Lakehouse in Action

Design Strategies, Common Pitfalls, and Real-World Gains

Talat Nauman

Lead Data Architect — Public Sector Education

Data & AI Architecture Sydney | 04 Mar 2026

The Lakehouse Isn’t the Innovation Anymore — How You **Govern**, **Own**, and **Evolve** It Is

What We Actually Built — A Practitioner's Architecture

Consumption and Intelligence

- Portal
- Self-Service Analytics
- Governed Reporting
- Applications
- **AI Agents and Models**

Servicing — Semantic and Presentation Layer

Zero-copy federation — no data duplication

- **Data Marts**
- **Certified Metrics**
- **Reports and Dashboards**
- **APIs**

Lakehouse — Business Domain Data Models

- **Bronze** — Organised by source
- **Silver**
- **Gold**

Organised by business domain

- Stream Compute
- Lake Compute
- ML and Data Science
- Integration and Exchange

Data Lake Storage

- Hot Tier — Frequently Accessed Data
- Warm Tier — Occasionally Accessed Data
- Cold Tier — Rarely Accessed / Archive

Ingestion and Orchestration

- Parameterised ELT
- Event-driven triggers
- Reusable pipelines
- Self-service config

Real Time

- Shortcuts
- CDC
- Mirroring
- Event Stream

Source Systems

- Operational databases
- SaaS applications
- Files and extracts
- Cloud data Lakes
- Streaming events
- APIs

Domain Ownership (Data Mesh)

- Domain teams own data products
- Self-serve ingestion
- Federated governance

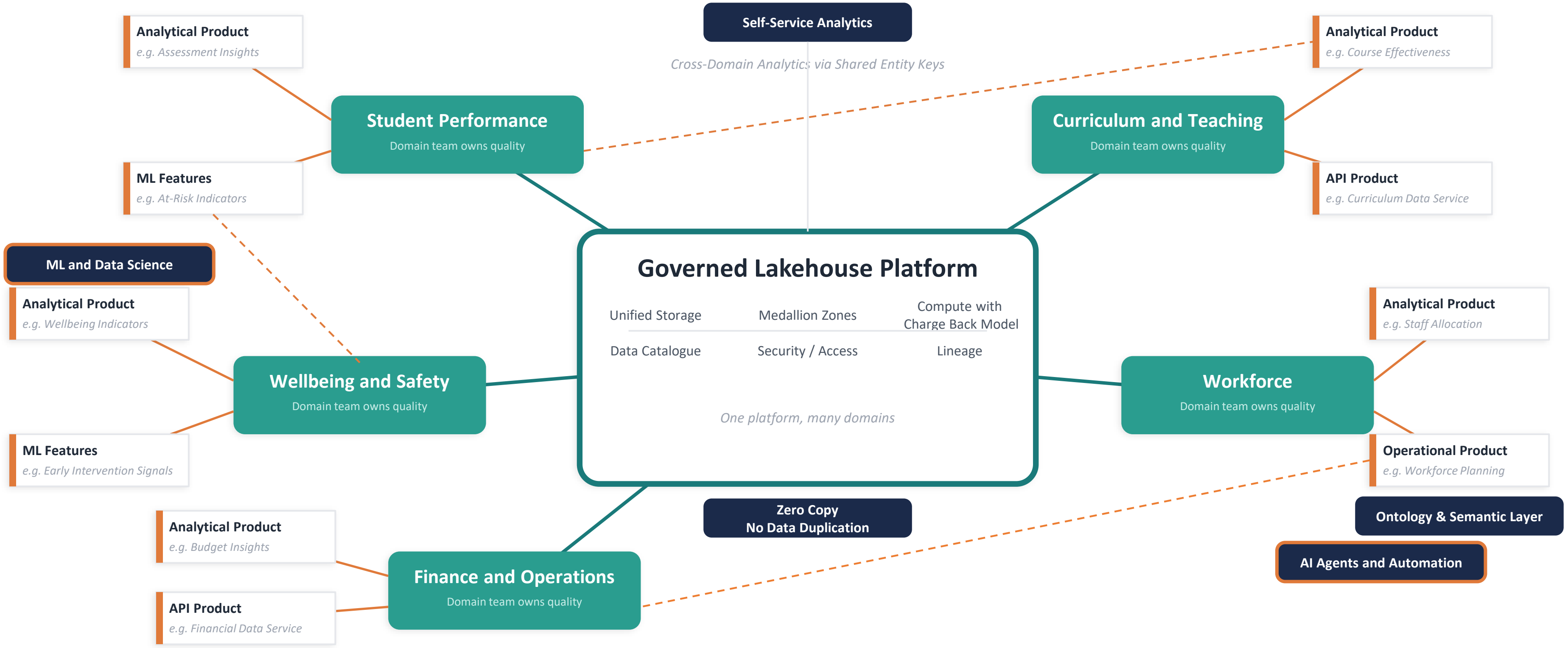
Platform Quality and Operations

- DevOps and CI/CD
- Monitoring and Observability
- SLOs and Reliability
- Cost Management

Data Governance Foundation

- Data Quality
- Access Management
- Metadata Management
- Master Data Management
- Lineage

From Central Lake to Domain Ownership — The Mesh Operating Model



Federated governance — domains own locally, platform enforces globally

Federated Governance Framework

- Data Quality Standards
- Access and Security Policies
- Metadata and Catalogue
- Shared Entity Keys
- Lineage and Auditability

Six Decision Areas That Define Lakehouse Success

ORGANISATIONAL ARCHITECTURE

1. Business-aligned use case sequencing
2. Data ownership and product boundaries
3. Governance by design

OPERATIONAL ARCHITECTURE

4. Performance and reliability architecture
5. Trusted consumption and semantic layer
6. AI-readiness as a design requirement

Organisational Architecture: Sequencing, Ownership, Governance

1

Use case sequencing

Start where the pain is, not where the engineering is easy. Prove value in 2–3 domains, then expand.

2

Data ownership

Domain teams own data as products with SLAs — not a central team bottleneck. Mesh is an operating model, not a technology.

3

Governance by design

Embed from sprint one. Classification, access control, lineage, auditability. Retrofitting costs 5–10x more.

Architect: Ownership boundaries and control-point decisions

Business: How you avoid the platform becoming a queue

Operational Architecture: Performance, Trusted Consumption, AI-Readiness

4 Performance & reliability

Design for operational SLOs, not reactive tuning. Workload isolation, compaction, observability from day one.

5 Trusted consumption & semantic layer

Bridge the gap between technical schemas and business concepts. One vocabulary for analysts and AI agents.

6 AI-readiness

Governed data + business ontology = foundation for safe, scalable AI. A design requirement, not a future concern.

BUSINESS CONCEPTS

"Attendance rate" · "At-risk student" · "Course improvement"

▼

SEMANTIC LAYER / ONTOLOGY

Maps business language to technical assets and metrics

▼

TECHNICAL ASSETS

Tables · Columns · Pipelines · Certified datasets

The Translation Gap

What Breaks in Practice

These are the patterns that repeatedly create delays, performance issues, and trust erosion.

Pitfalls: Migration Thinking, Late Governance, Wrong Metrics

PITFALL	CONSEQUENCE	PREVENTION
Treating lakehouse as a technology migration **Lift & Shift or Uplift?**	Same bottlenecks, new platform. 6–12 months rework.	Hybrid approach — Retire legacy fast and reduce risks Modernize incrementally with new patterns
Governance implemented too late	5–10x retrofit cost. Security friction. Lost confidence.	Redesign ownership — not just pipelines. Governance from sprint one.
Success = data landed, not data trusted	High ingestion, low adoption. Investment without value.	Track certified outputs, usage, and decision-use cases.

Pitfalls: Reactive Performance, Building Ferraris for Push-Bike Riders

PITFALL	CONSEQUENCE	PREVENTION
Performance treated as tuning, not architecture	User trust declines faster than engineers expect.	Design for SLOs, observability, compaction from day one.
Building a Ferrari when users are on push bikes	High capability, low adoption. Platform nobody uses.	Meet users where they are. Invest in onboarding, ontology and enablement alongside engineering.

Four Lenses for Communicating Lakehouse Value

Risk & Trust

Governance and security posture

“Months to weeks” for access approvals

Delivery Agility

Faster onboarding of new sources

“Weeks/months to days/weeks” for time-to-market

Adoption & Confidence

Business trust and platform usage

“Data as value enabler, not cost to justify”

AI-Readiness

Governed foundation for safe AI

“Agents grounded in governed, semantic data”

From Data Landed to Data Trusted — and AI-Ready

BEFORE		AFTER
Data seen as a cost to justify	▶	**Data seen as a value enabler**
Cross-domain questions require weeks of analyst work	▶	**Cross-domain questions answerable through semantic layer**
AI pilots build their own data pipelines	▶	**AI agents query governed, certified data products**
“How much data is in the lake?”	▶	**“How much data is trusted, consumed, and informing decisions?”**

What Success Looks Like

A successful lakehouse is **not defined by ingestion or data volume**

It is defined by **trusted consumption, governed scale, and operational reliability**

Its long-term value is realised when it becomes a **foundation for faster delivery and AI-readiness**

THREE TAKEAWAYS

1

Start with decision frameworks, not vendor features

2

Design governance and operability from day one

3

Build for the user you have today, not the user you wish you had

Questions + Discussion

"How is your organisation measuring lakehouse success beyond data landed?"

Talat Nauman

Lead Data Architect — Public Sector Education

https://www.linkedin.com/in/talatnauman/