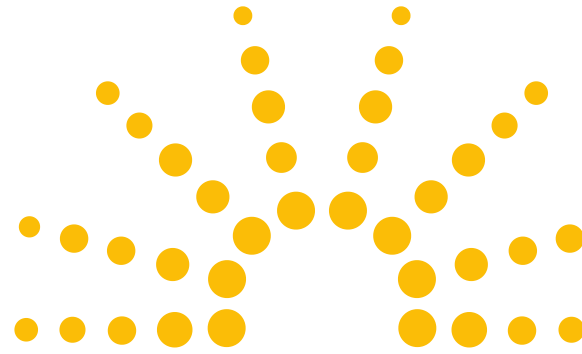




CWX WHITE PAPER · DATA GOVERNANCE & TENANCY

Governance *by design*

An opinionated approach to data governance, compliance and tenancy on the modern cloud data platform.

AUTHOR

Daniel Joanes
Sr. Director of Data & AI

PUBLISHED

October 2026

Governance that lives in the platform, not in people's calendars.

Most enterprises already have the ingredients of good data governance: a catalog, a glossary, access controls and people who care. What they usually lack is a mechanism that keeps those agreements true after the meeting ends.

This paper sets out how CWX approaches governance, compliance and tenancy on cloud data platforms. It is opinionated on purpose. The patterns come from building data and governance platforms for large Canadian banks and for high-scale retail and e-commerce businesses, where getting it wrong shows up as failed audits, duplicated work and numbers that don't match.

Three ideas run through everything that follows.

- **Mechanisms over good intentions.** Definitions, policies and access rules are written as code and enforced by the platform.
- **Catch it early.** A conflicting metric or an unapproved use of personal data should fail in the pull request, not surface in a board deck.
- **Access follows purpose.** People get data through time-bound initiatives with a stated justification, not permanent personal entitlements.

CONTENTS

01	The problem with good intentions	3	05	Metrics as contracts	7
02	Five opinions	4	06	Workflows, not favours	8
03	Compliance by design	5	07	Getting there	9
04	Tenancy through data initiatives	6	A-C	Reference architectures & references	10

Governance is only as good as the *last* conversation.

Two dashboards both report “demand”, and the numbers disagree. Someone notices. A governance lead gets the two authors in a room, and after some back and forth they settle on a definition.

Demand, they agree, is every transaction received before midnight, less any cancellations made before midnight. The definition, the SQL and a description go into the catalog. Everyone moves on.

Then a new analyst joins in another region. They’re busy, they don’t know the catalog exists, or it never occurs to them to look. They write their own version of demand, it ships, and the cycle starts again — often months after the first round of alignment.

Nobody did anything wrong. The analyst did what good analysts do and found a way to get the job done. The problem is structural: the agreement lived in a conversation and a catalog entry, and neither sits in the path of the work.

Good intentions don’t scale. Mechanisms do.

When governance relies on people remembering to check, it becomes a tax rather than a benefit. Teams route around it, the governance function spends its days brokering meetings, and the platform’s long-term investments compete for the same bandwidth as everyday requests from the business.

WHAT WE SEE IN PRACTICE

The bottleneck is rarely tooling. Cloud providers supply excellent building blocks — catalogs, policy tags, masking, lineage. The design work is deciding how they fit together, where they plug into the development lifecycle, and who is accountable for what.



The opinions that shape how we *build*.

Every platform makes trade-offs between freedom and control. These are the ones we make, and why.

1 Mechanisms over good intentions

If a rule matters, the platform enforces it. Documentation records the rule; it shouldn't be the thing that polices it.

2 Fail early, and fail in CI

The cheapest place to catch a conflicting definition or a policy breach is the pull request. The most expensive is production.

3 Policies are code

Versioned, reviewed and tested like any other change. Codifying a policy forces a first-principles question — what is this actually protecting us from? — and the answer is usually less restrictive than the review board assumed.

4 Host sensitive data — don't hide it

Locking personal data in a corner nobody can reach isn't compliance. It just disables the use cases that need it. Hold the data with the controls that let the right people use it.

5 Developer experience is the control surface

The platform owner decides the degrees of freedom. Inside them, work should be fast and pleasant. Outside them, the answer is a clear, automatic no — or a workflow.

The right people, the right data, and no human *gatekeeper*.

Compliance by design means the workflows, approvals, access groups and controls that compliance requires are built into the platform, so no administrator has to decide by hand who can see what.

It rests on metadata. Once every sensitive column carries a classification, policy attaches to the classification rather than to individual tables. That's what makes it scale: a new table holding email addresses inherits the right masking and access rules the moment it's tagged.

STEP 1

Classify

STEP 2

Tag

STEP 3

Enforce

STEP 4

Evidence

- **Classification at ingestion.** Automated scanning, with LLM-assisted tagging for the context scanners miss, proposes classifications that data owners confirm.
- **Policy follows the tag.** Column-level policy tags, dynamic masking and row-level access policies apply wherever classified data lands.
- **Purpose-bound access.** Sensitive data is granted through an approved initiative (chapter 04), never as a standing personal entitlement.
- **Evidence by default.** Every grant, approval and query is logged and mapped to the control it satisfies.

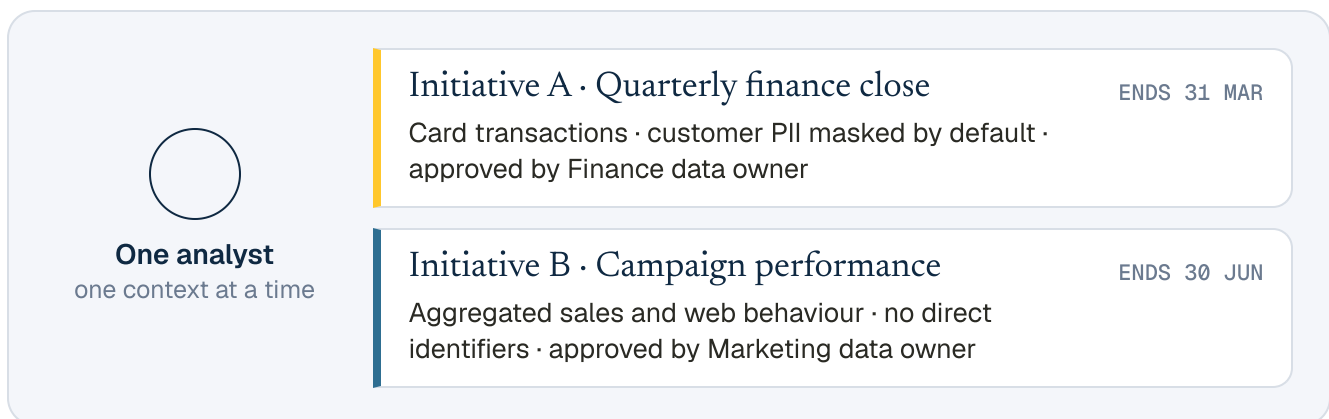
AUDIT READINESS

With the right metadata tied to the right rules, reporting against ISO 27001, SOC 2 or privacy regulation becomes something you run, not something you mobilize for. We've seen audits pause whole teams for weeks while people combed through data dictionaries. It doesn't need to work that way.

Access follows *purpose*, not people.

The same analyst might need card-transaction data for a finance report in the morning and nothing of the sort for a campaign analysis in the afternoon. Personal entitlements can't express that. Initiatives can.

A data initiative is a governed unit of work with a purpose, a data scope, an accountable owner and an end date. Access is justified and approved once, at the initiative level. People join initiatives and take on a role inside them. When they switch work, they switch context — and the access that belonged to the other context stays behind.



- **Isolation by default.** Each initiative gets its own workspace, so data used for one purpose doesn't quietly drift into another.
- **A better question for approvers.** "Is this purpose and scope reasonable?" is far easier to answer well than "should this person see this table?"
- **No single points of failure.** Skills move between initiatives. You're never stuck with one person who is allowed to do fraud analytics.
- **Access expires.** Initiatives end or are recertified, and access is revoked automatically.

A THIN HARNESS, NOT A NEW PLATFORM

Cloud platforms don't provide initiatives out of the box. We build them as a light layer over native identity, project and policy primitives — no new data store and no proprietary runtime. Appendix B shows the reference design.

Certified metrics, enforced at the *pull request*.

Once the business has agreed how a metric is defined, it should be very hard to drift from that definition by accident.

We treat certified metrics and data products as contracts: a definition, an owner, a schema, quality expectations and the consumers who depend on it, all held as code next to the transformations that produce them. The catalog is generated from that source of truth, so it can't quietly fall out of date.

The CI pipeline then does the remembering. When an analyst introduces a calculation that looks like an existing certified metric, the build flags it and points them to the certified version. When a change would break a contract downstream teams rely on, it doesn't merge without the owner's sign-off.

FAIL governance / metric-conflict

models/apac/daily_demand.sql defines a measure named demand

A certified definition already exists: **commercial.demand**

Owner: Commercial Finance · certified 2026-03-14

→ Reference the certified metric, or request a regional variant.

FIG. 1 · WHAT A NEW ANALYST SEES ON THEIR FIRST PULL REQUEST

The point isn't to block analysts. It's to catch the misunderstanding on day one, kindly and automatically, rather than months later in a reconciliation meeting.

OWNERSHIP STAYS WITH THE BUSINESS

A governance council still decides who owns each metric and how it's defined. The platform simply makes sure that decision survives new people, new teams and new regions.

Every meaningful change triggers the right *workflow*.

Good HR systems solved this years ago. You can't change your own title, pay or manager; the system routes the request to whoever is allowed to approve it and records the outcome. Data deserves the same discipline.

We map the changes that matter to the workflows they should trigger. Low-risk work stays frictionless. Higher-risk work goes to an accountable person, with the context they need to decide quickly. The thresholds should fit the business — a beverage retailer and a bank shouldn't have the same ones — but every change leaves a trace.

CHANGE	WHAT HAPPENS	APPROVER
New model in a sandbox	Proceeds immediately; classification scan runs in the background	None
New model in a governed domain	CI checks contracts, metric conflicts and policy	Peer reviewer
Change to a certified metric	Held for approval; downstream consumers notified	Metric owner
Request for classified data	Initiative created or extended, with justification	Data owner + privacy
Policy exception	Time-boxed exception logged against the control	Risk & compliance

TABLE 1 • AN ILLUSTRATIVE CHANGE-TO-WORKFLOW MAP

Governance usually runs on three tracks: business ownership, the administrative work of aligning and documenting, and the digitized workflow that holds it all in place. The first two depend on the third. Without it, council decisions and catalog updates have nowhere to live once the meeting is over.

A practical path from today to *business as usual*.

Few platforms can jump straight to enforcement. Adding controls to existing assets breaks things downstream, so the first job is getting to a clean, known state — then turning the dial from advice to enforcement.

PHASE 1

Discover & classify

Inventory the estate. Apply an 80/20 taxonomy to the assets that matter most. Shortlist the metrics worth certifying.

PHASE 2

Codify

Write policies, metric contracts and the initiative model as code. Run CI checks in warn-only mode so teams can adjust.

PHASE 3

Enforce & operate

Switch checks from warn to block. Automate compliance reporting. Run with a lightweight governance council.

The discovery phase is where most of the effort goes, and it's worth being honest about that. At one bank we started from roughly 800,000 tables with no taxonomy at all. Tagging all of it was never the goal; finding the small share that carried real risk and real business value was.

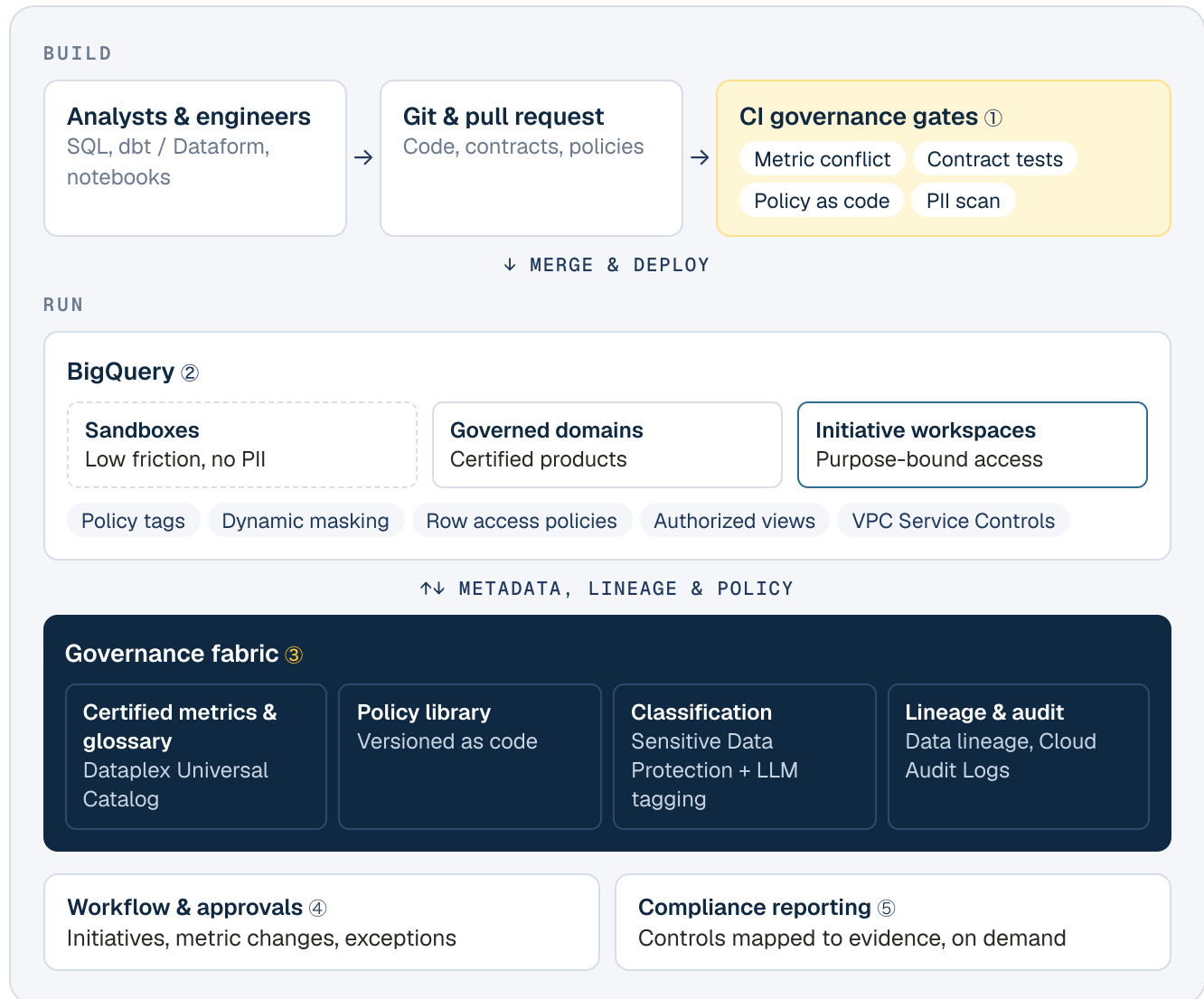
Where we'd start

Pick one domain with a visible pain point — usually a contested metric like demand — and one sensitive dataset that's currently locked away. Prove the pattern on both in a matter of weeks, show the business what changed, then expand domain by domain.

Build the mechanism once, and the alignment outlives the meeting.

The governance fabric on Google Cloud

How certified definitions, policy and classification plug into the development lifecycle and the warehouse. The same shape applies on other clouds with equivalent services.



① **Gates.** Run on every pull request; start in warn-only mode, then block.

③ **Fabric.** One horizontal layer of definitions and policy, applied consistently to every zone.

② **Zones.** Freedom is highest in sandboxes and narrows as data becomes certified or sensitive.

④–⑤ **Outputs.** Approvals and audit evidence are by-products of normal work, not separate projects.

Initiative-based tenancy

The lifecycle of a data initiative, and how it isolates access to source domains. Built entirely on native identity, project and policy primitives.

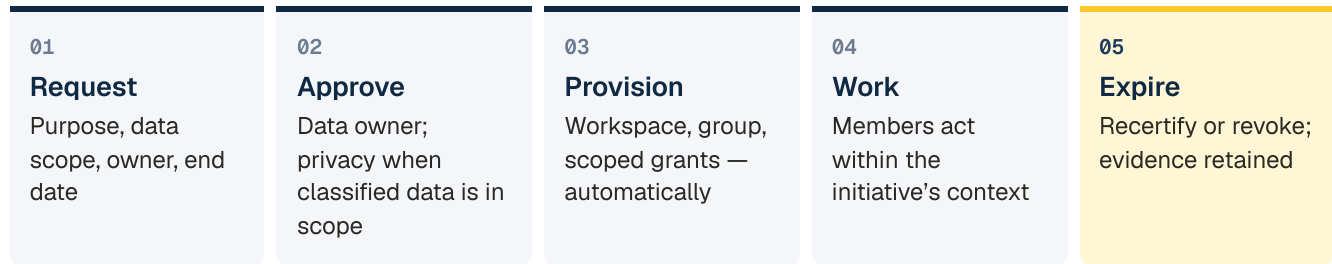


FIG. 2 · INITIATIVE LIFECYCLE

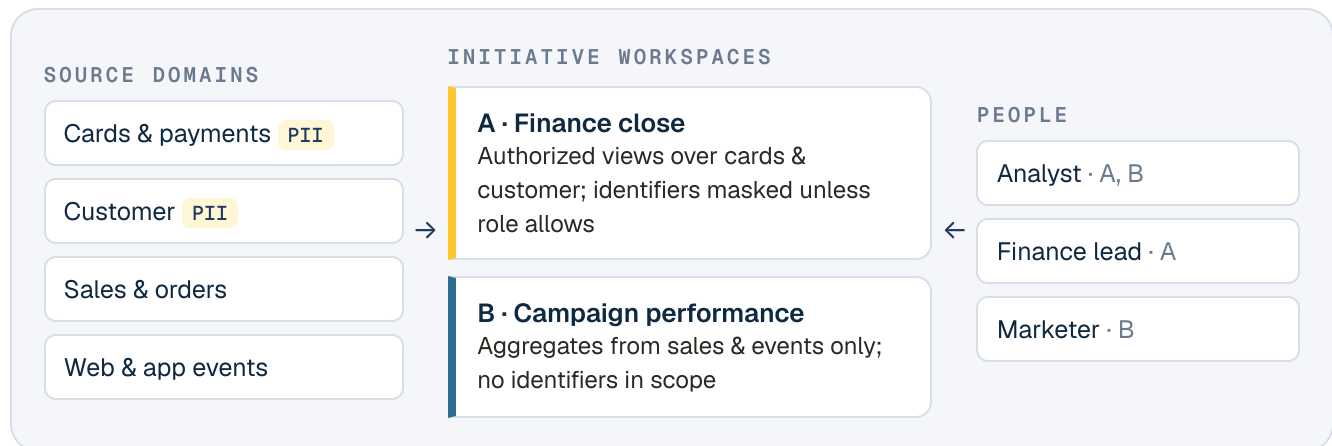


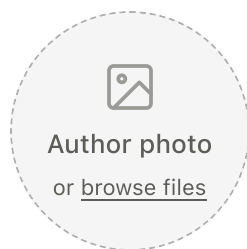
FIG. 3 · ISOLATION MODEL – ONE CONTEXT ACTIVE PER SESSION

CONCERN	GOOGLE CLOUD PRIMITIVE
Workspace	A project or dataset per initiative, created from a template
Membership & roles	Cloud Identity groups; IAM conditions with expiry
Sensitive access	Policy tags, data masking, authorized views, row access policies
Registry & workflow	Initiative records and approvals as code, applied through CI
Evidence	Cloud Audit Logs exported to BigQuery for reporting

Proven in regulated, high-scale environments

The patterns in this paper come from production platforms, not slideware. A selection of organizations where our teams have designed and built data platform and governance capabilities:

 Scotiabank logo or browse files Scotiabank BANKING	 Loblaw Digital logo or browse files Loblaw Digital RETAIL & E-COMMERCE
 Client logo or browse files Client name INDUSTRY	 Client logo or browse files Client name INDUSTRY



ABOUT THE AUTHOR

Daniel Joanes

Senior Director of Data & AI, CWX

Daniel came to consulting from product engineering. He spent several years at a large e-commerce company in Asia working on catalogs of 100 million products and 300 million active users, then built the loyalty platform for a payments company in Japan. At CWX he has led the design of data and governance platforms for major Canadian banks — bringing the pace of product teams to regulated environments.



We help regulated enterprises build data and AI platforms with governance *woven through every layer.*

To continue the conversation, contact Daniel Joanes or your CWX account team.

CWX . AI

