

# Transforming Enterprise Data with Lakehouse Architecture at Scale

Richa Solanki | Ally Bank

Conf42.com Prompt Engineering 2025 | November 06, 2025



# The Data Explosion: A Growing Challenge

Global data volumes are projected to reach **175 zettabytes by 2025**, with nearly one-third requiring real-time processing capabilities.

Enterprises today struggle with managing diverse data types—structured, semi-structured, and unstructured—across fragmented systems that create operational complexity and governance gaps.



# Traditional Approaches: The Dual Challenge

## Data Warehouses

**Strengths:** Strong governance, ACID compliance, structured analytics

**Limitations:** Limited scalability, high costs, rigid schemas

## Data Lakes

**Strengths:** Massive scale, flexible storage, cost-effective

**Limitations:** Data swamps, weak governance, no transactional support

- When used in isolation, these architectures create significant gaps in scalability, governance, and organizational agility—forcing enterprises to choose between reliability and flexibility.



# Enter the Lakehouse Architecture

Data Lakehouse Architecture represents a **unifying paradigm** that combines the best of both worlds: the reliability and governance of data warehouses with the massive scalability and cost-efficiency of data lakes.

This architecture eliminates the traditional trade-offs, enabling enterprises to handle diverse workloads—from batch analytics to real-time streaming—on a single, unified platform.

# Core Capabilities of Lakehouse Architecture

1

## ACID Transactions

Full transactional consistency across massive datasets, ensuring data integrity at scale

2

## Schema Enforcement

Flexible schema-on-read and schema-on-write capabilities with validation and evolution support

3

## Real-Time Analytics

Low-latency queries on streaming and batch data without complex ETL pipelines

4

## Unified Governance

Centralized metadata management, data lineage tracking, and compliance controls



# Multi-Zone Storage Model



## Raw Zone

Ingests data as-is from all sources. Preserves original formats with minimal transformation for auditability.

## Refined Zone

Cleanses, validates, and standardizes data. Applies business rules and data quality checks.

## Curated Zone

Produces analytics-ready datasets. Optimized for specific use cases and consumption patterns.

# Data Vault Design Patterns

## Why Data Vault?

Data Vault modeling provides the flexibility and auditability required for enterprise-scale lakehouses, supporting historical tracking and schema evolution without disrupting downstream systems.

01

### Hubs

Core business entities with unique keys

02

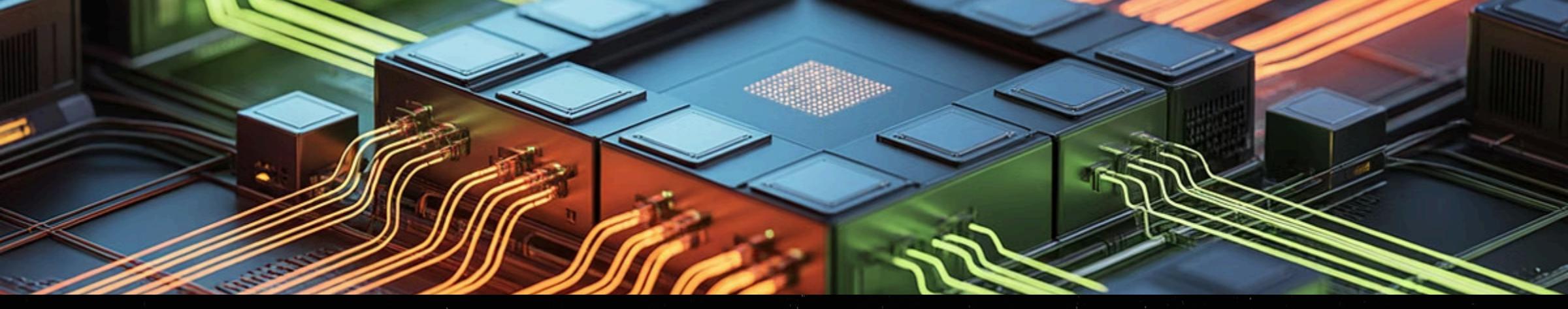
### Links

Relationships between business entities

03

### Satellites

Descriptive attributes with full history



# Performance Optimization Strategies



## Columnar Formats

Leverage Parquet or ORC for compression and query performance. Reduces I/O by reading only necessary columns.



## Intelligent Caching

Cache frequently accessed data in memory or SSD tiers. Accelerates repeated queries without redundant computation.



## Dynamic Partitioning

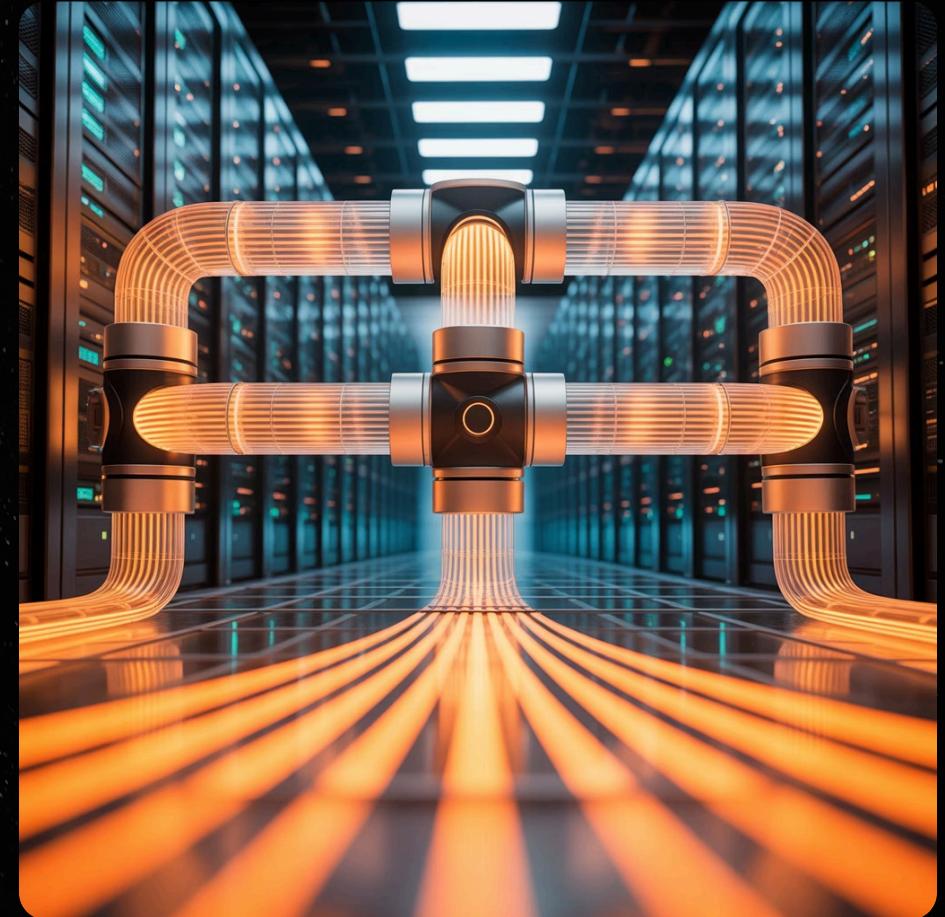
Partition data by time, geography, or business dimension. Enables efficient data pruning and parallel processing.

# Reducing ETL Overhead

Traditional architectures require multiple ETL pipelines moving data between systems, creating **complexity, latency, and points of failure**.

Lakehouse architecture supports **ELT (Extract-Load-Transform)** workflows, where transformation happens in-place on the unified platform. This eliminates redundant data copies and reduces pipeline maintenance.

The result: faster time-to-insight and significantly lower operational overhead.



# Automated Governance Practices



## Data Quality Automation

Automated validation rules detect anomalies, enforce quality standards, and trigger alerts for data issues before they impact downstream analytics.



## Lineage & Audit Trails

Complete tracking of data movement, transformations, and access patterns ensures compliance and simplifies root cause analysis.



## Access Control & Privacy

Role-based access controls, data masking, and encryption ensure sensitive data remains protected while enabling appropriate access.

# Measurable Business Outcomes

**60%**

## Reduced Maintenance

Less time spent on pipeline management  
and data reconciliation

**45%**

## Faster ETL Workflows

Streamlined data movement and  
transformation processes

**40%**

## Lower TCO

Decreased total cost of ownership through  
consolidation and efficiency

Organizations implementing lakehouse architecture report significant improvements in operational efficiency, data quality, and cost optimization while maintaining enterprise-grade governance and compliance.



# Real-World Architecture at Scale

01

---

## Data Ingestion Layer

Batch and streaming ingestion from diverse sources

02

---

## Storage & Compute

Decoupled architecture enabling independent scaling

03

---

## Governance & Catalog

Unified metadata and policy enforcement

04

---

## Analytics & Consumption

Multiple tools and interfaces for diverse use cases

# Implementation Considerations

## Technology Selection

Choose platforms that support open table formats (Delta Lake, Iceberg, Hudi) for vendor flexibility and future-proofing your architecture.

## Migration Strategy

Plan a phased approach starting with non-critical workloads. Build expertise and confidence before migrating mission-critical systems.

## Team Skills

Invest in training for data engineers and analysts. Lakehouse architecture requires understanding of both warehouse and lake concepts.

# Key Takeaways



## Unified Architecture

Lakehouse eliminates the warehouse vs. lake trade-off, providing enterprise governance with cloud-scale flexibility



## Reduced Complexity

Multi-zone storage and ELT workflows dramatically simplify data pipelines and reduce operational overhead



## Business Value

Organizations achieve measurable improvements in cost, performance, and time-to-insight while maintaining compliance

# Thank You

Richa Solanki | Ally Bank

Conf42.com Prompt Engineering 2025

