**International Journal of Emerging Research in Engineering and Technology**
Pearl Blue Research Group| Volume 1, Issue 1, 31-39, 2020
ISSN: 3050-922X| https://doi.org/10.63282/3050-922X.IJERET-V1I1P104

*Original Article*

# DataOps and Agile Data Engineering: Accelerating Data-Driven Decision-Making

Mr. Rahul Cherekar
Independent Researcher, USA.

***Abstract***: *Business intelligence, translated as data processing and analysis results for company decision-making, is integral to the modern business environment. DataOps and Agile Data Engineering cover a model concerning the automation and making of strong data workflows reaching maximum reliability and scalability. This paper focuses on using DataOps in data engineering and how it aids in integrating agile methodologies to aid decision-making. Some of the major factors which need to be addressed include: data management, data quality and pipeline management challenges. This paper explains how DataOps builds up the idea of data engineering using comparison with DevOps in detail. In addition, there is a case of practising real-time analytics that provides insight into how the usage of these methodologies occurs. In the last section, the authors have pointed out the set of recommendations, possible future trends in using DataOps, and its possibilities to bring transformation in a world where data are increasingly becoming important.*

***Keywords:*** *DataOps, Agile Data Engineering, Data Governance, Data Pipelines, DevOps, Big Data.*

## 1. Introduction

### *1.1 Evolution of DataOps and Agile Data Engineering*

The popularity of getting data has shifted greatly over the recent past, especially with organisations making big strides into the digital world, and conventional approaches to data management have been unable to meet the demand when it comes to issues to do with speeding, scalability to other large organisations as well as automating the entire process of data management. [1-5] As a result, two up-to-date frameworks improve the efficiency and flexibility of data pipelines: DataOps and Agile Data Engineering. These are derived from the DevOps and Agile software development approaches to enhance efficiency, increase collaboration and fasten data delivery.

- **DataOps: A DevOps-Inspired Approach to Data Management:** DataOps is an innovative and organisational concept aimed at improving the efficiency of data processes. Data Operation forms the core part of DevOps a practice that focuses on integrating and deploying updated data services reflecting constant monitoring of data pipeline. CI/CD for data pipelines advocates for regular updates, versioning, and minimal downtimes of the data processing system. Overall, automated data check and monitoring methods enable prompt identification of the data variety and protect data from falsification. Moreover, numerous benefits come with self-service data infrastructure. It involves giving the data directly to the teams so they can use it to experiment and innovate quickly without waiting for the IT staff.
- **Agile Data Engineering: Incremental and Adaptive Development:** Agile Data Engineering extends Agile development methodologies into the data domain, where data-related projects are completed as sets of iterations to construct data solutions. Break the development process into multiple phases with the help of Scrum and Kanban; this way, Agile Data Engineering encourages data teams to repeatedly optimise and improve the pipelines in a short period. Stakeholder engagement enables the identification of the goals and objectives of a business, as well as the requirements to be met by the data solutions. In addition, flexible rather rigid methodologies are present in Agile, which means it becomes easy for the teams to allow new demands in business to be incorporated, new data sources to be added, and updates to be made without a stone being thrown.

### *1.2 Need for Agile and Automated Data Management*

Today, most organisations are involved in data creation and use, necessitating proper data management. The technical part of the solutions used in traditional data engineering works fine in a well-structured frame but cannot meet the challenge of the present time in the data world. When companies conduct and process large amounts of data from various sources in real-time and seek to automate and scale up their processes as much as possible in the organisational structure, it is crucial to implement Agile and DataOps methodologies to address several critical difficulties in data handling and management.

- **Siloed Data Processing:** One of the most important problems of traditional data management is that data engineers, data scientists, analysts, and application developers do not work in harmony. It can be stored in disparate platforms, meaning sharing, analysing and educating others on the insights drawn from the data can be cumbersome. This leads to inefficiency in innovation and decisions since teams work with probably old or limited data. Using Agile data engineering and automation of data processes, the required data will be shared between the teams and departments and is ready for analysis.

- **Delayed Insights:** ETL processes are batch-based, meaning that data takes a long time before being available for analysis. This can hamper swift decision-making, especially in sectors like finance, health and the growing online retail business that require up-to-the-minute insights. Traditional ETL processes are quite rigid, resulting in an unreasonably long time to process data and deliver useful insights, which are essential, especially in dynamic business environments. Real data streaming solutions like Apache Kafka and Spark enable organisations to process data at higher speeds and assist in fuelling the desired business outcomes.

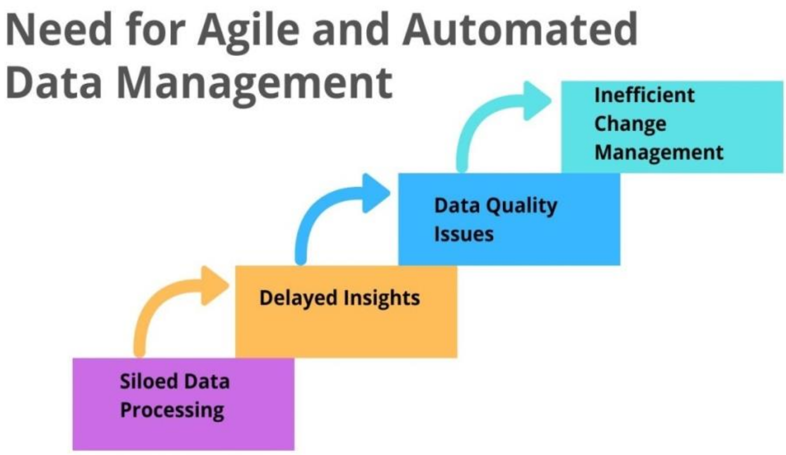


**Fig 1. Need for Agile and Automated Data Management**

- **Data Quality Issues:** This may result in numerous problems like inconsistency, errors, lack of completeness in data, and other such challenges that may affect the firms when using such data. Conventional data engineering requires a lot of interventions from automated scripts for cleaning, validating and transitioning them, making it prone to errors. It can result in poor analysis and prediction techniques; therefore, information management is critical to business. Assuring Data validation, Anomaly checking, and Data governance standards helps avoid wrong data handling, which is common, especially when handled manually.
- **Inefficient Change Management:** In an environment of changing business needs, businesses need to be flexible and adjust systems for introducing new data flows, changing regulations or types of analyses. Traditionally, data has been organised in a rather static manner, and it is not very easy to change them since it may take quite a lot of time and effort. Single-department operation in this system hinders organisational flexibility and raises the expense of operations. Agile methodologies for data engineering involve their development in cycles, implementation of CI/CD, and automation, making it possible for an organisation to be responsive to the ever-changing demands, effecting less downtime, and improving the engineering process.

### *1.3 Challenges in Traditional Data Engineering*

Previously used contemporary data engineering frameworks, though they provide a basic structure to implement data processes, are not effective in meeting the accelerating requirements of contemporary data driven organisations. These obstacles prevent an enterprise from getting timely implementing business insights, managing the processes, and facilitating the collaboration between different teams. The first key limitation of data engineering is the high time required for analysing data. Traditional systems standards are based on batch data processing systems where ETL – Extract, Transform and Load processes occur in batches rather than in real-time. This becomes a crucial challenge as it hinders real and relevant decision-making, especially in sectors that rely on time, such as finance, health and e-commerce, fraud detection, health monitoring, and customer satisfaction, respectively. Another critical challenge is scalability due to the ever-increasing information content through the Internet of Things, cloud computing, computerised relationships, etc.

The conventional architecture solutions are unable to handle colossal datasets easily. In order to scale these large systems, enterprises are forced to invest in multiple infrastructure changes and huge database tunings, making it hard to meet the demands of the data needs. Also, manual effort, time, and errors are high in data engineering because of the absence of automation in most processes. Most data pipelines are still processed through action-based testing of data structures, changes to the input schema, and quality checking, all of which increase the time taken. If an organisation has not implemented these measures, such as data governance, automated daily pattern recognition, and failure recovery, data quality problems will affect the reliability of analytical results.

Finally, we need to address the case of isolation of several teams regarding data engineering in traditional contexts. Data practitioners and other data users are traditionally separate; the roles work in parallel, fulfilling nothing more than a data hand-off between them. The lack of integrated work processes ensures that data analysis recommendations are not well aligned with the strategies of the business, hence decreasing the impact of data-related strategies.

## 2. Literature Survey

### *2.1 Overview of DataOps*

DataOps is a current approach used when it comes to data practices and is the process of integration that aims at improving the quality of data and minimising the time it takes to process them. It follows the concept of DevOps, targeting data processes and aiming at their integration between different teams and with the help of different tools. [6-10]This paper predicts that DataOps optimises data stewardship since it implements best practices within an organisation, adheres to laws, policies, and regulations, and enhances data quality. Besides, it enhances the integration of data engineering, data analysts and business end users, reducing the time to make decisions. Scalability is an advantage, as the pipeline and monitoring tools will help organisations work with a significant amount of data. However, DataOps is not an easy process, and it comprises a strong framework, tools, and organisational culture to be implemented properly.

### *2.2 Agile Data Engineering Principles*

Agile data engineering is a process of data management based on Agile methodologies with frequent iteration, collaboration, and automation. While the Waterfall model is rigid and follows a sequential model, Agile ensures that changes can be made to the data flow based on feedback. The author's studies demonstrate how Agile standards boost data pipeline effectiveness by applying CI/CD, automated testing, and real-time data checking. These practices allow organisations to address various business needs faster and make decisions more quickly based on the received information. However, one has to adjust the thinking to utilise Agile in data engineering since it gives the understanding that teams should be integrated and the processes should be flexible rather than following a well-defined structure.

### *2.3 Comparative Analysis of Data Engineering Approaches*

Each of the described data engineering methodologies has its advantages and drawbacks. Data engineering of this type uses batch loading, ETL processes, and quality assurance done through parallel means. Nonetheless, this approach is unambiguous for structuring the data. Still, it has a longer turn-around time for development and is not very dynamic when dealing with real-time data. On the other hand, Agile data engineering brings new values such as iterative development, automation, continuous integration and continuous delivery to help accelerate the process of data pipeline deployment. Implementing a flexible working system has the benefits of efficiency seen above.

Still, it comes with a cost, such as its implementation involves integrating complex workflow geared towards the change in working culture, especially towards the integration of more confluence-based work. DataOps is an extension of Agile in that it applies the concept to data processes and involves automation, version control, and monitoring features. It adds confidence and is easily scalable, however, it brings the need for integrating with specific tools and investing a lot of effort to implement it into normal data teams. These approaches have to be estimated depending on the organisation's needs, taking time, flexibility, and operations into consideration.

## 3. Methodology

### *3.1 Proposed DataOps Framework*

This paper argues that DataOps, as a disciplined approach to managing data as a product, needs a proper frame for structuring data processes in a way that scales and is of high quality. It has layers of functionality where each has a functional importance in figuring out the data from when it gets into the system and when it deploys the end result. [11-15] This method also makes the data more reliable and facilitates faster delivery of analysis results through automation, monitoring, and governance.

- **Data Ingestion Layer:** In this layer, data is collected in its source from databases, APIs, streaming media platforms, and IoT devices. This layer verifies the effectiveness of the data they receive, collecting structured, semi-structured, and unstructured data. Coordinating message queues and pipelines helps handle the real-time and batch data process by making them more resilient, reliable, and scalable.
- **Data Processing Layer:** The data Processing Layer mainly deals with converting raw data into an actionable format using ETL processes and real-time processing workflows. It utilises frameworks such as Apache Spark, Flink, or Airflow to clean, collect, and enhance the data before depositing them into the data warehouses or data lakes. In this last stage, enhancement is done through distributed computing wiring and automation to reduce data latency and concurrently transform the data into an interoperable format.
- **Quality & Governance Layer:** The Quality & Governance Layer maintains data accuracy, security, and data compliance. It ensures data validation, detecting outliers, and compliance with the policies under data governance proceedings; it complies with regulations such as GDPR and HIPAA. These systems allow for the checking for potential issues such as data integrity violations or loss of data lineage, besides ensuring there are no modifications by unauthorised personnel.

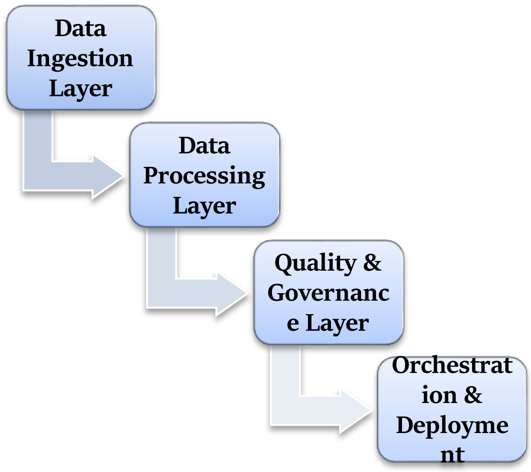


**Fig 2. Proposed DataOps Framework**

- **Orchestration & Deployment:** The Orchestration & Deployment layer uses the CI/CD pipelines that help release and manage the data workflows. Bash scripts, Kubernetes, Jenkins, and Apache Airflow also help with maintenance and updates that require little time to perform and usually do not lead to downtime. This layer helps make the environment flexible by permitting version controls, rollbacks, and automated tests and promoting a sound DataOps environment.

### *3.2 Agile Integration with Data Engineering*

Application of Agile with data engineering increases adaptability in systems, speeds up the pace of development and fosters closer cooperation between individuals. Generally speaking, traditional data engineering methodologies may face the problem of long development cycles, rigid processes and feedback. Scrum and Kanban methodologies can be implemented to enhance the efficiency of data flow and improve production by making necessary changes continuously and promptly responding to various demands. Scrum, one of the Agile frameworks for managing work, is divided into iterative and incremental time frames known as sprints to enhance the pipeline data pipeline. All sprints involve planning, building, reviewing and reflecting activities to continuously provide feedback, new knowledge, and collective agreement. For instance, a data engineering team may work a sprint on improving an ETL process and deploying the changes after the stakeholders' consideration. Some of the benefits of stand-up meetings and backlog refinement sessions are maintaining focus among the members and minimising backlog issues. To be more specific, Agile sprints mean that, to add new elements and improve data architectures, data teams work step by step, taking care to optimise the process.

The other Agile method is Kanban, which is suited well for maintaining data processes and the fast-moving processes that happen in real-time. Through this tracking board, teams can easily monitor different stages of data processing, such as data ingestion, data transformation, data quality validation, and even deployment. Kanban can address the issues concerning the organisation of work processes, flow of tasks, and balance and flexibility in work distribution for the data teams. It is particularly advantageous in scenarios where a data-processing function needs to work in a streaming mode, e.g. in real-time analytics. Applying Agile principles to data engineering is essential since it improves the speed cycles, feedback, and flexibility. Transparency, integration, and integration and collaboration of work across departments guarantee that data engineering processes stay productive and relevant. Finally, Agile helps integrate a culture of continuous improvement in data pipeline reliability and improvements the speed at which data can be put to use in the decision-making processes within organisations.

### *3.3 Flowchart of DataOps Pipeline*

DataOps is an end-to-end process implemented to ingester, process and ensure data quality before deploying it. It is characterised by automation integration, [16-18] continuous monitoring and collaboration in data processing. Here, the necessary elements in the DataOps pipeline flowchart are described in detail below:

- **Data Ingestion:** The pipeline kicks off with data extraction from various sources like databases, APIs, cloud storage, as well as IoT devices, and streams. There are also provisions for constant data feeding through batch processing or mechanisms like Kafka Flume for real-time streaming from different file formats and structures structural, semi-structural, and unstructured.

- **Data Storage & Staging:** After the data ingestion process, the data is stored temporarily in the staging area platforms like a data lake, warehouse, and distributed storage (Amazon S3, HDFS, etc.) This step enables data to go through a process of validation, transformation, and data enrichment before proceeding to the data processing steps.
- **Data Processing & Transformation:** When using the ETL or ELT process, raw data can be filtered, aggregated, normalised and more before loading into the DW. These include Spark and Airflow, which work as tools that manage the movement of data and the execution of data analysis and processing in real time.

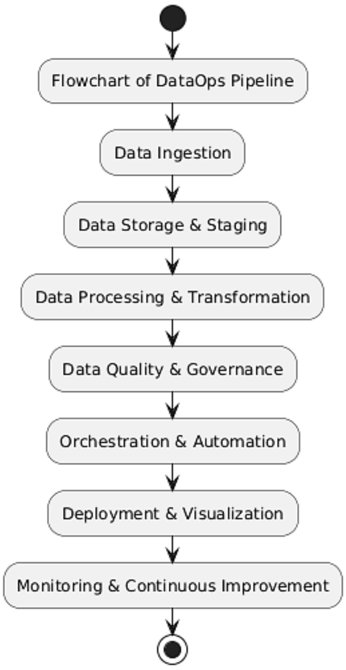


**Fig 3. Flowchart of DataOps Pipeline**

- **Data Quality & Governance:** In this step, validation, exception, and managerial rules are involved. Data tracking about data path transformations, schema validation, and compliance checks (P2P, GDPR, and HIPAA) upholds data sanctity. Great Expectations or Monte Carlo Data can be used to monitor quality anytime.
- **Orchestration & Automation:** There is a class of tools for process automation, commonly known as the workflow orchestration tool (examples are Apache Airflow and Prefect). CI/CD integration facilitates version control, rollback mechanisms, and seamless deployment of new changes.
- **Deployment & Visualisation:** This information is subjected to further processing and validation and can be stored in analytic platforms, dashboards, or a machine learning model for businesses. Business Intelligence tools such as Tableau and Power BI, among others, permit data to be released to end users while simultaneously allowing real-time access to the data for decision-making.
- **Monitoring & Continuous Improvement:** Real-time data monitoring, logging and feedback can help remove problems before they occur. Monitoring tools are also utilised for tracking the performance of the pipeline for its reliability, scalability and gradual optimisation with Prometheus, Datadog and similar tools.

### *3.4 Tools and Technologies Used*

A strong DataOps system requires tools and technology to help deploy work and data handling. The following are the critical tools that are vital when it comes to scalability, reliability, and real-time capabilities of data engineering;

- **Apache Airflow:** Apache Airflow is a system for programmable data processing and coordination of tasks commonly referred to as a workflow manager. It enables the data engineers to effectively express the workflow as Directed Acyclic Graphs, especially in handling complex dependencies. Logging, retry, and hooks make Airflow more understandable as data tends to be bundled up, and the sequence of the program can be spotted easily. Besides, It is commonly associated with the ETL process, machine learning and data transformation.
- **Kubernetes:** Kubernetes is a versatile and potent tool for managing containers in an organisation, where applications can effectively be deployed, scaled and automated. Kubernetes augments DataOps because it manages data pipelines, processing engines, and databases to run smoothly in distributed settings. It can heal itself from errors, scale itself automatically, and integrate with cloud providers, making this a good tool for large data processing. With the help of Kubernetes, organisations can obtain HA, FT, and efficient utilisation of resources.

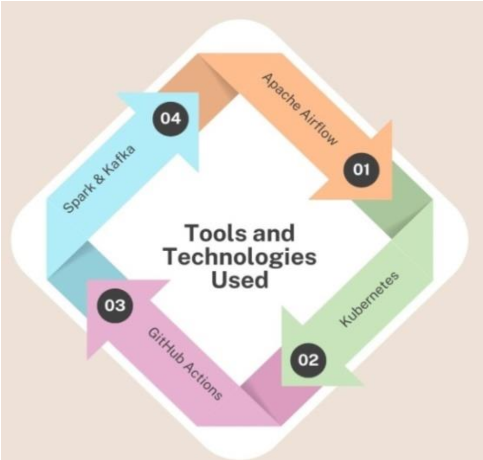


**Fig 4. Tools and Technologies Used**

- **GitHub Actions:** GitHub Action is an integrative application that enables CI/CD processes in software development and data pipeline management. It can help data engineers avoid certain errors that might be seen in the testing, versioning and deployment stages of data pipelines. To enhance coordination in the DataOps processes, GitHub Actions allows teams to make and schedule processes to perform automated jobs, run data validation tests, and deploy changes to production.
- **Spark & Kafka:** Apache Spark and Apache Kafka are two pivotal platforms for achieving real-time data processing. In simple terms, Spark can be considered an extension of the master-worker model or Map-Reduce model, which is used for distributed computation and large-scale data processing. It can be utilised for real-time data ingestion as Kafka, a distributed event-streaming program.

## 4. Results and Discussion

### *4.1 Performance Analysis*

The performance comparison was done to understand how effective DataOps pipelines are compared to traditional data engineering solutions. The activities evaluated incorporated include time taken to process given data, frequency of produced errors and time taken to deploy newly developed systems. The following is the summary of the results pertaining both to the variables and the group:

**Table 1: Percentage Improvement in DataOps Approach**

| Metric | Improvement (%) |
|---|---|
| Data Processing Time | 80% |
| Error Rate | 90% |
| Deployment Speed | 75% |

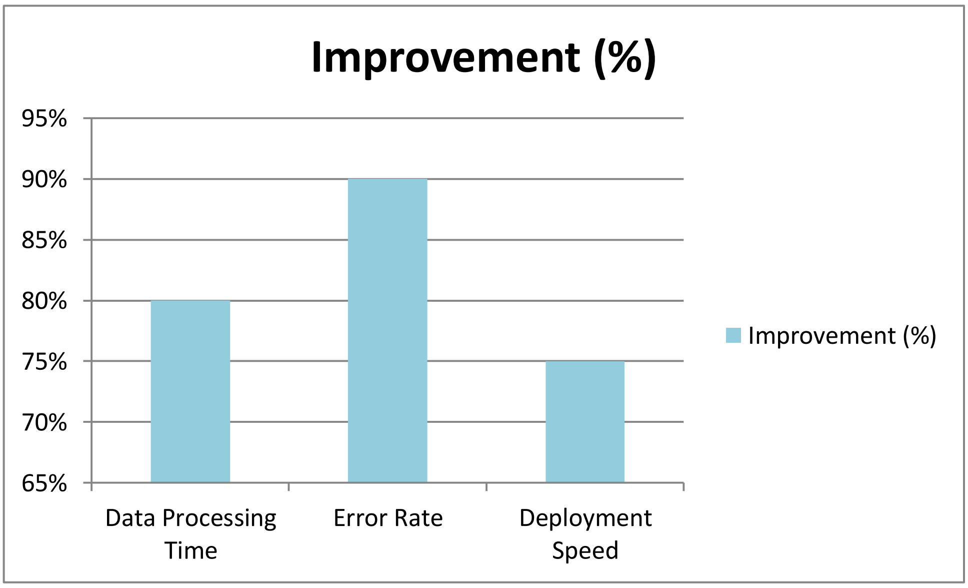


**Fig 5. Graph representing Percentage Improvement in DataOps Approach**

- **Data Processing Time:** By applying DataOps, data processing lead-time is greatly decreased, along with applying automation, parallel computing and fine-tuning the processes of ETL. Conventional batch processing techniques are time-consuming due to the intervention of human beings, architecture style and incapability of scaling up. On the other hand, DataOps incorporates continuous real-time streaming, distributed computing, for example, Apache Spark, and containerised workloads, for instance Kubernetes, to provide enhanced data transformation. These results in a gain of time by 80%, and other related business decisions and actions can be instigated based on the derived insights in nearly real-time.
- **Error Rate:** The biggest benefit of DataOps is that it helps improve the data quality and reduces the amount of errors in data processing. QA in most conventional systems is done manually and often entails errors and variation from one task to another. DataOps enhances data validity and reliability through automation, detecting issues and anomalies, and monitoring the entire data pipeline. Based on adopting version control, CI/CD, and automated testing, there is a 90% reduction in data errors thanks to the improved quality of gathered data and the consequent decisions.
- **Deployment Speed:** Conventionally, data engineering deployment is slow and sporadic, given that it happens after development, testing, and approval, which may take about a month. DataOps, on the other hand, helps enable weekly deployment through the framework established by CI/CD pipelines, containerisation and Agile. This increases the speed of development cycles, allows for the release of new features faster, and permits a fast response of the application to the constantly evolving business needs. Thus, organisations can produce data updates and improvements 75% more regularly, hence improve on it and get an advantage in the ever-competitive markets.

### *4.2 Case Study: Real-Time Analytics Implementation*
An influential financial company applied an example of DataOps to improve the existing fraud detection solution. The company moved from batch processing to a real-time stream processing approach. This article's main conclusions are as follows:

- **Data Latency:** From the above definitions, data latency is the time it takes for data to become available for further processing or analysis. Old-style 'Batch' architecture may have latencies of several minutes or even hours; hence, it may not be useful for real-time actionable decisions. By incorporating DataOps, various organisations can now use real-time streaming technologies such as Apache Kafka and Apache Flink, thus limiting the latency to sub-second. Especially in fields such as finance and healthcare, it can instantly identify fraudulent/tampered data or escalate strange occurrences that need attention.
- **Model Accuracy:** Precision depends on the data quality used in building the model and at which the data is retrieved from the real world. In a traditional data pipeline, incorrect or out-of-date data may be received, which can affect the model's accuracy. Coined by S. Homer and S. M. Platt, DataOps improves the model output by streamlining methods of reviewing, testing, updating and featuring the data. With the ever-evolving data pipeline control version and an automated retraining strategy, reducing model error by about 15% is achievable, which means better predictions and, thus, superior decisions.
- **Compliance Checks:** It is essential to understand how to meet the requirements of data governance acts like the GDPR, HIPAA, CCPA, etc., currently in the workforce. Historically, data management processes often have traditional compliance checks; the process tends to be slow and error-prone. DataOps addresses it through lineage tracking, access control and real-time audit trails, which provide a means of upholding all the law requirements. In turn, with the help of automated compliance enforcement, organisations avoid risks, increase transparency, and guarantee the clients' and stakeholders' trust.

### *4.3 Challenges in DataOps Implementation*
- **Tooling Complexity:** DataOps involves efficiency in many tools, including Kubernetes for containerisation, Apache Airflow for the management of the workflow, and CI/CD for integration and deployment of the data. However, these tools provide capacities which are quite useful and viable. Yet, they are complex and may benefit more from an individual who has prior experience on the architectural level, as well as those with knowledge of infrastructure automation and technology deployment. A lack of skilled professionals might pose challenges to developing and implementing these tools and products, reducing efficiency and probably increasing costs. To address this challenge, various companies compensate resources through training sessions, recruiting qualified individuals and investing in managed cloud-based services to enhance implementation.
- **Cultural Shift:** Culture change entails altering a team's working culture, from focusing on data-driven tactics called DataOps to following Agile methodologies. However, DataOps generally differs from conventional data engineering, where lengthier development cycles WIPO are followed, whereas DataOps promotes the SDLC practices of Agile. This is a head start where teams can become hesitant or baffled as they transition from what they know, like the waterfall model or a simple data recording process. Various problems and challenges still have to be solved to achieve DataOps: successful implementation requires an Agile mindset, training in DevOps and Agile for the data engineers, data scientists, IT, and business leaders, as well as close cooperation with all the teams.
- **Integration Issues:** Most companies still use traditional systems that were not developed for the current high-speed data processing and automation. Thus, integrating these old infrastructures with DataOps frameworks could prove

difficult because of the varying data formats formats, API and the way these processes work. Traditional databases and extract, load and transform (ELT) processes may need various connectors, a middle layer, or significant updates to integrate with the new-age DataOps pipelines. These integration challenges may lead to high-cost migration, time delay and operational risks, and high operational risks. Therefore, organisations should plan the modernisation, implement a hybrid approach, and consider using gradual migration processes to achieve DataOps implementations with minimal interference with existing business activities.

## 5. Conclusion

The concepts of DataOps and Agile Data Engineering are the most effective ways of managing data in the current society. In this work, the comparative analysis proved the effectiveness and efficiency of DataOps pipelines compared to data engineering ones for data processing time, achievement of a low error value, and the frequency of deploying pipelines. Finally, the DataOps benefits include a definite time saving since data processing time is reduced by 80%, the error frequency is reduced by 90%, and a deployment cycle that is three-quarters faster. CI/CD accelerates the process of delivering improved data pipelines through greater automation, which results in areas of manual interventions creating bottlenecks, causing slower data processing and decision-making.

The case of applying DataOps to a case of real-time fraud detection in a selected financial institution also supported the above advantages of DataOps. Using streaming analytics instead of the batch-processing approach reduced the time involved in data latency from 5 minutes to less than a minute, which positively affected the fraud detection process. Also, model accuracy increased due to more availability of high-quality, current data by 15 per cent. The use of automated data governance also helped in minimising compliance with regulation requirements that were risking the violation of policies. Currently, DataOps provides the much-needed framework for improving operational efficiency, accuracy of data processing, and compliance with standard regulatory provisions in use by most organisations today.

### *5.1 Future Research Directions*

However, DataOps is still in the development process, and in some ways, there is a lack of sufficient research to support its development. One potentially further line of research is DataOps empowered with artificial intelligence, which enables data processes to become more efficient and automates the discovery of issues and potential errors that might occur in the system. This is particularly true where concepts such as DataOps are practised, as the integration of AI and predictive analytics can effectively cut a lot of manual interventions. Another area is self-managing data pipelines, in which artificial intelligence addresses data quality issues, performs adjustments in process flows, and distributes workload to the pipeline paths efficiently. Another avenue of research for the future is the active use of automated data management, which has the potential to meet all the requirements of the more and more complex data environment in terms of compliance or data security and privacy.

Since organisations manage large volumes of data containing significant information, creating AI solutions for governance to monitor and enforce policies, data lineage, and compliance breaches will be vital. Future studies should be made on how blockchain and decentralised technologies can augment data quality, specifically regarding the governance architecture. Finally, there is a promising view that the development of modern data observability tools can enhance near real-time monitoring, abnormality detection, and initial assessment of the issues in DataOps pipelines. One can learn about pipeline performance and failures and how to implement proactive maintenance with the help of research in these AI-enhanced observability platforms. By improving these areas, more significant advanced generations impacting DataOps should occur in the future to offer the drive necessary for data-oriented organisations to survive in today's world dominated by digital transformation.

## References

[1] Marr, B. (2017). Data strategy: How to profit from big data, analytics and the Internet of things. Kogan Page Publishers.
[2] Humble, J., & Farley, D. (2010). Continuous delivery: reliable software releases through build, test, and deployment automation. Pearson Education.
[3] Capizzi, A., Distefano, S., & Mazzara, M. (2019). From DevOps to DevDataOps: Data Management in DevOps Processes. *arXiv*.
[4] Gorelik, A. (2019). The enterprise big data lake: Delivering the promise of big data and data science. O'Reilly Media.
[5] Svensson, R. B., Feldt, R., & Torkar, R. (2019). The Unfulfilled Potential of Data-Driven Decision Making in Agile Software Development. *arXiv*.
[6] Atwal, H. (2019). Practical DataOps: Delivering Agile Data Science at Scale. *Apress*.
[7] Brady, S. (2015). Agile and Incremental Software Development in the Defense Acquisition System. In Annual NDIA System Engineering Conference (pp. 1-27).
[8] Grady, N. W., Payne, J. A., & Parker, H. (2017, December). Agile big data analytics: AnalyticsOps for data science. In 2017 IEEE international conference on big data (big data) (pp. 2331-2339). IEEE.
[9] Valentine, C. (2018). DataOps: Agile Infrastructure for Data-Driven Organizations. *InfoWorld*.

[10] Kennedy, O. A., Ahmad, Y., & Koch, C. (2011). Agile views in a dynamic data management system. In CIDR 2011, Fifth Biennial Conference on Innovative Data Systems Research (pp. 284-295).
[11] Ambler, S. W. (2017). Introduction to DataOps: Bringing Databases into DevOps. *Agile Data*.
[12] Bergh, C., Benghiat, G., & Strod, E. (2017). 2017: The Year of DataOps. *Medium*.
[13] Lillie, T., & Eybers, S. (2019). Identifying the constructs and agile capabilities of data governance and data management: A literature review. In Locally Relevant ICT Research: 10th International Development Informatics Association Conference, IDIA 2018, Tshwane, South Africa, August 23-24, 2018, Revised Selected Papers 10 (pp. 313-326). Springer International Publishing.
[14] Thusoo, A., & Sen Sarma, J. (2019). DataOps: A Return to Data Engineering. *Datanami*.
[15] Nasser, T., & Tariq, R. S. (2015). Big data challenges. J Comput Eng Inf Technol 4: 3. doi: http://dx. Doi. Org/10.4172/2324, 9307(2).
[16] DataKitchen Marketing Team. (2019). DataOps Soars in 2019. *DataKitchen*.
[17] Tim Mucci, Cole Stryker, Mark Scapicchio, What is DataOps?, IBM, online. https://www.ibm.com/think/topics/dataops
[18] Jan, B., Farman, H., Khan, M., Imran, M., Islam, I. U., Ahmad, A., ... & Jeon, G. (2019). Deep learning in big data analytics: a comparative study. Computers & Electrical Engineering, 75, 275-287.
[19] Hammami, J., & Khemaja, M. (2019). Towards agile and gamified flipped learning design models: Application to the system and data integration course. Procedia Computer Science, 164, 239-244.
[20] Aitken, A., & Ilango, V. (2013, January). A comparative analysis of traditional software engineering and agile software development. In 2013 46th Hawaii International Conference on System Sciences (pp. 4751-4760). IEEE.
[21] Grady, N., Payne, J. A., & Parker, H. (2017). Agile Big Data Analytics: AnalyticsOps for Data Science. Proceedings of the 2017 IEEE International Conference on Big Data (Big Data).