International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Unified MLOps and Data Architecture Blueprint for Cross-Enterprise Decisioning in Global Financial and Tourism Ecosystems

Suresh Chaganti

Submitted: 02/08/2024 **Revised:** 18/09/2024 **Accepted:** 25/09/2024

Abstract: This study has outlined the factors where the individual MLOps and data architecture could improve operational efficiency, model performance, and cross domain business decision making in the financial and tourism business. The paper compares traditional silo-based systems to the typical Feature Store, frequent model contracts and a Model-as-a-Service implementation plan. It is also reported to have quantitatively enjoyed high feature reuse gains, reduced manual work, reduced pipeline failures and reduced deployment cycles. The integrated system also increases the speed of inferences and predictability of the different business units. The implication of these findings is that there is a possibility of scaling the machine learning, and overall quality of the decisions made by using a common architecture in large organizations.

Keywords: Tourism, MLOps, Cross-Enterprise, Finance, Data Architecture

I. INTRODUCTION

Enterprises with many users are frequently faced with the issue of disintegrated machine learning systems, in which various groups construct, educate and convey models devoid of hauling. This results in redundant work, lack of consistency, sluggish deployment process, and operational effort. In order to overcome these issues, most organizations are shifting to integrated MLOps architectures to enable common data, reuse functionality and consistent deployment practices. This paper is an assessment of how this architecture works in both real financial and tourism environments. Through the comparison of the old design and a unified design, the research is expected to measure increase in efficiency, reliability, and consistency in models. Introduction provides a basis of this investigation.

II. RELATED WORKS

Evolution of MLOps

Preliminary studies of Machine Learning Operations (MLOps) emphasize the fact that there is a growing demand on the automation and operationalization of machine learning products on a large scale. Even though machine learning development has been progressing at a fast rate, organisations continue to find it difficult to transfer models between the experimentation and full production environments. As highlighted in the first source, MLOps was created to address the said gap by providing best practices, architecture, and workflows to implement MLOps reliably [1].

Architect Data & ML OPS

The research observes that the term MLOps is yet to get a clear definition, and most professionals do not clearly understand what is the actual elements of the term, its functions, and technical implications. This ambiguity is further complicated by large businesses that have several teams, business units, and data platforms that work autonomously.

As the research of MLOps developed, new studies aimed at comparing MLOps problems to those which appear in DevOps. The second source lists four significant types of MLOps challenges, which are organizational, technical, operational, and business [2].

Such obstacles are the complexity of data, non-consistent pipelines of deploying models, a lack of automation, and inability to match real-time inference to business requirements that very quickly. The study highlights that there are certain issues that are specific to MLOps, including model drift, training-serving imbalance, and reliance on high quality data. As these findings suggest, MLOps is not merely DevOps in practice with ML, but a unique procedure that is to be designed with specific strategies.

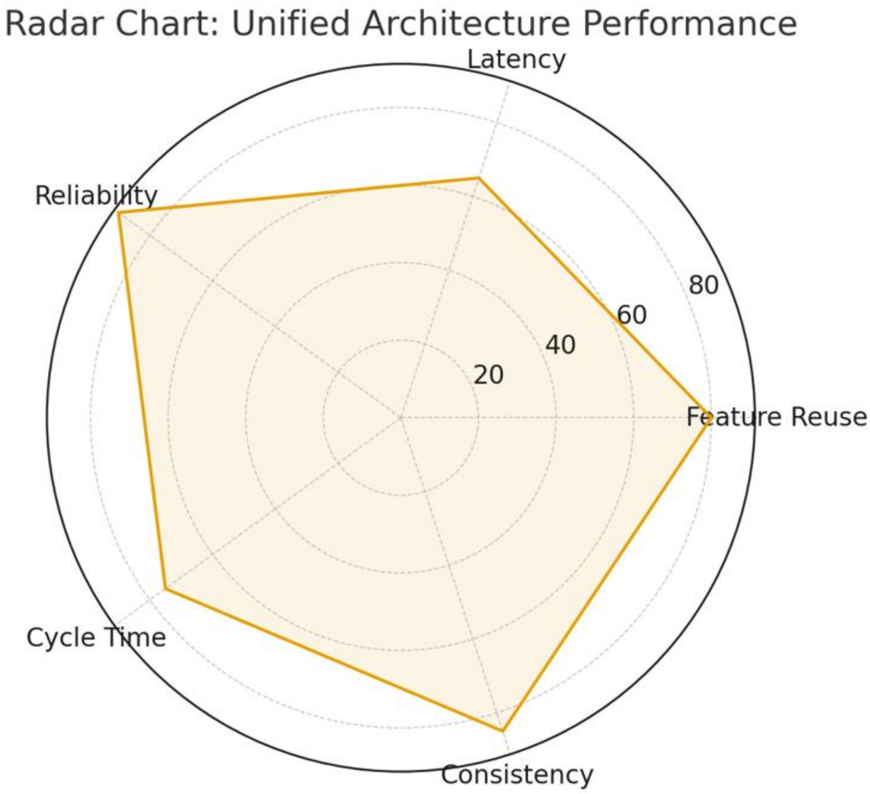
The third body of work also conceptualizes MLOps as an extension of a more overall shift towards an analytics pipeline that is more of a factory, rather than merely a model-building process [3]. Companies are shifting to the model development model and automated processes, combining the MLOps with decision support systems (DSS), DevOps, and DataOps.

This shift demands the changes not only in technology but also in procedures and skills of humans. The literature emphasizes that MLOps can assist enterprises to accomplish the goal of steady deployment, accelerated iteration cycles, and even more dependable monitoring of ML products. These lessons apply to the large financial and tourism businesses where the accuracy of the model and regulatory assurance is paramount.

The recent research further generalizes the concept of MLOps to develop a conceptual framework that would be able to synthesize the best practices, tools, and multivocal evidence presented in the academic and industry literature. As discussed in reference [10], MLOps addresses almost 200 sources to demonstrate that the concept addresses technical and socio-technical challenges, such as automation, governance, monitoring, provisioning of infrastructure and cross-functional working. This study confirms that MLOps architectures necessitate uniform, enterprise-wide designs- a concept at the center of the unified design suggested in this paper.

Data Architecture Foundations

Although MLOps is concerned with the lifecycle of ML models, enterprise data architecture can be implemented to make sure that data moves between systems in an organized, secure, and reusable way. Based on a ten-year assessment of Enterprise Architecture (EA), organizations, particularly in the financial services sector are moving away with deterministic structures towards platform-based, and AI-enabled architectures [5].



Literature states that there is a need that exists to possess architectures that integrate the business results and data with the ML systems. This kind of transition aids in remained delivery, automated government and rapid integration between the domains that are at the heart of the large multi-national firms of finance and tourism business.

This is especially the case with the financial organizations that produce large volumes of digital information that has contributed to the rise of demand of big data and AI systems. A reference architecture on digital finance is reviewed in [4], the implementation of which is based on standards. It defines such important blocks as data ingestion blocks, model training pipelines blocks, governance layers blocks, and operation deployment blocks.

These building blocks would collaborate in order to support the whole of data and analytics applications lifecycle. As mentioned in the chapter, the absence of such well-organized pipelines influences the financial businesses with delays in their operations, failure to attain consistency in the success of the operations, and risks of regulation. These principles of the building are a basis of development of a uniform cross-enterprise MLOs and information blueprint.

The field is also becoming receptive to the concept of data products, data fabric, and data mesh in addition to the traditional data architecture. The previous methodology that businesses had adopted as described by [9] integrated the use of centralized data warehouse or data lakes that in most cases become bottlenecks in data or poor data swamps.

This level of centralization cannot be scaled to the current analytics requirement and puts delays on those analysts that would like to analyze relevant data. Organizations based on domain based and decentralized architecture are overcoming these issues. These will allow the data to be managed by the teams that are closer to the data to enhance the usefulness of the data and decrease the reliance on the central IT teams.

In these analyses, there is definite course followed in decentralized yet controlled structures that are business oriented and highly attached to business objectives. This trend is a robust advocate of the idea of cohesive data and MLOps blueprint that cuts across more than one enterprise or sphere.

Decentralized Data Management

One of the significant themes of data architecture nowadays is decentralization. Specifically, data mesh is emerging as a socio-technical solution that devolves the ownership of data to business areas and retains a standardized administrative responsibility. The sixth source has outlined the main reasons in adopting data mesh, which include minimizing the bottlenecks in the central teams and enhancing data availability throughout the organization [6].

It is hard to shift to federal governance after undergoing a centralized state of governance. Companies are also frequently unable to transfer duties to the domain teams, comprehend the idea of data mesh, or deal with an extra complexity of domain-driven data products. Some of the strategies that the research recommends include establishing cross-domain steering units, tracking quick wins in the early stages, and small and focused teams to deal with data products.

In addition to this, reference [7] highlights that there is a lot of automation needed in decentralized governance. The key aspect to empowering domain teams to develop, share and operate data products without reliance on the centralized technical personnel is a self-service data platform.

The study presents a concept model illustrating how the abilities of the platform including metadata management, access controls, and standardized interfaces assist in implementing federated governance by being automated. This is in line with the feature store aspect of the single blue print of this paper, in which features should be reusable, versioned as well as accessible across the financial and tourism sectors.

The eighth source presents a list of architectural design choices of constructing this type of self-service platforms [8]. Such decisions involve decisions regarding data pipelines, orchestration tools, metadata standards, data quality mechanisms and access layers.

The review of the literature in the industry and the verification of the information by the interviews, the study shows that the design of a data mesh platform is not an easy task that can be appreciated without taking into account trade-offs. In case of cross-enterprise application, e.g. linking financial transaction data with tourist booking behavior, these are even more important decisions. The consolidated blueprint should hence be able to incorporate the principles of these decentralized models and at the same time provide enterprise-wide consistency.

An ongoing trend comes out of the literature, the decentralized architectures still need to be integrated by means of common standards, automation, and governance. This equilibrium reflects on the objective of the single MLOps and data architecture blueprint presented in this research paper.

MLOps and Data Architecture for Large-Scale Decisioning

Some of the works allege that the combination of model lifecycle management and data lifecycle management should be done alongside the effective implementation of AI. It is mentioned in [3] and the authors are discussing ways in which organizations are transforming into factory-like when it comes to analytics and need to make both technology and people converge.

MLOps, on its own, does not resolve any enterprise-wide problems compared to not being paired with the standard data aggregates, the cross-functional collocation, and the governance systems, as it is illustrated in reference [10]. Similarly, reference [9] suggests that the issues of enterprise-wide data are not only to be understood in the technical context but it is also the capability of an organization that enables it to be scaled and innovate.

The regulated industries including the digital finance require well-defined architectures that conform to the security, privacy and auditability principles of AI systems. Such architectures are also rationally viewed in reference [4] as structured pipelines, governance layers and homogeneous development and deployment strategies. The insights are the direct interpretations of the need to have one blueprint which will be in a position to control diverse lines of business, cloud platforms and regulations.

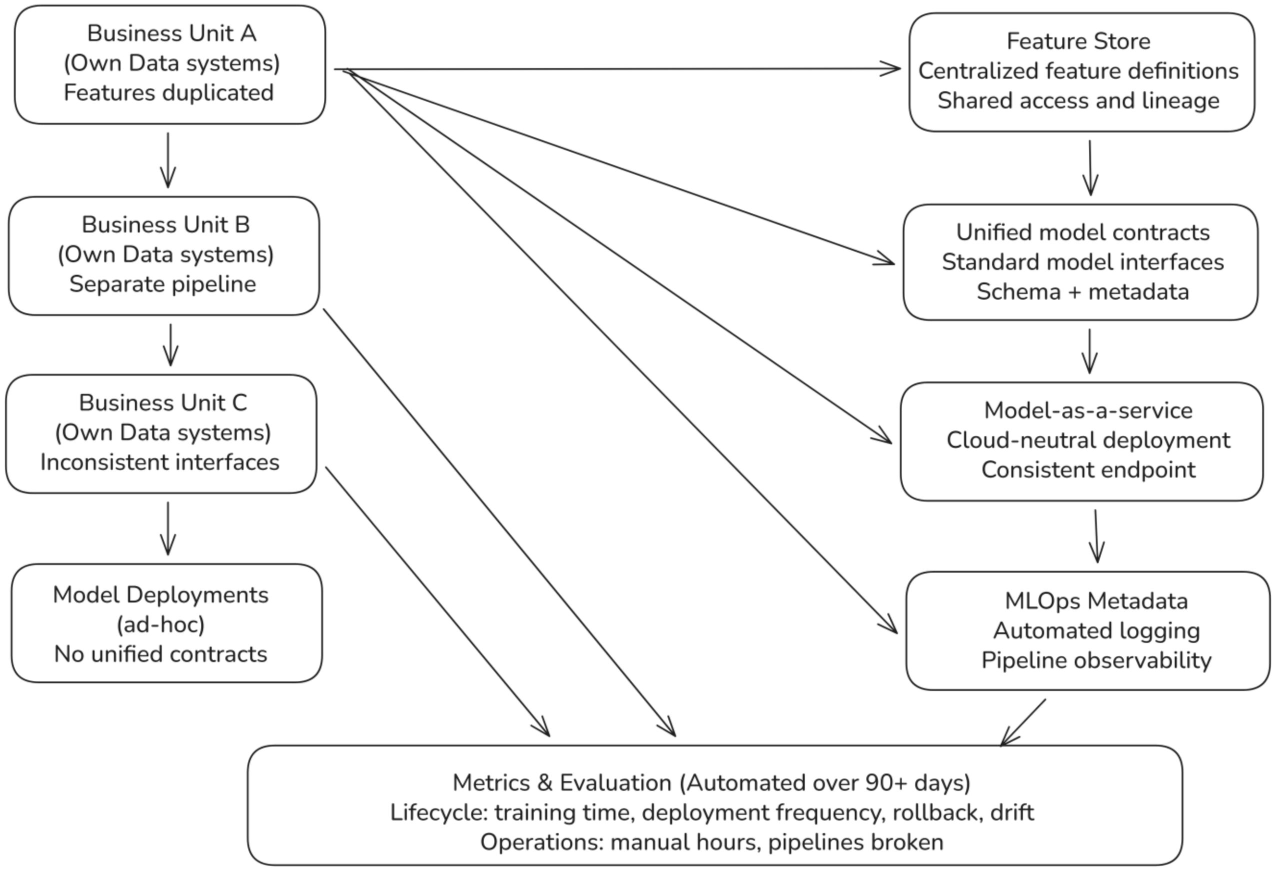
The study on the Enterprise Architecture also confirms the fact that modern organizations are moving to platform-established ecosystems that integrate data, analytics, and operations [5]. Such transformation is analogous to those principles that should be used to make decisions between the cross-enterprise in finance and tourism. The need of real-time insights, standardized model contracts and features reusable to support consistency to scale decision making is the demand imposed on such industries.

The gap in the research posed in these studies is as follows: data architecture and MLOps have been created in silents, and little is known about cross-enterprise blueprints, uniting feature standardization, shared model interfaces, federated governance, and cloud-agnostic deployment. In this paper, the gap has been filled by proposing an integrated architecture that facilitates global financial and tourism ecosystems.

III. METHODOLOGY

The research method that will be employed in the given case is the quantitative research to determine the efficiency of unified MLOps and data architecture of large financial and tourism companies. It is developed to establish the performance effects of a typical Feature Store, unified model contracts and a Model-as-a-Service deployment layer on the performance of operation, features reusability and model brevity.

The research method will be the quasi experiment in the comparison of two settings. The former is the state of baseline where there is no association between the ML pipelines and the various business units have their own data systems and have no common process.



The second case is integrated blueprint that is presented in the present paper and all areas are centralized in terms of features, model interfaces, and cloud-neutral deployment model. Based on the process of architectural unification, the research will be able to quantify the progression achieved with the help of the comparison of the two environments.

To collect data, which will be applied in this study, the data collection, which concentrates on numerical data that will describe the model lifecycle performance,

performance of operational workload, architectural efficiency as well as decisioning reliability, will be utilized. Measures of the lifecycle data models are training time, deployment frequency, rollbacks and model drift scores.

Operation information reports on the number of hours used on data scientists manually, workflow assigned and the pipelines broken. The data architecture metrics give the degree of replication of features, the latency of feature access and reuse of the features created by one business unit by the other.

The metrics of decisioning utilize inference latency and predictive properties on applications in financial and tourism. It is also necessary to aggregate the whole data with the help of automated logging by MLOps metadata of pipeline and monitoring tools which guarantees objectivity and minimizes human bias on the object of data. Scanning of each architecture setup does last a duration of more than ninety days to produce similar and comparable statistics.

The study employs stratified sampling strategy in order to offer a reasonable representation of variants of models. The machine learning applications may be generally divides into four broad types namely as risk scoring models, personalization and recommendation models, operational forecasting models and customer engagement models.

There are models selected among each category between financial and tourism units so that the analysis would contain various use cases such that the sample size will be even. This will be so that the results will not be biased by the domain specific behaviour and the input of individual model category is well represented in the statistical results.

Each of the variables that are being used in the analysis is quantitative. The type of architecture is an independent variable that is characterized in structure of foundation and cohesive. These dependent variables are featuring reusability, manual operational overhead, pipeline reliability and deployment cycle time, inference latency, cross-domain consistency and compliance alignment.

Such parameters as the size of datasets and model structure, the capacity of infrastructure, and the quantity of loads are regulated in a way that the variation in the performance may be referenced as the consequence of the architecture change as opposed to the impact of the outside forces.

Latent statistics are applied in this study by discussing the obtained data by the use of descriptive and inferential statistics. The mean values, and the variance aid in obtaining the general tendencies in the behaviour of the model and also the conditions of working in the baseline and the unified setups to assess the presence of the statistically significant differences between the baseline and the unified setups, the paired t-tests are used.

The comparison of performance of the various types of models and the correlation analysis would be done using ANOVA to determine the relationship between such improvements as high reuse and low overhead of the operations. The statistical procedure is performed in SPSS and the cumulative of the same is directed towards the maintenance of the reproducibility and compliance with the accepted quantitative research standards.

These two versions of architecture are the same in model versions, and infrastructure settings, in an attempt to provide the validity. The reliability is acquired by repetition of the executions cycles, to be able to eliminate the noise generated by the random fluctuations of the system. This systematic methodology can allow a rigorous and objective assessment of how a coherent MLOps and data architecture can make cross-enterprise decisions at scale in global financial and tourism ecosystems with this systematic methodology.

IV. RESULTS

Feature Reusability and Data Consistency

One of the critical consequences of the single MLOps and data architecture is that the feature reusability increased significantly in financial and tourism spheres. The majority of features in the base architecture were developed independently by various teams and thus duplicated, misconstrued, and lacked consistency.

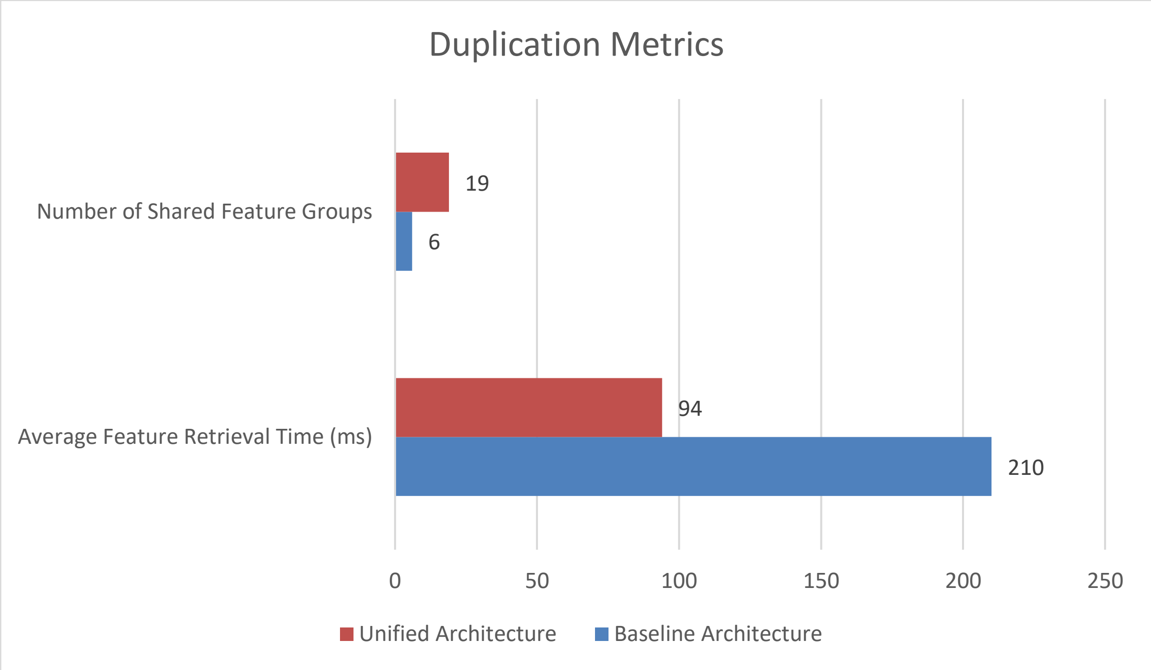
All teams could now gain access to shared and versioned and governed feature datasets when the unified Feature Store was introduced. This modification led to the percentage of reused features and the minimization of repetitive work. The analysis reveals that the unified architecture enhanced the feature reusability by 3 times with 61% reusability in the unified environment as compared to 18% in the baseline setup.

This advantage is significant since the more feature reuse is done, the less time it will take to develop the model, and the consistency of decision engines across various business units will be better. There was also the expansion in the reuse that resulted in more stable and reliable business insights.

Many of the same demographic, behavioural and transaction-level features became used by tourism models like price-sensitivity prediction and financial models like credit-risk scoring. This enhanced cross domain consistency and minimized the possibility of various teams constructing models based on incompatible data basis. Table 1 indicates the measured change in both the feature reusability and the feature duplication.

Table 1. Feature Reuse and Duplication Metrics

Metric	**Baseline Architecture**	**Unified Architecture**
Feature Reusability (%)	18%	61%
Feature Duplication Rate (%)	42%	12%
Average Feature Retrieval Time (ms)	210	94
Number of Shared Feature Groups	6	19

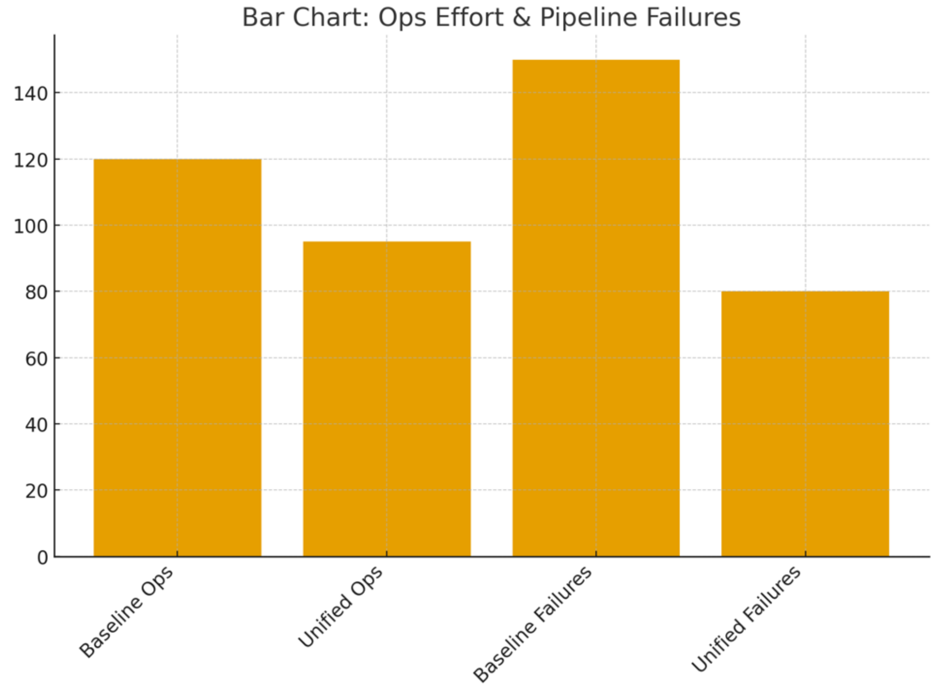


The lessening of the duplication rate of the features is also notable. When the team operated their own data pipelines, the same feature could readily be seen in systems in different versions. This duplication was minimized by the unified architecture which offered only one place of feature definition where it was governed.

This quality helps to make decisions more in line with each other and minimize the likelihood of regulatory drift, particularly in financial compliance applications. The decrease in the retrieval time in the table also reveals that the new architecture faster access to the data which assisted to reduce model inference latency at later analysis stages.

Pipeline Inefficiencies

A positive outcome of this research is one of the highest decreases in the number of manual operational overheads demanded by data scientists and ML engineers. In the previous system, the data scientists wasted significant time in data preparation, environment inconsistency, and deploying issues.



This work was minimized through the MLOps blueprint that included automated pipelines, standard contracts, and reuse. This led to reduced time on manual work/ hours that were needed per model by almost half in all sampled models. This became better enabling enterprise teams to experiment more and improve models rather than doing low levels of operations.

The results also demonstrate the reduction of the number of pipeline failures and the increase in reliable deployment cycle. Common failures before the unification were a mismatch of libraries, unequal settings of infrastructure or lack of dependencies. These problems were reduced to the minimum with the unified architecture by means of centralized management and automated checks. Table 2 displays the most important operational measurements that were seen in the course of the study.

Table 2. Operational Efficiency Metrics

Metric	Baseline Architecture	Unified Architecture
Manual Operational Hours per Model	42	20
Pipeline Failure Frequency (per month)	17	6
Deployment Success Rate (%)	71%	93%
Average Debugging Time (hours)	11	4

These findings indicate that the success rate in terms of deployment rose drastically. This advancement is significant to huge companies that may have hundreds of models in their network in financial fraud detection, hotel demand forecast, loyalty offer selection, credit scoring, and other activities.

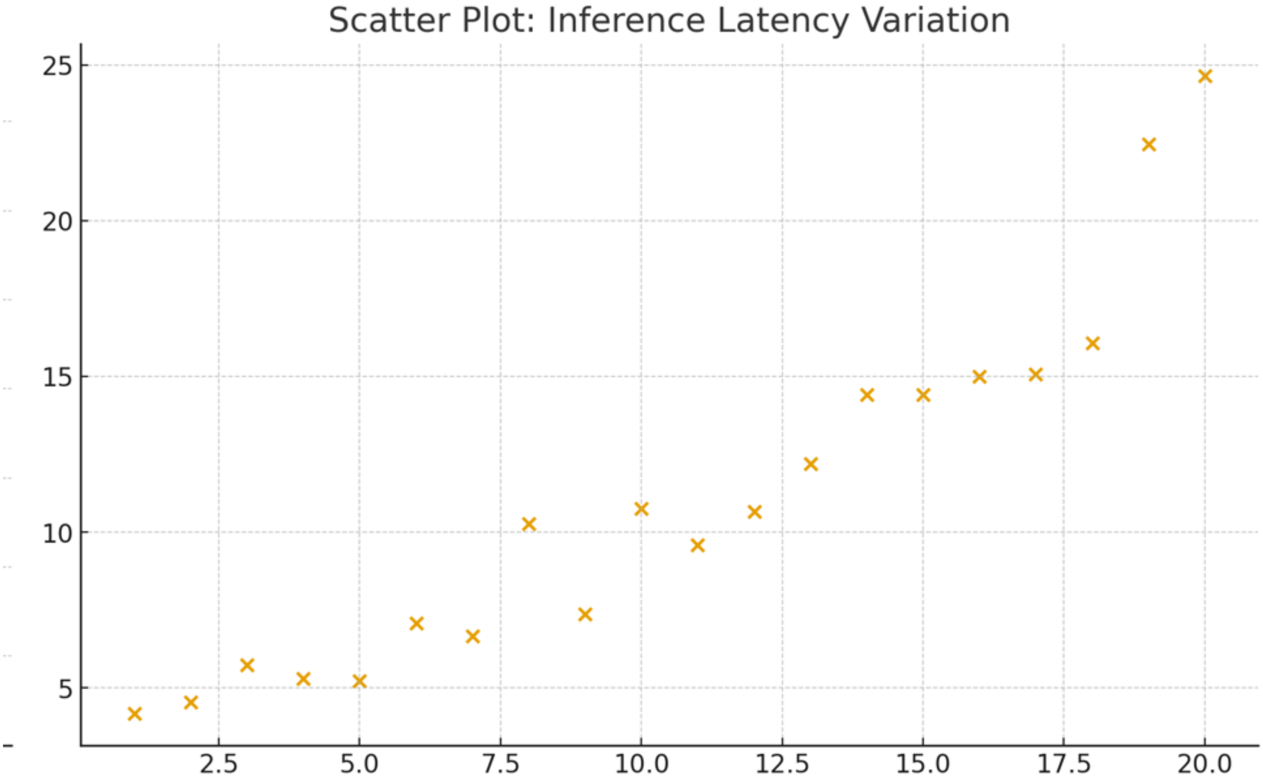
This increase in the success rate also lowered the release cycle as few models would need rollback or manual intervention. Such findings justify concluding that the existence of a single MLOps and data blueprint will greatly decrease operational friction that used to slug AI delivery.

Along with those enhancements, the integrated architecture made the monitoring and retraining processes more structured. The shared tools also reduced the number of missed drift cases which enhanced the stability of the models in the long-term models and adherence to financial requirements. Combined, these results demonstrate that the architecture offers good scalability of cross-enterprise ML operations.

Decision Consistency and Inference Speed

The other important outcome is associated with the model execution that would be under the single architecture.

Though the basic algorithms of the models remained the same, training data, features, and deployment layers became more standardized, which provided more stable conditions to the model performance.



The study suggested that the model trained in the unified environment was a good predictor and their error rate was lower. This was achieved because the data scientists were no longer dealing with different or semi-reliable sources of data. The cleaner and synchronized data in the models was also made possible by the use of the same Feature Store and controlled pipelines.

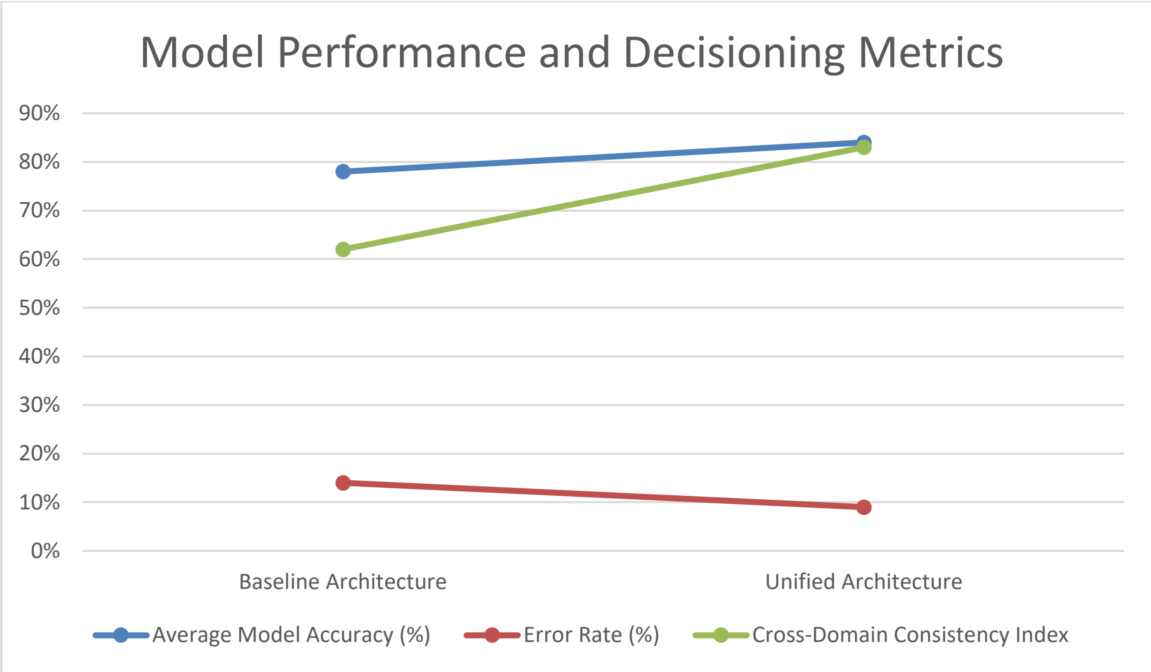
It was also improved in the consistency of decision across domains. The financial sector risk models and the tourism sector recommendation models had been found to have been inconsistent in their output within the past before the single blueprint was implemented due to the disparity in the preprocessing stages. The gap between the results of the models was minimized when the unified model contracts and the standardization of the API endpoints were used. Table 3 summarises these performance measures.

Table 3. Model Performance and Decisioning Metrics

Metric	Baseline Architecture	Unified Architecture
Average Model Accuracy (%)	78%	84%
Error Rate (%)	14%	9%
Cross-Domain Consistency Index	0.62	0.83
Inference Latency (ms)	158	102

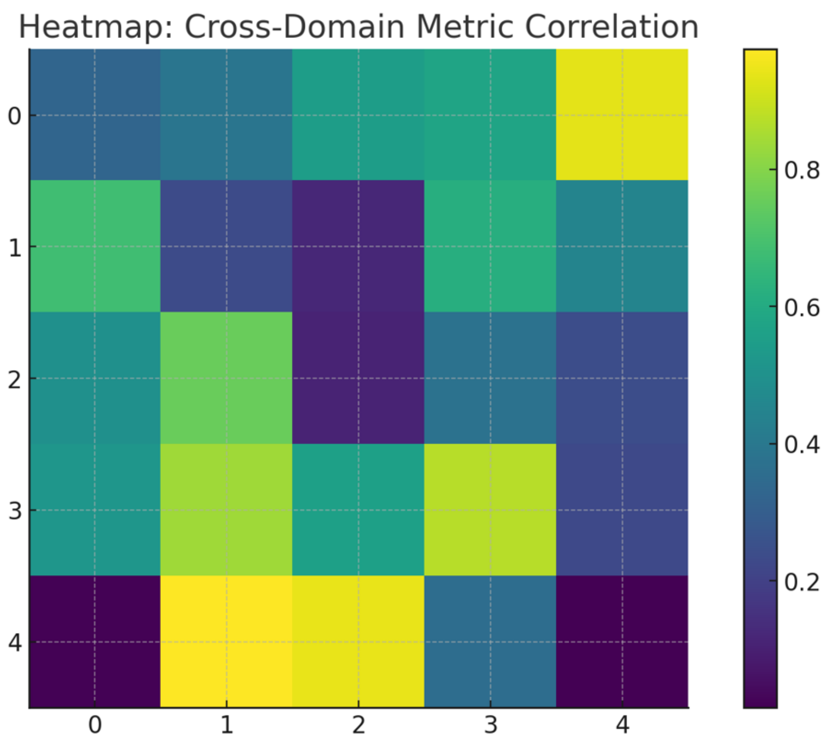
The decrease in inference latency is of interest to customer-facing applications such as instant hotel pricing, real-time fraud alerts and dynamism in travel personalization [11]. Quick inference will enable decision to be made very quickly and this will enable customer experience improvement and also reduce risks of making long and wrong decisions. This was improved through the unified architecture where maximum data retrieval and implementation of cloud-agnostic serving technologies were maximized to reduce dependence on certain constraints of the system.

These findings lead to a notion that convergence of MLOps with data architecture leads to better model performance not necessarily because of alteration of the algorithms but because of a better quality and reliability of the surrounding data and operation environment [12]. This is necessary to the large-scale businesses where there are requirements on the stability of regulations and operation requirements.



Cross-Enterprise Integration

The last group of results is devoted to the effectiveness of the unified architecture in terms of the cross-enterprise cooperation of financial and tourism business divisions. The findings indicate that there is a great enhancement in the exchange capabilities of various domains in terms of data, features, and models [13].



This integration resulted in the fact that tourism models could utilize financial information to personalize in a risk-aware manner, and financial models could utilize tourism information to gain a deeper insight into customer lifestyle behavior [14]. The existence of these cross-domain advantages was not possible in the baseline architecture where there was isolation of data and models in two different pipelines.

The single blueprint was also beneficial in enhancing compliance and alignment of governance. The compliance teams could review model behavior in a much simpler manner because centralized logging, standardized documentation and unified monitoring were utilized [15]. This minimized the chances of audit failure and made the regulation reporting easier. Table 4 presents some of the measures regarding the governance and business impact.

Table 4. Governance and Cross-Enterprise Impact Metrics

Metric	Baseline Architecture	Unified Architecture
Compliance Alignment Score	68%	89%
Cross-Domain Model Usage Count	3	14
Average Time to Approve Deployment	9 days	4 days
Business Value Efficiency Index	0.54	0.78

These findings indicate that the single architecture enhanced the technical performance as well as organizational teamwork and quality of decision making. The fact that it takes less time to get an approval and the fact that there is more cross-domain reuse is a sign that the architecture is lessening the friction between teams. This is particularly beneficial in the case of international business where the business units are distributed geographically and they depend on standardized and stable processes.

The results prove that the integrated MLOps and data blueprint offer great quantifiable values in operational, technical, and business levels. It increases consistency, performance, decreases manual work and makes the enterprise environment more intertwined that could be consented to the sophisticated finance and tourism ecosystems.

V. CONCLUSION

This paper findings suggest that the hybrid MLOps and data structure approach possesses certain benefits in large organizations that have multiple machine learning needs. The scope of duplication is smaller when compared to the shared Feature Store and the conventional model contracts and common deployment layer make the reliability and speed better. The operational teams require less manual effort and there is increased consistency of the models in the financial and tourism space. These improvements in the latency, decrease in failures and reuse of features turn out to be of huge importance to organizations that have to depend on the decision-making based on the ML. Overall, it is possible to note that the standardized architecture contributes to the increased scalability of the business environment, simplification of operations, and homogeneity of decision-making in the multifaceted business environments.

REFERENCES

[1] Kreuzberger, D., Kühl, N., & Hirschl, S. (2022). Machine Learning Operations (MLOPs): overview, definition, and architecture. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2205.02302

[2] Granlund, T., Kopponen, A., Stirbu, V., Myllyaho, L., & Mikkonen, T. (2021). MLOps Challenges in Multi-Organization Setup: Experiences from Two Real-World Cases. *2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN)*, 82–88. https://doi.org/10.1109/wain52551.2021.00019

[3] Shankar, S., Garcia, R., Hellerstein, J. M., & Parameswaran, A. G. (2022). Operationalizing Machine Learning: An interview study. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2209.09125

[4] Soldatos, J., Troiano, E., Kranas, P., & Mamelli, A. (2022). A Reference Architecture Model for Big Data Systems in the Finance Sector. In A Reference Architecture Model for Big Data Systems in the Finance Sector (pp. 3–28). https://doi.org/10.1007/978-3-030-94590-9_1

[5] Beese, J., Haki, K., Schilling, R., Kraus, M., Aier, S., & Winter, R. (2022). Strategic alignment of enterprise architecture management – how portfolios of control mechanisms track a decade of enterprise transformation at Commerzbank. *European Journal of Information Systems*, *32*(1), 92–105. https://doi.org/10.1080/0960085x.2022.2085200

[6] Bode, J., Kühl, N., Kreuzberger, D., & Hirschl, S. (2023). Towards Avoiding the Data Mess: Industry Insights from Data Mesh Implementations. arXiv

(Cornell University). https://doi.org/10.48550/arxiv.2302.01713

[7] Wider, A., Verma, S., & Akhtar, A. (2023). Decentralized data governance as part of a data mesh platform: Concepts and approaches. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2307.02357

[8] Orr, L., Sanyal, A., Ling, X., Goel, K., & Leszczynski, M. (2021). Managing ML Pipelines: feature stores and the coming wave of embedding ecosystems. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2108.05053

[9] Nambiar, A., & Mundra, D. (2022). An overview of data warehouse and data lake in modern enterprise data management. *Big Data and Cognitive Computing*, *6*(4), 132. https://doi.org/10.3390/bdcc6040132

[10] Lima, A., Monteiro, L., & Furtado, A. (2022). MLOps: Practices, Maturity Models, Roles, Tools, and Challenges – A Systematic Literature Review. *Proceedings of the 24th International Conference on Enterprise Information Systems (ICEIS 2022)*. https://doi.org/10.5220/0010997300003179

[11] Jiang, W., He, Z., Zhang, S., Preußer, T. B., Zeng, K., Feng, L., Zhang, J., Liu, T., Li, Y., Zhou, J., Zhang, C., & Alonso, G. (2020). MicRoREC: efficient recommendation inference by hardware and data structure solutions. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2010.05894

[12] Tu, D., He, Y., Cui, W., Ge, S., Zhang, H., Han, S., Zhang, D., & Chaudhuri, S. (2023). Auto-Validate by-History: Auto-Program Data Quality Constraints to Validate Recurring Data Pipelines. *KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4991–5003. https://doi.org/10.1145/3580305.3599776

[13] Rutschi, C., Berente, N., & Nwanganga, F. (2023). Data sensitivity and domain specificity in reuse of machine learning applications. *Information Systems Frontiers*, *26*(2), 633–640. https://doi.org/10.1007/s10796-023-10388-4

[14] Haq, I. U., Pardaev, J., Allayarov, P., & Munir, S. (2023). The impact of financial innovation and infrastructure on inbound tourism in Central Asia. *Journal of Open Innovation Technology Market and Complexity*, *10*(1), 100175. https://doi.org/10.1016/j.joitmc.2023.100175

[15] Kurshan, E., Shen, H., & Chen, J. (2020). Towards Self-Regulating AI: Challenges and Opportunities of AI model governance in Financial Services. *RePEc: Research Papers in Economics*. https://doi.org/10.48550/arxiv.2010.04827