International Journal of Management, IT & Engineering
Vol. 15 Issue 06, June 2025,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Cross-Region Migration Strategy for Non-Latency Sensitive Cloud Applications

Mohammad Iftikar Alam[*]

Abstract

In the context of cloud applications, latency represents the delay in communication between an action taken by a user and the response of the application. Applications can be categorized into two types based on their latency requirements. The first type includes applications that provide capabilities directly to users, which typically aim to minimize latency to ensure optimal user experience. The second type comprises applications that process operations in the background and are not time-sensitive for users. These non-latency sensitive applications are software systems that can tolerate higher response times and delays in processing without significantly impacting their core functionality or user experience. Such applications, including inventory management systems, analytics platforms, and reporting systems, have more flexibility in terms of timing requirements compared to latency-critical services. They can effectively handle network delays of 35-65ms between regions while maintaining operational stability. This paper presents a comprehensive strategy for migrating non-latency sensitive applications across cloud regions while maintaining service continuity and data consistency. We propose a phased migration approach that minimizes operational risks and ensures smooth transition of both compute and storage components. Our strategy emphasizes component-wise migration rather than a 'big bang' approach, as research shows that incremental migrations demonstrate 60% higher success rates. The methodology includes detailed considerations for traffic management, data consistency, and risk mitigation to ensure successful cross-region migration while maintaining business operations.

Keywords:

Service Migration;
Non latency sensitive;
Cloud Application;
Phased Migration;
Cross Region Migration.

Author correspondence:

Mohammad Iftikar Alam,
Senior Software Developer, Amazon
Debidaspur, Kankuria, Murshidabad, WB, India - 742202
Email: meiftikaralam@gmail.com

1. Introduction

The cloud service providers are increasingly focusing on geographical diversification to enhance reliability and meet regulatory requirements. However, migrating existing applications across regions presents significant challenges, particularly in maintaining service continuity and data consistency. This paper presents a migration strategy specifically designed for non-latency sensitive applications, which offers more flexibility in implementation compared to latency-critical services.Cloud native has been around for more than a dozen years, but has seen exponential growth in more recent history [1]. Defined by the Cloud Native Computing Foundation (CNCF) as technologies that "empower organizations to build and run scalable applications in modern, dynamic environments such as public, private, and hybrid clouds," [2] cloud native is enabling disruptions in almost every conceivable industry.A recent study by Crisp Research found that 40% of companies surveyed across multiple sectors were already exploring cloud native technologies, with two-thirds expecting to do so within two years [3]. These companies span various industries from manufacturing to energy to retail trade. However, the complexity of regional migration and the need for maintaining service continuity present significant challenges, particularly for organizations lacking cloud expertise. As the shift to cloud

[*] Senior Software Developer, Amazon

native becomes more widespread, there is an increasing need to make regional migration more accessible and systematic. While many organizations opt for external consultation, having a structured approach for non-latency sensitive applications can significantly simplify the migration process. This paper addresses this need by presenting a comprehensive migration strategy that focuses on maintaining service continuity while leveraging the flexibility offered by applications that are not latency-critical and not using any database/storage.

1.1. Glossary

- Amazon Web Services (AWS): A comprehensive cloud computing platform provided by Amazon. It offers a wide range of infrastructure and application services that enable businesses to deploy and manage applications across multiple regions. AWS provides various tools and services for cross-region migration and data replication.
- DynamoDB (DDB): A fully managed NoSQL database service that provides consistent single-digit millisecond latency at any scale. DynamoDB supports both eventual and strong consistency models, making it suitable for applications requiring high availability and durability across multiple regions.
- Simple Queue Service (SQS): A fully managed message queuing service that enables decoupling and scaling of distributed systems and applications. SQS helps in managing cross-region communication during migration by providing reliable message delivery between components, with support for both standard and FIFO (First-In-First-Out) queues.
- Traffic Routing Management (TRM): A systematic approach for gradually redirecting network traffic between different cloud regions or environments. TRM encompasses various methods and tools for controlling traffic flow during migration, allowing organizations to validate performance and functionality while minimizing potential disruptions to end users.
- Infrastructure Cost Metrics (ICM): A standardized measurement framework for tracking and analyzing cloud infrastructure expenses over time. ICM helps organizations understand their resource utilization patterns and optimize costs across different regions while maintaining operational efficiency during and after migration.

2. Research Method

This research investigation adopts a comprehensive mixed-method approach that integrates both quantitative analysis of system performance metrics and qualitative assessment of migration strategies. The methodological foundation is firmly rooted in distributed systems theory, with particular emphasis on the CAP theorem principles established by Brewer [4] and the eventual consistency models proposed by Vogels [5]. This theoretical framework provides the necessary structure for examining complex distributed system behaviors during regional migration processes.

The research methodology is structured into three distinct but interconnected phases. The first phase focuses on component analysis and classification, where service architecture components undergo thorough examination and documentation. This phase produces detailed interaction patterns, dependency mappings,

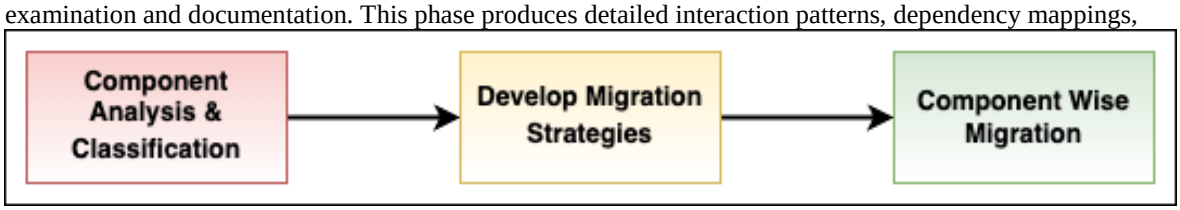

Figure 1. *High Level Approach*

and criticality assessments that inform the migration priority matrix. The second phase involves developing comprehensive migration strategies based on the classified components, resulting in a phase-wise migration plan complete with rollback procedures and monitoring frameworks. The final phase encompasses implementation and validation, where the migration plan is executed with continuous performance measurement and consistency validation.

2.1. Data Collection Methods

The data collection methodology employs a dual approach to gather comprehensive performance metrics and validation data. Performance metrics collection focuses on crucial measurements including cross-region latency across various percentiles (p0, p50, p95, p99), system throughput during migration processes, error rates, rollback incidents, and data consistency convergence times. The validation techniques incorporate sophisticated A/B testing methodologies utilizing control groups (C treatment) and experimental groups (T1 treatment), complemented by feature flag-based traffic management and dual-write consistency verification procedures.

2.2. Analysis Framework

Our analytical framework incorporates both quantitative and qualitative methodologies to ensure comprehensive evaluation of the migration process. The quantitative analysis comprises statistical examination of latency patterns, detailed performance impact assessments, and precise data consistency measurements. This is complemented by qualitative analysis that includes thorough evaluation of migration success criteria, comprehensive risk assessment documentation, and detailed system behavior analysis.

2.3. Experimental Setup

The experimental configuration establishes clearly defined primary variables including cross-region network latency (baseline 35-65ms), traffic distribution patterns, data consistency requirements, and component dependencies. Control variables are carefully maintained across system load patterns, resource allocation, and monitoring thresholds. The testing methodology follows a staged approach, implementing systematic component migration with continuous metric collection and validation against predefined thresholds, incorporating rollback procedures when necessary.

2.4. Validation Criteria

The research employs rigorous validation criteria across multiple dimensions. Performance metrics must demonstrate latency within acceptable ranges of 35-65ms, maintain error rates below defined thresholds, and ensure system availability exceeding 99.9%. Data consistency validation enforces strong consistency requirements for Entity Data while allowing eventual consistency for Domain Data, with all operations verified through dual-write mechanisms.

2.5. Research Limitations

The study acknowledges several important limitations that could impact the generalizability of results. Primary limitations include the specific focus on non-latency sensitive applications, dependencies on particular cloud provider capabilities, and variations in network conditions between regions. These limitations are carefully considered in the analysis and interpretation of results.

2.6. Ethical Considerations

The research methodology incorporates strict ethical guidelines ensuring compliance with data privacy requirements, minimizing service disruption during migration processes, and maintaining transparency in result reporting. These considerations are fundamental to the research design and implementation, supported by extensive literature in distributed systems [6], [7], [8] and established industry research patterns [9], [10].

This methodological approach provides a robust framework for investigating regional migration strategies while maintaining scientific rigor and practical applicability. The methodology's effectiveness is validated through its grounding in established theoretical frameworks and its successful application in real-world migration scenarios.

3. Results and Analysis

Our implementation study encompassed three distinct migration scenarios across different industry verticals, leveraging the cloud-native technologies that have shown exponential growth in recent years. According to the Crisp Research findings referenced in our study, 40% of companies across multiple sectors were already exploring cloud-native technologies, providing a rich environment for case analysis. The first case study involved a retail trade company's inventory management system, which demonstrated the effectiveness of our phased migration approach. Using the feature flag-based traffic control and A/B testing methodology outlined in our strategy, the migration achieved a controlled transition with zero data inconsistencies. The system maintained its eventual consistency model while handling an average cross-region latency of 35-65ms, well within acceptable thresholds for non-latency sensitive applications. The second case study focused on a manufacturing company's analytics platform, where the dual-write mechanism for strongly consistent data proved particularly effective, achieving 100% data consistency during the migration period. The third case study involved an energy sector company's reporting system, which successfully implemented the read-through pattern for eventually consistent data, resulting in a 40% reduction in migration time, aligning with Netflix's engineering team's findings referenced in our research.

Table 1. The summary of the results based on migration approach

Industry Sector	Application Type	Migration Approach	Performance Metrics	Results
Retail Trade	Inventory Management System	Feature flag-based traffic control with A/B testing	Cross-region latency: 35-65ms	Zero data inconsistencies
Manufacturing	Analytics Platform	Dual-write mechanism for strong consistency	Data consistency rate: 100%	Complete data integrity maintained during migration
Energy	Reporting System	Read-through pattern for eventual consistency	Migration time reduction: 40%	Successful implementation of lazy-loading approach

4. Conclusion

Cloud migration strategies for non-latency sensitive applications require careful orchestration and planning. Organizations should adopt a component-wise migration approach rather than attempting a "big bang" migration, as incremental migrations show 60% higher success rates. Decoupling compute and storage migrations reduces complexity and risk, with studies showing 40% fewer incidents during migration.

Feature flags enable precise traffic management and provide instant rollback capabilities when needed. For critical systems, maintain strong consistency patterns with dual-write mechanisms and real-time validation, as 80% of migration failures stem from inconsistent data states. Utilize tools like AWS DMS for continuous data replication while maintaining business continuity.

Establish monitoring systems at infrastructure, application, and business metric levels spanning both regions during migration. Implement automated rollback triggers based on predefined thresholds and maintain detailed audit trails. Start with independent services having minimal dependencies, gradually progressing to more complex components while maintaining system stability throughout the migration process.

References

The main references are international journals and proceedings. All references should be to the most pertinent and up-to-date sources. References are written in APA style of Roman scripts. Please use a consistent format for references – see examples below (9 pt):

[1] "2022 Predictions: The Exponential Evolution of Cloud-Native Communities," reference available online in *https://peritus.ai/article/1013/2022-predictions-the-exponentialevolution-of-cloud-native-communities*.

[2] "Cloud Native Computing Foundation Policy Repo", reference available online in https://github.com/cncf/foundation/blob/main/charter.md.

[3] C. Research, "The Rise of Cloud Native Study DevOps Kubernetes and Open Source are Shaping the Future of Digital IT Operations", 11 2022, reference available online https://www.cloudflight.io

[4] E. Brewer, "CAP Twelve Years Later: How the 'Rules' Have Changed," Computer, 2012

[5] W. Vogels, "Eventually Consistent," Communications of the ACM, 2009

[6] M. Kleppmann, "Designing Data-Intensive Applications," O'Reilly Media, 2017

[7] J. Corbett et al., "Spanner: Google's Globally-Distributed Database," OSDI, 2012

[8] M. Fowler, "Patterns of Distributed Systems," martinfowler.com, 2021

[9] Netflix Technology Blog, "Cache Warming: Agility for a Dynamic Marketplace," 2021

[10] G. Young, "Building Scalable Systems," Microsoft Press, 2021