NetApp — Data in Motion: Secure Multi-Cloud Storage and Replication Pipeline

Birla Institute of Technology and Sciences, Pilani

Team Members:

- Pranak Shah (2022B2A71344P)

- Aditya Raj (2022B2A41583P)

- Kush Natani (2022B4AA1288P)

- Dhruv Gupta (2022A3PS1206P)

1. Introduction

As organizations continue to scale operational workloads and analytics across distributed infrastructure, data increasingly needs to move across multiple cloud environments. While major cloud platforms offer reliable object storage, they operate independently and do not natively provide mechanisms for secure, consistent, and verifiable cross-cloud data synchronization. This leads to challenges involving vendor dependency, limited data portability, security compliance complexity, and operational risk. NetApp — Data in Motion addresses these gaps by implementing a unified data replication pipeline that moves encrypted data across cloud storage providers without exposing sensitive content to any provider. The system enforces confidentiality at the application layer and uses cryptographic verification to ensure that replicas across cloud environments are always consistent. This solution is fully containerized and deployable in local or enterprise environments, making it both practical and extensible.

2. Problem Statement

Enterprises working in regulated or high-reliability environments must ensure that their data remains available, portable, and secure across infrastructure providers. Storing data in only a single cloud exposes organizations to vendor lock-in, outage vulnerability, and compliance violations. Meanwhile, traditional multi-cloud management strategies rely on manual replication scripts or third-party synchronization tools that do not guarantee encryption consistency or replication integrity. Key difficulties include: • Different security and IAM models across providers • Lack of standardized encryption enforcement • Risk of confidentiality loss if cloud vendors access plaintext data • No native cross-provider checksum verification • Operational burden of monitoring replication status manually A viable multi-cloud solution must unify these concerns under a single control plane. Our system does so by shifting all security and replication logic to the API layer, before data touches any cloud environment.

3. System Architecture Overview

The system is composed of four key subsystems working together to provide secure and deterministic replication: 1. Unified Control Plane (FastAPI): Handles user interaction, encryption, hashing, and replication logic. 2. Multi-Cloud Storage Layer: Data is stored across MinIO (S3-compatible), Azurite (Azure Blob), and FakeGCS/GCS (Google Cloud Storage). 3. Cryptographic Security Layer: All data is encrypted using AES-256 before transit and storage, ensuring clouds never see plaintext. 4. Observability and Dashboard Layer: A Streamlit UI visualizes replication status, data integrity, and provider availability in real time.

Architecture Diagram: +--------------------------+ | Streamlit UI | | Live Replication Status | +-----------+---------------+ | v +---------+---------+ | FastAPI | | Control Plane | +---------+---------+ | --- | | | v v v MinIO (S3) Azurite (Blob) FakeGCS / GCS

4. Data Flow Workflow

1. A client uploads a file to the FastAPI control plane. 2. The file is encrypted using AES-256 and assigned a SHA-256 cryptographic hash. 3. The encrypted object is replicated to MinIO, Azurite, and FakeGCS/GCS. 4.

Each provider acknowledges successful upload. 5. The dashboard displays data integrity status using hash comparison across clouds. This ensures confidentiality, integrity, and traceability throughout the data lifecycle.

5. Security and Encryption Design

The security model follows a zero-trust approach: cloud providers are treated as untrusted storage engines that hold encrypted blobs without access to decryption keys. The encryption is performed at the application layer using AES-256 symmetric encryption. Integrity is verified using SHA-256, ensuring all replicas match bit-for-bit. Because keys never leave the control plane, cloud compromise does not expose data contents.

6. Replication Reliability & Failure Recovery

The replication process is resilient to temporary outages. If a cloud provider becomes unavailable, the system retries replication periodically until the provider is reachable. Additionally, verification endpoints allow real-time consistency checking across providers, ensuring early detection of drift or replication failure. Idempotent operations prevent duplication or overwrite issues.

7. Dashboard and Observability

The Streamlit dashboard provides real-time insight into the multi-cloud system. It displays available storage endpoints, replication status of each stored object, and SHA-256 consistency verification. This improves operational visibility and simplifies compliance auditing.

8. Conclusion

NetApp — Data in Motion provides a secure, extensible, and verifiable approach to multi-cloud data mobility. By anchoring trust in cryptographic enforcement rather than cloud provider policies, the system ensures confidentiality and data integrity across environments. The architecture is modular and ready for extension to real cloud vendors, Kubernetes orchestration, predictive access optimization, and adaptive replication policies.