FUTUREWEI

Whitepaper

LUMOS: Lake Unified Multi-cloud and On-premises Solution for the Financial Services Industry

Trademarks and Permissions

Notice

The purchased products, services, and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services, and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees, or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

FUTUREWEI® TECHNOLOGIES, INC.

Boston Research Center

Address: 111 Speen Street, Suite 114
Framingham, MA 01701
United States of America

Website: http://www.futurewei.com/

Executive Summary

As the financial industry accelerates its digital transformation—driven by demands for real-time services, AI-driven insights, and regulatory compliance—it faces critical limitations with current public cloud and hybrid cloud strategies. Key pain points among these are **single points of failure**, **insufficient service-level guarantees**, and **inability to maintain control over mission-critical data and applications**.

This whitepaper introduces the LUMOS (Lake Unified Multi-cloud and On-premises Solution) architecture: an **on-premises-anchored** unified model that places an on-premises anchored cross cloud data lake at the core of the FSI's digital infrastructure. Unlike cloud-first or lift-and-shift hybrid approaches, this strategy anchors data, governance, and resiliency on-prem, while selectively extending data, compute and services to public clouds for elasticity, AI, and innovation.

An **on-premises-anchored multi-cloud architecture** is the only approach that can meet the high standards of operational resiliency, auditability, and restart capability demanded by financial institutions for their most critical workloads. Relying solely on the public cloud cannot ensure uninterrupted service in the face of provider outages, geopolitical tensions, or evolving compliance requirements.

Key differentiators of the LUMOS architecture include:

- Centralized Data Governance
  A single on-premises anchored cross cloud data lake ensures data sovereignty, compliance, and acts as the definitive system of record.
- Adaptive SLOs and Mobility
  Tiered reliability ensures graceful service degradation from core to edge, with rapid application and data mobility for quick recovery and compliance.
- Built-in Disaster Recovery and Observability
  Integrated disaster recovery is standard, supported by unified identity management and seamless observability across all environments.
- AI and Analytics Ready
  Fully supports AI, machine learning, advanced analytics, and containerized workloads without sacrificing control or security.

This approach enables FSI service providers to minimize vendor lock-in, comply with evolving regulations such as DORA and FINRA, and maintain 24/7 business continuity—even during disruptions. LUMOS provides a future-proof foundation for secure innovation, dependable operations, and confident compliance, backed by the resiliency that only an on-premises-anchored architecture can ensure.

Introduction

When the Public Clouds aren't Enough: Challenges Facing Financial Institutions

Single Points of Failure and Insufficient SLAs

Despite being positioned as highly available and resilient, public cloud SLAs (e.g., 99.99%) are often not high enough to meet the stringent, tiered Service Level Objectives (SLOs) required by modern banking systems (e.g., 99.9999%). In addition, cloud SLAs typically lack the granularity needed to support distinct workloads, from real-time transaction processing to mission-critical batch analytics. Public cloud platforms remain vulnerable as a single point of failure (SPOF). Outages in critical managed services, cloud regions, or identity providers can disrupt essential banking operations, exposing a fundamental weakness in relying solely on a single provider.

For mission critical platforms (e.g., core banking), which demand the highest levels of uptime, performance, and consistency, public cloud alone often cannot meet these expectations without incurring excessive cost or engineering overhead. The lack of support for tiered, application-specific SLOs forces banks to overbuild infrastructure across all workloads, resulting in inefficiencies, increased complexity, and higher operational costs.

To meet the SLOs essential for financial-grade reliability, banks require an architecture that goes beyond the limitations of current public cloud capabilities.

The Cost of Downtime: Why Cloud Outage Compensation Falls Short

The situation becomes even more problematic during downtime. While public cloud providers promote their SLAs, these agreements usually amount to service credits—offering little real compensation for business disruption or revenue loss. In practice, they serve more as future discounts than as genuine protections against the consequences of service outages.

For mission-critical banking systems, that's simply not sufficient. Outages affecting real-time transactions or core operations can lead to reputational harm, regulatory consequences, and significant financial loss—**none of which are meaningfully addressed by standard cloud SLAs**.

This disconnection between SLA coverage and actual business risk highlights the need for banks to reassess their cloud strategies and adopt architectures that offer enforceable service guarantees aligned with the high FSI operational standards.

The Mobility Myth: Why Multi-Cloud Application Mobility Is Harder Than It Should Be

Given the limitations of public cloud SLAs, FSI service providers increasingly adopt multi-cloud or hybrid cloud architectures to improve resilience and control. However, many current multi-cloud

strategies rely on simply lifting and shifting monolithic applications into virtualized or containerized environments. While this may offer superficial modernization, it does little to enable genuine application mobility across cloud platforms.

Applications frequently remain tightly coupled to specific infrastructure, services, or APIs tied to a single cloud vendor. This dependency limits the ability to move workloads dynamically in response to cost optimization, performance improvements, or evolving regulatory requirements.

These dependencies undermine the core benefits of multi-cloud, limiting the ability to dynamically move workloads based on cost efficiency, performance needs, or shifting regulatory demands. As a result, organizations face reduced resilience and constrained operational flexibility. Without seamless application mobility, it's difficult to meet service-level objectives (SLOs) and organizations remain bound by the limitations of a single provider's SLA.

Lacking Data Consistency and Mobility

Data mobility remains a critical constraint in multi-cloud and hybrid environments. Transferring data between public clouds—or between cloud and on-premises systems—is often slow, costly, and operationally complex. Even more challenging is the task of maintaining consistent application and data states across distributed environments.

Without strong consistency, organizations face the risk of **data drift**, where discrepancies emerge among data copies. This can lead to application failures**,** inaccurate regulatory reporting, broken customer experiences, and misleading analytics.

The absence of a unified data architecture severely limits true multi-cloud agility. When application logic and data are tightly coupled to specific platforms or inconsistent across regions, service disruptions become inevitable—especially in dynamic environments where resiliency and compliance are paramount.

Insufficient Cloud Security and Compliance

Although many cloud providers highlight certifications like GDPR, ISO 27001, and SOC 2, these standards often fall short of meeting the more rigorous regulatory and operational requirements of the financial services industry. Cloud-provided security controls frequently lack seamless integration with enterprise-grade security frameworks and legacy audit systems. Additionally, challenges such as multi-tenancy risks, unclear shared-responsibility models, and limited visibility into cloud provider operations create significant concerns for CISOs and compliance teams striving to maintain full oversight and control.

Regulatory Pressures and Risks of Vendor Lock-In

Financial regulators across jurisdictions — including FINRA, EBA, and DORA — increasingly emphasize **operational resilience, vendor diversification, and exit strategies**. Relying heavily on a

single cloud provider poses systemic risks, especially when no clear fallback or migration path exists. In this environment, banks must demonstrate that they can shift operations in response to geopolitical shifts, cloud outages, or compliance requirements — a level of agility that current cloud architectures rarely support.

Key Requirements for a Future-Ready Multi-Cloud Strategy

Tiered SLO Support and Graceful Degradation

A future-ready architecture must support **tiered Service Level Objectives (SLOs)** aligned with the criticality of banking workloads. Core systems like payment processing or risk engines require the highest availability and recovery targets, while lower-tier services such as reporting or internal tools can tolerate lower SLOs. Crucially, the system should support **graceful degradation**—enabling partial service delivery under stress or failure scenarios, rather than full outages. This allows banks to maintain operational continuity while minimizing user impact.

On-Premises Anchored Multi-Layer App/Data Mobility

Achieving true application mobility requires a decoupled architecture, where business logic and service orchestration are abstracted from the underlying infrastructure. This involves adopting portable deployment models—such as containers, service meshes, and cloud-agnostic dependencies—that allow applications to move fluidly across environments.

Banks should be able to dynamically relocate services between public clouds and on-premises environments, guided by real-time considerations like latency, regulatory compliance, data sovereignty, or cost—without requiring significant re-engineering.

However, real agility goes beyond container portability. Data mobility must extend across multiple layers, including:

- Application-layer data
- Underlying storage systems
- Metadata and governance frameworks

To support seamless operations, the architecture should enable:

- Cross-environment cross-cloud data synchronization
- Tiered consistency models (e.g., strong, eventual, etc.)
- Policy-based data placement and orchestration

This ensures applications can run as close as possible to where the data resides, while complying with data residency and privacy regulations.

By treating the application itself as a form of data, and anchoring app/data mobility around on-premises control, financial institutions can meet the strictest SLO requirements—delivering performance, reliability, and control beyond what public clouds alone can offer.

On-Premises-Anchored Security and Compliance

An on-premises-centric model provides a **foundation of trust** and **control** for sensitive workloads. Banks can retain critical data, encryption keys, and policy enforcement locally while extending selective workloads to public cloud as needed. This hybrid control plane enables deeper integration with legacy systems, more predictable security postures, and easier audits—helping to meet strict requirements from central banks, auditors, and cybersecurity authorities.

Vendor Risk Mitigation Through Multi-Cloud

By design, a multi-cloud architecture reduces the **operational and strategic risks** associated with vendor concentration. Critical services can be deployed in active-active or active-passive configurations across cloud providers, with failover capabilities and tested exit strategies. This ensures business continuity in the face of cloud outages, geopolitical tensions, or changes in vendor pricing and policies. Multi-cloud also enables banks to negotiate better SLAs and reduce switching costs.

On-Premises Capabilities for Specialized Financial Needs

Certain banking use cases—like high-frequency trading, low-latency risk analytics, or air-gapped regulatory environments—demand the performance, control, and security of on-premises infrastructure. A future-ready approach doesn't replace these capabilities but strengthens them by integrating with cloud-native tools, GPU acceleration, and AI/ML frameworks. Rather than viewing on-prem as "legacy", it should be embraced as a high-performance, strategic component of a unified computing architecture.

Governance: Striking the Right Balance Between Control and Agility

Cloud agility is meaningless without governance. Banks must enforce consistent policy controls, identity frameworks, audit trails, and change management across all environments—on-premises and cloud. A unified governance layer allows for **federated control**: giving development teams the freedom to innovate, while maintaining compliance, risk management, and operational integrity at the enterprise level.

Architectural Foundations of LUMOS

The LUMOS (Lake Unified Multi-cloud and On-premises Solution) architecture reimagines hybrid cloud for the financial sector by placing an **on-premises anchored cross cloud data lake at the center** of the bank's digital core—surrounded by multiple public clouds and data centers used strategically for resiliency, elasticity, AI, and scale. This architecture treats data as the anchor, enforcing governance, sovereignty, and operational consistency, while using public clouds as execution environments, not data controllers.

Unlike traditional hybrid or cloud-first approaches, LUMOS prioritizes a strategy allowing FSI service providers to keep sensitive assets on-prem while dynamically extending data, compute and services to the cloud when needed. This design enables maximum resilience, security, auditability, and ability to restart.

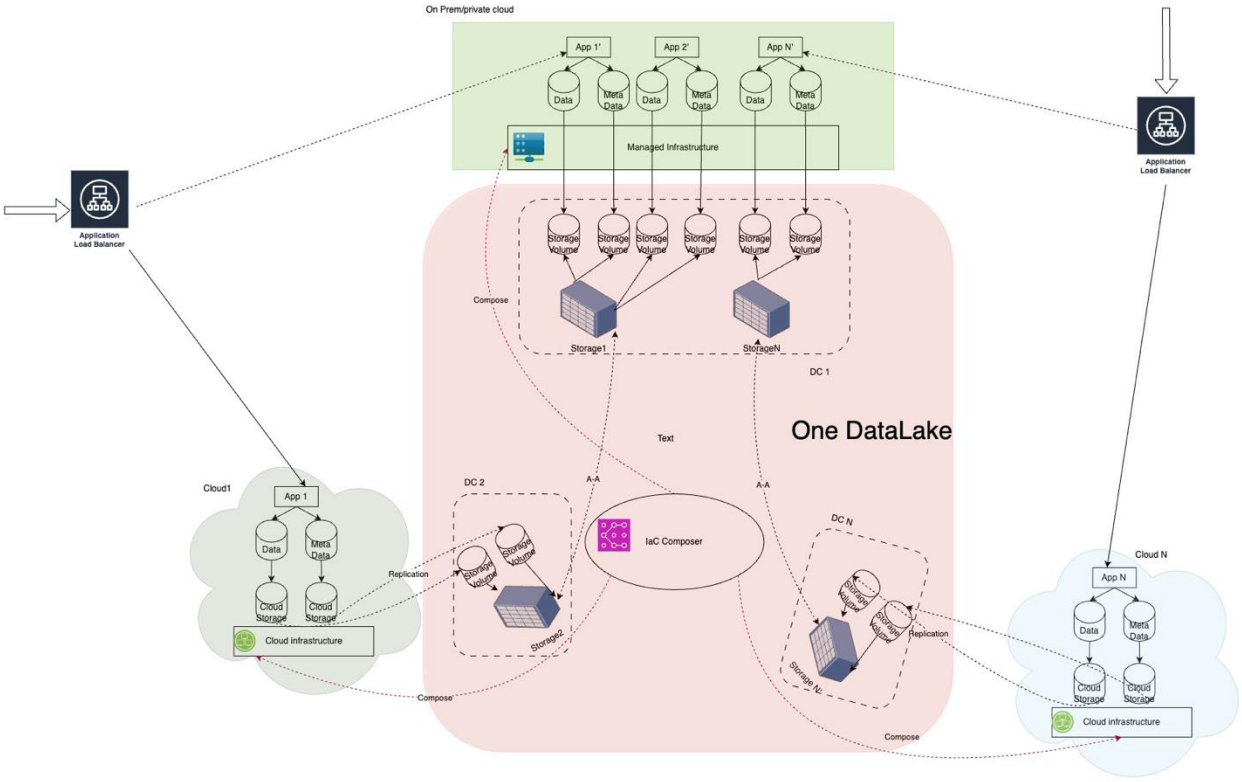

Figure 1 LUMOS overview. One data lake unifies multiple clouds with data and application mobility.

Key Characteristics:

- **Centralized, policy-driven on-prem anchored cross cloud data lake** as the system of record
- **Federated orchestration** of services across public and private clouds
- **Unified identity, access control, and observability** across environments

- **Seamless application and data mobility** across clouds and on-prem, enabling dynamic workload placement and disaster recovery

Service-Level Objective (SLO) Tiering Strategy

Not all workloads in banking workloads are created equal. A critical part of the One-Lake architecture is SLO tiering, where services are grouped based on their availability, latency, and recovery requirements:

Tier	Workload Type	SLO Profile
Tier 0	Core banking, clearing systems	99.999% uptime, sub-second failover
Tier 1	Payments, fraud detection	99.99% uptime, high availability
Tier 2	Business ops, analytics	99.9% uptime, standard DR
Tier 3	Batch jobs, test environments	Best-effort, flexible SLOs

This tiering allows for granular infrastructure planning, cost optimization, and smart service placement across on-prem and cloud. However, most of public cloud providers can't truly guarantee such high level resilience. Below are shown the gaps:

Tier	Workload Type	FSI Target SLO	AWS Equivalent (Approximate)
Tier 0	Core banking, clearing systems	99.999% uptime, sub-second failover	❌ Not natively supported. AWS SLAs top out at 99.99%. Sub-second failover requires custom HA design with multi-region and cross-zone replication.
Tier 1	Payments, fraud detection	99.99% uptime, high availability	✅ AWS services like RDS Multi-AZ, EKS, and Elastic Load Balancer target ~99.95–99.99% availability. True 99.99% often requires multi-region, custom replication.
Tier 2	Business ops, analytics	99.9% uptime, standard DR	✅ Services like S3, EC2, Redshift, Athena typically offer ~99.9% availability. DR must be manually configured across AZs or regions.
Tier 3	Batch jobs, test environments	Best-effort, flexible SLOs	✅ Covered by default compute (EC2, Lambda, Batch) and storage services without HA. Suitable for non-critical jobs.

And the remedy is not what FSI customers want(service credit is offered and it is financially much less than actual damage caused by service interruptions- potential millions level).

Remember, only on-prem can truly offer highest resiliency for financial applications.

Closing the Gaps: Building a Resilient, Unified Multi-Cloud Strategy

While interest in multi-cloud adoption continues to grow, many current implementations fall short due to persistent architectural and operational gaps. These include vendor lock-in, inconsistent observability, fragmented control planes, and data movement challenges. However, these issues are not insurmountable—with the right design principles, organizations can unlock the full potential of multi-cloud.

- Vendor lock-in: Proprietary APIs and services prevent true portability. To break free from vendor lock-in, applications must be decoupled from proprietary services and APIs. Embracing open standards, portable container orchestration platforms, and abstraction layers enables true workload mobility across cloud providers.
- Limited observability: Inconsistent monitoring and logging across environments hinders troubleshooting and compliance. To address limited observability, enterprises can implement a centralized telemetry pipeline that normalizes logs, metrics, and traces across environments. This unified visibility enhances troubleshooting, security monitoring, and compliance reporting.
- Disjointed control planes: Lack of a unified governance model across on-prem and cloud platforms. The issue of disjointed control planes can be solved by adopting a policy-driven governance layer that spans both on-prem and cloud environments. This creates consistency in access control, compliance enforcement, and operational workflows.
- Data silos and latency: Moving or synchronizing data across environments is costly and inconsistent. These challenges create operational friction, reduce resilience, and limit the flexibility that multi-cloud aims to provide. Data silos and latency can be mitigated by using distributed data fabrics or real-time data replication technologies that ensure consistency without compromising performance.

By proactively addressing these gaps with a holistic architecture, financial services organizations can move beyond the limitations of current multi-cloud models—achieving the resilience, agility, and compliance that modern workloads demand.

Designing for Resilient App and Data Mobility

In addition, LUMOS architecture embeds mobility and resilience into its foundation:

- App mobility is achieved through container-native deployments, service meshes, and platform-agnostic APIs.
- Data mobility is enabled via storage layers that abstract physical location and offer real-time replication, caching, and synchronization crossing multiple cloud although anchored in customer on-prem data center.
- Orchestration frameworks monitor service health and dynamically redirect workloads to optimal locations based on policy and context.
- Resilience is achieved not just through redundancy, but by designing for failure—graceful degradation, automated recovery, and SLO-aware fallback.

Together, these capabilities ensure banks can respond in real-time to outages, regulatory shifts, or performance degradation—without compromising service integrity.

Storage-Led Data Mobility to Enable App Portability

In the LUMOS architecture, data mobility serves as the foundation for both application and infrastructure mobility. By consolidating data into a unified lake and extending access seamlessly across on-prem and public clouds, banks can effectively decouple applications from underlying infrastructure—enabling flexible, on-demand workload deployment wherever it's needed most.

This is made possible through several key enablers:

- Treating applications and infrastructure as data, integrating them into the broader data mobility strategy.
- A unified storage layer that spans across environments, delivering consistent data semantics regardless of location.
- Global namespace and metadata services, which abstract the physical location of data and enable location-independent access.
- Policy-driven replication and caching, optimizing for performance, regulatory compliance, and data sovereignty.
- APIs and integration with container orchestration platforms (e.g., Kubernetes CSI), ensuring tight, seamless app-to-data binding.

With this architecture, banks gain the ability to dynamically relocate services in response to shifting cost structures, performance demands, regulatory changes, or resilience goals—unlocking true agility in a multi-cloud world.

Resilience Is Built In, Not Bolted On

In banking, resilience isn't optional—it's a foundational requirement. The architecture ensures continuous service availability by embedding disaster recovery (DR) directly into the runtime fabric:

- Active-active and active-passive deployments across multiple environments ensure continuous service.
- Real-time replication and automated failover orchestration minimize disruption.
- Built-in support for tiered recovery strategies aligned to SLO classifications.
- Unified monitoring and observability to trigger intelligent rerouting or load balancing.

By making DR a **core architectural capability**—not an afterthought—this architecture guarantees 24/7 availability, operational resilience, and the confidence to innovate.

Security by Design, at Every Layer

In a unified, multi-cloud environment, security must be embedded throughout the entire execution stack. Containerization provides an ideal foundation for delivering secure, isolated, and auditable services at scale.

Key capabilities include:

- Signed and immutable container images, validated through CI/CD pipelines to ensure integrity and prevent tampering.
- Robust runtime protection, leveraging sandboxing, SELinux, AppArmor, and seccomp profiles for defense-in-depth.
- End-to-end encrypted service communication, enabled by mTLS and governed through a policy-based service mesh.
- Federated identity and RBAC enforcement, ensuring consistent access control across on-prem and cloud environments.
- Unified data security controls, delivering consistent encryption, access policies, and auditability across all deployment locations.

With this approach, financial institutions can uphold the highest standards of trust, security, and regulatory compliance—while confidently scaling services in complex, hybrid cloud environments.

AI-Ready by Design: Powering Intelligent Banking at Scale

AI and machine learning are rapidly becoming foundational to modern banking—from real-time fraud detection to hyper-personalized customer experiences. The LUMOS architecture is purpose-built to support AI adoption at scale, embedding intelligence into the core infrastructure.

Key enablers include:

- The One-Lake architecture, which centralizes data governance and enables seamless data mobility across clouds—streamlining data access and unlocking the full potential of multi-cloud AI ecosystems.
- Integrated GPU clusters and native support for ML pipelines, positioned close to data sources to accelerate training and inference.
- High-throughput, low-latency storage access, ensuring efficient data flow for demanding AI workloads.
- Federated learning capabilities, allowing model training on sensitive data without compromising privacy or breaching residency requirements.
- End-to-end ML model lifecycle management, integrated into DevOps pipelines to support continuous iteration and deployment.

By embedding AI readiness directly into the infrastructure, LUMOS empowers banks to rapidly scale intelligent services—while maintaining strict control over data and compliance.

Summary

Today's financial services industry (FSI) providers face mounting challenges in their digital transformation journeys. These include balancing innovation with regulatory compliance, managing operational risk, ensuring data sovereignty, and optimizing cost—all within a rapidly evolving technology landscape. Navigating these complexities requires careful tradeoffs across infrastructure, application design, and service delivery.

LUMOS is engineered to address these challenges head-on. It is a unified architecture that spans on-premises and multi-cloud environments, purpose-built to meet the high demands of the modern financial enterprise. Rather than layering solutions after the fact, LUMOS embeds resilience, security, portability, and AI readiness directly into its design—delivering foundational capabilities that support both day-to-day operations and long-term innovation.

Key Architectural Capabilities:

- Resilience by Design
  LUMOS ensures continuous availability through built-in disaster recovery capabilities, including active-active and active-passive deployments, real-time data replication, automated failover orchestration, and intelligent observability. These features support 24/7 service continuity and minimize the impact of disruptions.
- Security Everywhere
  Security is integrated at every layer of execution. LUMOS leverages containerization for service isolation, CI/CD pipelines for image validation, mTLS and service mesh for encrypted communication, and federated identity with role-based access control across all environments. Data protection is consistent across on-prem and cloud, ensuring full compliance with regulatory requirements.
- True Application and Infrastructure Mobility
  By decoupling applications from infrastructure and standardizing control across environments, LUMOS enables seamless workload mobility. Organizations can dynamically shift services in response to cost, performance, compliance, or availability needs—avoiding vendor lock-in and enhancing agility.
- Unified Data Fabric
  The One-Lake data architecture centralizes governance while enabling access across environments. Through global namespaces, consistent metadata services, and policy-based replication, LUMOS abstracts physical data location and ensures performance, sovereignty, and operational efficiency.
- AI-Ready Infrastructure
  Designed with intelligence in mind, LUMOS supports integrated GPU clusters, federated learning frameworks, and high-throughput, low-latency storage—empowering institutions to deploy AI/ML services at scale without compromising privacy or control. ML lifecycle management is built into DevOps pipelines, supporting continuous improvement and operationalization.

In Summary, LUMOS transforms fragmented cloud strategies into a cohesive, secure, and future-ready platform. By embedding critical capabilities into the core of the architecture, it empowers financial institutions to innovate with confidence, respond to regulatory change with agility, and operate with the resilience and efficiency required in today's high-stakes financial landscape.