Journal of Technological Innovations

Est. 2020

Enabling Effective Multi-Cloud Data Management with DIVA: Architecture and Capabilities

Vamshi Krishna Dasarraju

Abstract

Multi-cloud strategies have gained traction among organizations seeking to mitigate vendor lock-in, optimize costs, enhance disaster recovery, and leverage diverse cloud services. However, managing data across multiple cloud providers presents formidable security, governance, performance, and scalability challenges. Legacy single-cloud approaches must be revised to address the complexities of multi-cloud environments. This paper introduces DIVA (Data Intelligence and Visibility Architecture), a novel framework designed to address these challenges through advanced architectures, governance models, and tailored strategies.

Introduction

In contemporary IT landscapes, multi-cloud data management refers to the orchestration, security, optimization, and governance of data distributed across various public cloud platforms. This includes database services (e.g., Azure SQL DB, Amazon RDS) and cloud-native storage solutions (e.g., Azure Blob Storage, Amazon S3). Effective multi-cloud data management necessitates capabilities such as comprehensive data discovery, lineage tracking, secure migration, access control, governance enforcement, performance monitoring, and cost optimization.

Challenges and Legacy Approaches

Traditional approaches designed for on-premises or single-cloud environments need help to meet the demands of multi-cloud setups. They often exhibit shortcomings in data discovery, manual processes for data movement, inadequate performance monitoring tools, primitive migration capabilities, disjointed security policies, and dependency on cloud vendorspecific services.

The DIVA Architecture: Overview

DIVA introduces a holistic approach to multi-cloud data management to address legacy model deficiencies. DIVA is structured around seven core capabilities, each aimed at enhancing visibility, governance, and operational efficiency across AWS, Azure, and GCP cloud environments:

1. **Shared Metadata Store:** Central catalog incorporating technical, business, operational, and usage metadata for unified data management.
2. **Data Map Visualization:** Interactive visualization of data flows, lineage, and relationships across multi-cloud platforms.

3. **Usage Dashboards:** Analytics and reporting for optimizing costs, monitoring performance, and identifying trends across cloud providers.

4. **Policy Engine:** Centralized enforcement of security, access controls, regulatory compliance, and data lifecycle automation.

5. **Migration Factory:** Automated, policydriven data migration and synchronization between cloud platforms.

6. **Management Portal:** Unified interface for orchestrating data management workflows and optimizing resource utilization.

7. **Standard Services:** Standardized functionalities for data lifecycle management, disaster recovery, data science collaboration, and search capabilities.

Key Components of DIVA

Shared Metadata Store: A comprehensive catalog aggregating technical, business, operational, and usage metadata. Enables rapid data discovery, lineage tracing, and compliance auditing across multi-cloud environments.

Data Map Visualization: This tool provides intuitive maps of data flows and relationships, aiding in migration planning, consolidation strategies, and anomaly detection across cloud systems.

Usage Dashboards: Offers real-time analytics on cost breakdowns, performance metrics, and resource utilization across AWS CloudWatch, Azure Monitor, and GCP/GCO Monitoring.

Policy Engine: Centralizes governance policies, ensuring consistent enforcement of access controls, data security measures, and regulatory compliance across heterogeneous cloud infrastructures.

Migration Factory: Facilitates seamless data movement and synchronization through automated workflows, supporting common migration patterns (e.g., relational database replication, blob storage mirroring) across AWS, Azure, and GCP.

Management Portal: Serves as a unified control plane for managing multi-cloud data operations, providing administrators with insights into resource utilization, performance optimizations, and policy enforcement.

Common Services offers standardized functionalities for managing the data lifecycle, disaster recovery, data science collaboration, and search capabilities across AWS, Azure, and GCP.

Technical Architecture of DIVA

DIVA adopts a hub-and-spoke model centered around a MongoDB-based Shared Metadata Store. Microservices built on Spring and Quarkus frameworks provide modular functionalities such as policy evaluation, migration orchestration, optimization, and analytics. Integration agents facilitate seamless interaction with AWS, Azure, and GCP services, ensuring real-time synchronization of metadata and operational commands.

Governance and Performance Strategies

DIVA emphasizes autonomous governance through continuous policy evaluation, event-triggered compliance checks, and automated remediation actions. It supports low-code policy modeling for defining and enforcing governance rules across multicloud assets, promoting early governance integration ("shift left") in development lifecycles.

Performance Enhancement and Future Directions

DIVA leverages event-driven architectures and serverless computing paradigms to enhance data processing speeds, scalability, and agility across cloud platforms. It advocates embracing cloud-native innovations and API-first principles to future-proof data management strategies against evolving cloud landscapes.

Conclusion

The DIVA architecture represents a significant advancement in multi-cloud data management, offering a comprehensive solution to the challenges of heterogeneous cloud environments. By integrating advanced governance models, real-time analytics, and scalable infrastructure, DIVA empowers organizations to achieve high-performance data operations across AWS, Azure, and GCP, thereby unlocking the full potential of multi-cloud strategies.

References

[1] B. Ghosh, "Novel Approach for Multi-Cloud Data Management," Medium, 25 December 2023. [Online]. Available: https://medium.com/@bijit211987/novel-approach-for-multi-cloud-data-management-e2e298c12f55.

[2] A. Smith, "6 Multi-Cloud Architecture Designs for an Effective Cloud Strategy," Medium, 1 March 2024. [Online]. Available: https://medium.com/stackademic/6-multi-cloud-architecture-designs-for-an-effective-cloud-strategy-9e5c256878a0.