed methodology, Section 5 describes the processes involved in the data-analytics pipeline, and Section 6 describes the challenges faced across the company in the adoption of DataOps. Finally,we conclude the paper with final remarks and some thoughts regarding future iterations of the research in Section 7.

2. Towards DataOps

Tools capable of managing business intelligence and advance analytics have become increasingly available and significant to companies looking to leverage the data they possess. The prospect of applying data analytics for different organizations has allowed to shape and evolve different techniques, tools and methodologies to analyze business data and assist managers to understand and assess the opportunities and threats in time, enabling timely decisions [6].

The main challenge of initiatives which handle a significant amount of data is to connect critical decision-makers to the correct data and to help them to understand what they can do with such data to make better decisions [7]. DevOps redefined the process in which companies develop and deliver their products by enforcing a more close collaboration between engineering teams to accelerate delivery.

2.1. DevOps within the DataOps context

Researchers [8] defined DevOps as "the practice of operations and development engineers participating together through the entire service lifecycle; from the design and development process towards production support". Figure 1 illustrates the lifecycle in a typical DevOps pipeline. In the context of data-analytics *Devs* focuses on algorithms and their performance and aim to get on-time software releases. On the other hand, *Ops* concerns the maintenance of the pipeline. It monitors the pipeline by putting in place the tools to manage the feedback from continuous experiments [9].

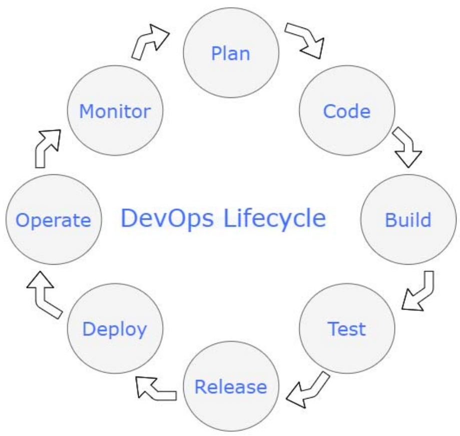

Figure 1. Continuity in the DevOps Lifecycle

2.2. Agile Analytics as a part of DataOps

Fast and precise analytics enable to make quick decisions allowing the company to be ahead of the competition. Companies require rapid and precise analytics to be able to withstand evolving market threats [10] . These analytics require systems that can incorporate quick changes in a structured way. Otherwise, several Ad-Hoc solutions created will end up running in parallel in systems due to the companies not being prepared to handle changes at a consistent pace [11].

Good practices for the adoption of DataOps in the software industry

Similar to agile methodologies, agile analytics are comprised by sets of rules, guidelines and principles, and its workflow can be used to perform data analytics. The aim of agile analytics is to avoid re-writing code and to use its workflow system to thread several analytic functions [12]. Agile analytics focuses on end goals and as a result enables data-driven predictions for decision making. The development and production of analytics systems using traditional systems can result in a high-quality product. However, much planning is required, and it can result in a high cost to implement [11].

The effort invested in building a data analytics application before getting any workable results can represent a significant setback for a company if the project fails. For this reason, the recommended approach is to have only sufficient planning, adoption to change, risks preparations and team involvement in order to achieve incremental development goals [13].

2.3. Methodology for software development within DataOps

The development of a DataOps system follows an iterative and incremental approach in its development to allow managers and stakeholders to prioritize user-valued features and track user's feedback during the development of the software [14]. A DataOps oriented team works with small iterations which typically range from one to three weeks and no more than four weeks. The small iterations following the continuous customer's feedback ensure that the project remains in satisfactory progress.

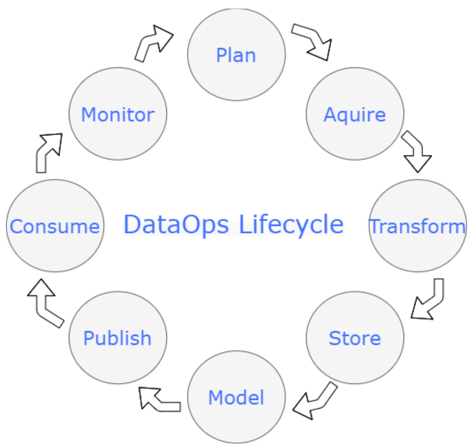

Figure 2. DataOps lifecycle, following the principles of DevOps

A practical approach to create business value is to follow a value-driven development (VDD) approach to get a valuable feature at the end of each iteration [15]. With VDD the development of the system is directly associated with its business economics values. Users are solely concerned about the outcomes of the system which addresses a business problem and disregard the complexities and steps taken in its development [16].

Every iteration of the development process should produce a software functionality according to the business requirements [17]. All features must be tested and appropriately debugged at every stage of the development. As the product evolves within every iteration, tests should be planned and performed to produce a high-quality product. The adoption of an Agile methodology requires that all routine processes be automated. Automated testing shifts the effort from routine tasks into more focus on the development of quality of the implemented functionalities [10].

Figure 2 illustrates the lifecycle of a typical DataOps process. DataOps follows some of the same principles of DevOps, but instead focuses on streamline data management processes to maximize the value of massive corporate data stores [5].

2.4. Actors in the implementation

Several different functional groups within an organization collaborate and work towards the implementation of DataOps [6], with the goal to deliver value. Figure 3 illustrate the main collaborators in the DataOps process in charge of generating business value:

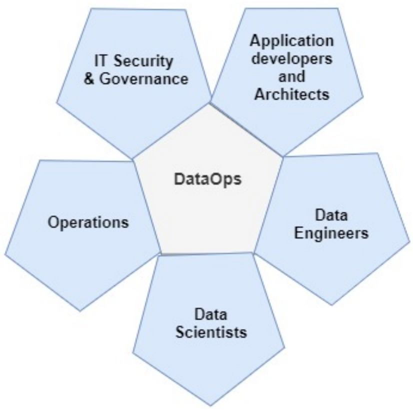

Figure 3. Cross-functional collaboration required for DataOps process

- *Data Scientists* apply data science techniques that include machine learning and deep learning models mostly through each iteration. They create rigorous and efficient models using frameworks available in the form of software libraries in Python or R, for example, alongside tools for processing big-data like Spark or Tensorflow, among others.
- *Data Engineers* deal with the aggregation of data and manage the datasets used for evaluating and training data models.
- *Developers and Architects* create end-to-end applications that include the models created by the data scientists.
- *Data Governance and IT* define the data access controls, allowing data scientists to access historical data for model training and benchmarking. Such controls can be accessed in a read-only state for model tuning and reflection.
- *Operations team (DevOps)* is responsible for deploying the applications into production environments to support service-level agreements.

Good practices for the adoption of DataOps in the software industry

2.5. Enterprise-grade platform requirements

DataOps focuses on people and process and requires an enterprise-grade platform capable of enabling the collaboration between teams and sharing the data and computational resources across all involved groups [6]. The technical requirements for a platform capable of supporting a DataOps process include:

- High availability and reliability
- Support for modern data science languages and tools
- Multitenancy
- Support for distributed architectures
- Support for linear and scalability growth within business and use cases

A single platform approach is a key to increase the agility, support a holistic approach and reduces the effort and time required to copy or move large datasets across siloed teams or process [6].

3. DataOps components

DataOps is a practice and methodology that can be split into sub-components that leverage existing analytics tools along with toolchain components. This sub-components address source control management, process management and effectively communicate groups to deliver a reliable data pipeline [13]. The four main components of a DataOps platform can be summarized as the following: data pipeline orchestration; assurance automation, quality control and monitoring; continuous deployment; deployment of the data science model.

3.1. Data Pipeline orchestration

The orchestrator is a software entity that is responsible for managing the processes, control the execution of the steps and handles the exceptions. It is the only one who knows when a pipeline should be executed at a given moment. Therefore, it should be the only one who can trigger that execution. DataOps requires a directed graph-based workflow that should contain every aspect of the data analytics production [18]. This should include the steps of the data analytics production, such as data gathering, access, integration, modelling and visualization.

3.2. Assurance Automation, Quality Control and Monitoring

The focus of assurance automation is to continuously monitor the production quality of both the data and the artefacts through automated tests [18]. All the changes made to the code have to be automatically tested during the deployment process. Integrated services to monitor and manage alerts should be included.

Good practices for the adoption of DataOps in the software industry

3.3. Continuous Deployment

Continuous deployment (CD) deals with the movement and continuous configuration of the code into a development environment in a seamless manner [4]. A good practice is the creation of sandboxes for the engineers to experiment proof of concept pipelines [18].

3.4. Deployment of the Data Science Model

Valuable insights need to be delivered quickly and continuously to ensure customer's satisfaction. Changes continually managed are the most effective way to generate value for the business. Hence, DataOps thrives on the frequent interaction between the data analytics team, the operations team and the customers [18].

4. Steps in the implementation of DataOps

Companies that successfully adopted the principles of DataOps experience notable improvements in regards to their ability to produce robust and adaptive analytics [19]. The following steps should be considered by organizations aiming for the implementation of DataOps.

4.1. Data Analysis tests and automated tests

Automated tests are an effective way to verify that all changes implemented in the pipeline are working correctly and not disrupting any functionality on the system. In order for the automated tests to work effectively[20], all tests need to be added incrementally, and every new feature is required to have a test case.

4.2. Use a version control system

The process required to transform the raw data into information must be carried through multiple stages and processes[21]. These processes can be serialized or in parallel. The pipeline is composed of scripts, configuration files, algorithms, containers, among other software artefacts. The artifacts[22] mentioned are in essence just code, and such code is subject to be modified and continuously improved, hence the necessity of tracking and maintain it through a version control system like Git. However, all the versions of the data need to be managed.

4.3. Branch and Merge

To ensure a smooth workflow[20] when every developer can make changes to the code without disrupting the work of another person in the team, the developer should pull a copy of the code to a local copy. The copy of the code is known as a branch where the developer works on it and after all the changes are made and tested then can be merged back to the trunk.

Journal of Physics: Conference Series 1694 (2020) 012032 doi:10.1088/1742-6596/1694/1/012032

4.4. Use Multiple environments

To be productive, a member of the data analytics team must have a copy of the data they need. Whenever the data set is too large to be distributed for a member of the team, a subset of the data required for the developer to work can be extracted. Any changes to a schema can lead to conflicts when the team is working in a production database. The data engineer developing or fixing a data model can get mixed up as new data emerges. Therefore, to minimize problems and isolate the impact of their work for the rest of the organization, every developer has to work on their environment.

4.5. Reuse and Containerize

A technique to boost productivity is the ability to reuse and containerize code; every intermediate step in the pipeline receives an output from a previous stage. A stage passed in the pipeline provides an input for an upcoming stage [23]. Working on the data analytics pipeline as a monolith can be unmanageable; the best approach is to divide the pipeline into smaller and reusable components. These components are more accessible to the team if they are segmented or containerized in Docker.

5. DataOps Processes

DataOps orchestrates, monitors and manages all the data factory. For this reason, without DataOps many of the most valuable resources a company can possess are susceptible to being wasted. DataOps is a people-driven practice, rather than a technology-oriented practice. Thus, it cannot be purchased as a tool or an application [24].

5.1. Initial Organization

5.1.1. Understanding business requirements Similar to the business requirement phase on a software development project, the same steps should be conducted in the context of business analytics[18]. During this stage, it is crucial to understand the overall process of the project. Moreover, it is worthy of consideration the sources of data, the acquisition process, the transformation of the data and the consumers of the data.

5.1.2. DataOps team formation The identification of prominent stakeholders is a primordial task. Some of the stakeholders can be end users, data analysts, data architects, developers, data scientists[25]. After identifying the key stakeholders, a team need to be formed with the representation of all identified members. This team should work collaboratively, starting with the definition of objectives.

5.1.3. Business consumers identification As the fourth principle [26] of DataOps suggests *team sport*, the business team and the information technology teams should be encouraged to work together in order to produce valuable insights from business data from various data sources. This step applies the principles of DevOps where it has done iteratively and not being *done once and forget it later*. As the consumers decide the requirements of the data analytics process in DataOps, evolving situations require that the business dynamics change with time alongside with the requirements for data analytics.

5.1.4. Objectives and Scope The objectives and scope can be clearly defined for an initial system. Many businesses currently are the combination of various business groups, and to consider every need can be a difficult task. The goal should be to leverage all available data in a reliable, efficient and cost-effective manner applying DataOps principles[27].

5.2. Data Management

5.2.1. Data Acquisition
Data sources. Internal and external sources of data need to be identified in this step alongside the challenges specific to the data acquisition for each.
Distribution process. Aspects such as security, accessibility, estimations for data download times, the available bandwidth for cloud data transfer and encryption mechanisms need to be evaluated[28].
Data Onboarding. All the identified data has to go through an onboarding process to be subject to standardized sanity checks to verify the security, the type of the data and encryption. This process should be under the control of the data analytics team before it is used. The onboarding process can be automatized and validated trough scripts.
Tools Identification. The capabilities for each tool should be evaluated to fit the purpose of context for every step of the pipeline.

5.2.2. Data Transformation
Challenges. The lack of involvement from the end-users in the insight creation process causes mistakes along the way due to engineers address the business requirements according to their understanding to develop the transformation logic. This causes the end product to several times to present wrong results and requires intervention to fix them in an ad-hoc manner due to the high cost in time and money needed to rebuild a model.
Transformation strategies. Applying agile and DevOps methodologies developers can iteratively build the data models limiting the risks and avoiding the loss of effort in the project. Multiple teams should be involved in the process to allow the quick iterations of the process and produce minimum viable products (MVP) in an agile manner. All data should be transformed in a manner that allows it to be reused in multiple scenarios as depicted in Figure 4, this data should be passed through the correct filters and tested to ensure its quality.

Good practices for the adoption of DataOps in the software industry

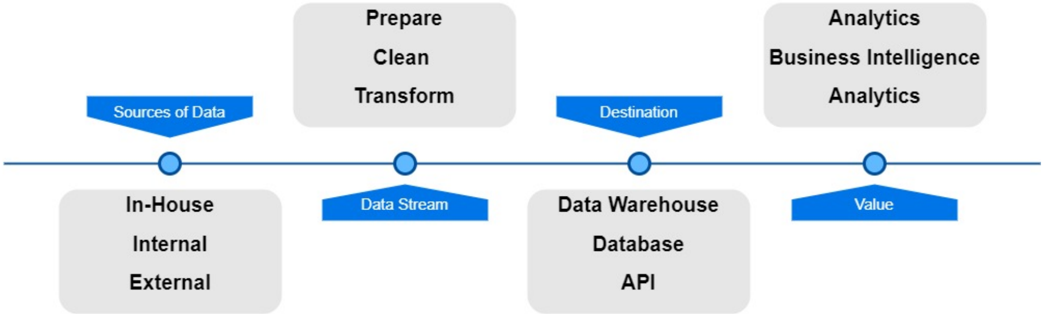

Figure 4. Data Journey across the DataOps pipeline

Workflow. A development, build, test and review flow should be followed. The resulting code for each iteration should be tested then analyzed by the appropriate stakeholders such as the data scientists, data analysts and architects. As new data sets are captured, it should follow the same cycle.

5.2.3. Data Repository Management
Challenges. It is crucial for the success of the data analytics that the repository be able to scale while providing security and well-established access control. However, with a lack of clear objectives, it is difficult for managers to carry out from the business units which leads to spread repositories across the lines of business in the enterprise.
Strategies. Following the guidelines for best practices on the management of datasets, the following strategies are proposed:

- Determine size
- Define architecture for security and storage
- Define hardware and software
- Monitoring and planning

In order to select the repository for the data first, it is critical to determine the potential size of it since any enterprise data volume may be immense and expected to grow exponentially. Therefore, careful analysis of the current size, future growth and security should be carried out.

5.2.4. Data Modeling A model is mostly created to solve a specific use case with little regard to reusability or growth. The team picks a problem and data, and after some analysis, an ad-hoc solution or quick prototype is created. Such a prototype would be implemented in a specific set of data or department, which defeats the purpose of DataOps at a larger scale. Therefore, a scientific approach is required to encourage growth and enabling insight creation.

Data can be selectively chosen from the repository to create different types of data models. The creation of a single data model for all cases would complicate the model.

Depending on the requirements, different techniques such as supervised, unsupervised, or semi-supervised learning methods can be used [29].

Despite the existence of established models and methodologies, the creation of an adequate trained model is a challenge. Well established techniques and practices should be available throughout the organization allowing them to be reused. Solving problems in isolation is highly discouraged since it will be time-consuming and counter-productive.

5.3. Quality Control

5.3.1. Review The review allows us to discover multiple problems that an enterprise-level mostly lay in silos. When a problem is solved, the details are not always available to other departments. Therefore, a significant amount of time is spent trying to resolve challenges.

In an enterprise, a single model repository should be accessible across all divisions and departments; thus, if a problem arises, it can be solved with an existing solution available. The review process minimizes the rework while enabling reusability and optimizing solutions with minimal possible effort.

5.3.2. Monitor and Plan Monitoring the steps continuously and in a disciplined fashion allows to improve errors and inadequacies in every step. All identified problems should be stored in the backlogs from the models. Several metrics can be collected to monitor the overall status of the project. This data will allow us to understand the consumer's interaction, identify patterns, assess existing threats, and evaluate the performance of the models.

As time passes, some of the data have to be purged, a proper strategy for archiving or purging has to be planned. Periodic data audits can be performed to help smooth this process to ensure that any unwanted data is purged.

6. Challenges in DataOps

The boom of DataOps is relatively new in the industry, and there is limited experience in the field and multiple research challenges. Some of the biggest challenges in the implementation of DataOps in a company can be summarized as the following:

- Goals and requirements continuously evolve, despite the efforts of the team to deliver solutions. Users will generate new requests continuously.
- Companies nowadays collect massive amounts of data through multiple devices and platforms. The majority of this data remains in silos, being stored separately and communicated through multiple APIs and services. The process of integrating this data is involved, lengthy and subject to bottlenecks.
- The data from operational systems is generally not structured for the creation of analytics. Also, a well-structured database has to be optimized for reads and aggregations. The data schema needs to be easily understood by humans with

descriptive names and intuitive connections between the content and the data tables.

- Errors in external and internal data are unavoidable; such errors can prevent the analytics pipeline to perform optimally. Data errors can be subtle and difficult to trace and resolve.
- Unfixed data errors in the published analytics dissatisfy stakeholders and may damage the reputation of the company and the data analytics team.
- With a data analytics team continually working on changes and improvements, a growing effort is required to validate and verify the integrity of the modifications. The effort required to maintain and assure the quality of the pipeline also increases with the complexity of the system.
- Some of the data processes can not be automatized and have to be performed manually regularly. These tasks tend to be prone to errors, consume more time and tedious.

Overcoming these challenges will require an overhaul to the methodologies and processes guiding the company's workflow [20]. The challenges in the research paper arise in the lack of a prototype to test the steps discussed above. Every step was verified according to the available literature on the topic. The implementation of the mentioned steps of DataOps described should not be directly applied to every company since every business' goals, and resources vary, and a generic solution is not recommended.

7. Conclusion

The original idea from the Big Data 1.0 movement suggested that the aggregation of massive volumes of data was itself valuable. It is now in our understanding that the aggregation is not enough, since this data does not bring any value until it is applied to impact growth or profitability through operational efficiency.

While DataOps continues to emerge as an organizational level practice for organizing teams in a collaborative manner to produce data applications, it represents a important trend that recognizes that data-intensive applications have their own set of considerations in regards of management and the security of large complex datasets while allowing an agile access for the people who need it.

Building a scalable, high performance and highly available system require a monitored and disciplined approach. The potential of a well maintained pipeline can help reduce the time and cost of making an important decision at any step of the progress due to its automated and monitored nature.

This paper gathered and highlighted good practices in DataOps reported in the academic literature and serves as a starting point for practitioners willing to implement DataOps in their companies. It also proposes a set of guidelines intended to approach an organizational shift towards a better data driven decision making. Moreover, it contributes to build a data-analytic system which will be an aid into the success of a

REFERENCES

business through the conversion of their data into valuable business insights. Finally, the research illustrates a comprehensive picture of the definition, the steps and challenges in the adoption of DataOps. Future research is intended to investigate real-world cases of the adoption of DataOps with the purpose of evaluating the effectiveness of DataOps.

References

[1] Pinkel C, Schwarte A, Trame J, Nikolov A, Bastinos A S and Zeuch T 2015 DataOps: Seamless End-to-End Anything-to-RDF Data Integration *The Semantic Web: ESWC 2015 Satellite Events* ed Gandon F, Guéret C, Villata S, Breslin J, Faron-Zucker C and Zimmermann A (Cham: Springer International Publishing) pp 123–127 ISBN 978-3-319-25639-9

[2] 2017 What is DataOps? | Platform for the Machine Learning Age | Nexla library Catalog: www.nexla.com Section: Data Operations URL `https://www.nexla.com/define-dataops/`

[3] Swanson B M What is DataOps? Everything You Need to Know library Catalog: blogs.oracle.com URL `https://blogs.oracle.com/datascience/what-is-dataops-everything-you-need-to-know`

[4] Ereth J 2018 DataOps-Towards a Definition. *LWDA* pp 104–112

[5] DataOps and the future of data management library Catalog: www.technologyreview.com URL `https://www.technologyreview.com/2019/09/24/132897/dataops-and-the-future-of-data-management/`

[6] Valentine C and Merchan W DataOps: an Agile Methodology for Data-Driven Organizations library Catalog: mapr.com URL `https://mapr.com/whitepapers/datops-an-agile-methodology-for-data-driven-organizations/`

[7] Thusoo A and Sarma J 2017 *Creating a Data-Driven Enterprise with DataOps* (O'Reilly Media, Incorporated)

[8] Demchenko Y, Zhao Z, Surbiryala J, Koulouzis S, Shi Z, Liao X and Gordiyenko J 2019 Teaching DevOps and cloud based software engineering in university curricula *2019 15th International Conference on eScience (eScience)* (IEEE) pp 548–552

[9] Benjamin B, Mireille B F, Sébastien M, Frédéric P and Günther J 2019 When DevOps meets meta-learning: A portfolio to rule them all *2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)* (IEEE) pp 605–612

[10] Ullah R 2019 *Agile Analytics: How organizations can benefit from the agile methodology for smoother delivery of data-driven analytics projects* masterthesis Universität Koblenz-Landau, Universitätsbibliothek

[11] Earley S 2014 Agile analytics in the age of big data

[12] Kuah S C 2018 *On the Construction of a Text Mining Workflow System* masterthesis

Journal of Physics: Conference Series **1694** (2020) 012032 doi:10.1088/1742-6596/1694/1/012032

REFERENCES

[13] Gupta R Components of the DataOps toolchain and best practices to make it library Catalog: www.ibmbigdatahub.com URL `https://www.ibmbigdatahub.com/blog/components-dataops-toolchain-and-best-practices-make-it-successful`

[14] Larman C and Basili V R 2003 *Computer* **36** 47–56

[15] Ktata O and Lévesque G 2009 Agile development: Issues and avenues requiring a substantial enhancement of the business perspective in large projects *ACM International Conference Proceeding Series* pp 59–66

[16] Souza E, Abrahão S, Moreira A, Araújo J and Insfran E 2016 Comparing value-driven methods: an experiment design. *HuFaMo@ MoDELS* pp 19–26

[17] Bäck A 2019 *Data driven decision making in Industrial Control Systems* Master's thesis

[18] Sahoo P R and Premchand A 2019 Dataops in manufacturing and utilities industries

[19] Khalajzadeh H, Abdelrazek M, Grundy J, Hosking J and He Q 2018 A survey of current end-user data analytics tool support pp 41–48

[20] Bergh C, Benghiat G and Stroud E 2019 *The DataOps Cookbook* 1st ed (Cambridge: United States of America)

[21] Held J, Stonebraker M, Davenport T H, Ilyas I, Brodie M L, Palmer A and Markarian J

[22] Capizzi A, Distefano S and Mazzara M 2019 From DevOps to DevDataOps: Data Management in DevOps processes

[23] Zhang Z 2020 *DevOps for Data Science System*

[24] Peterson B 2018 Hyping DataOps library Catalog: blogs.gartner.com Section: Data and Analytics Strategies URL `https://blogs.gartner.com/nick-heudecker/hyping-dataops/`

[25] Jabbari R, bin Ali N, Petersen K and Tanveer B 2016-05-24 What is DevOps? a systematic mapping study on definitions and practices *Proceedings of the Scientific Workshop Proceedings of XP2016* XP '16 Workshops (Association for Computing Machinery) pp 1–11 ISBN 978-1-4503-4134-9 URL `https://doi.org/10.1145/2962695.2962707`

[26] The DataOps Manifesto https://www.dataopsmanifesto.org/ URL `https://www.dataopsmanifesto.org/`

[27] Dörnenburg E **35** 71–75 ISSN 1937-4194 conference Name: IEEE Software

[28] Artač M, Borovšak T, Di Nitto E, Guerriero M and Tamburri D A 2016 Model-driven continuous deployment for quality devops *Proceedings of the 2nd International Workshop on Quality-Aware DevOps* pp 40–41

[29] Meissner R and Junghanns K 2016 Using devOps principles to continuously monitor RDF data quality *Proceedings of the 12th International Conference on Semantic Systems* pp 189–192