Fivetran + dbt Labs

From Data Chaos to Connected Intelligence



Thomas Sutton-Walsh
Account Executive
Fivetran + dbt Labs



Indrajeet Singh
Associate Director, Data & Analytics Engineering
University of Auckland

Agenda

Two parts

Fivetran + dbt Labs

- Fivetran & dbt Labs positioning:
  - Agents are the new primary data consumers.
  - Are you ready?

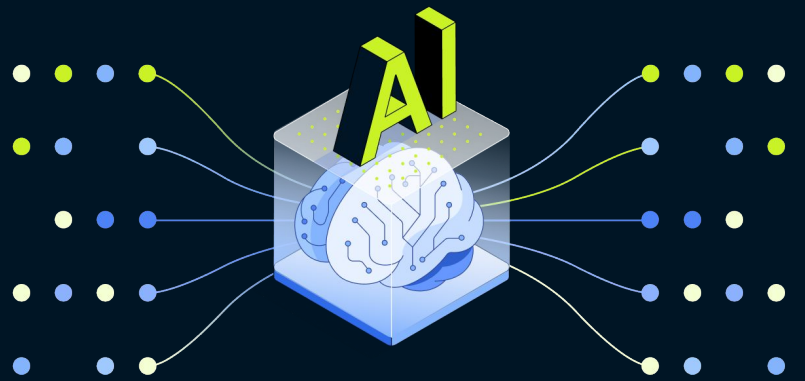
University of Auckland

- Data journey
- Crawl, Walk, Run, Sprint
- What's next?



Your data stack was built for humans,





but agents are the new primary data consumers.

Today’s data stack breaks at agent scale

01

Trust breaks

- Incomplete data
- Inconsistent metrics
- No lineage



02

Scale stalls

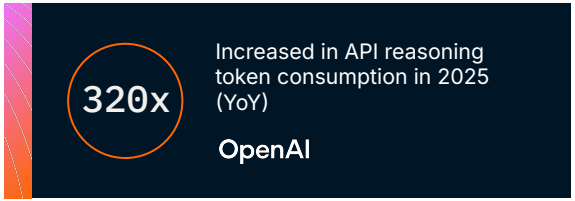
- Ungoverned data
- Fragmented context
- Locked in formats



03

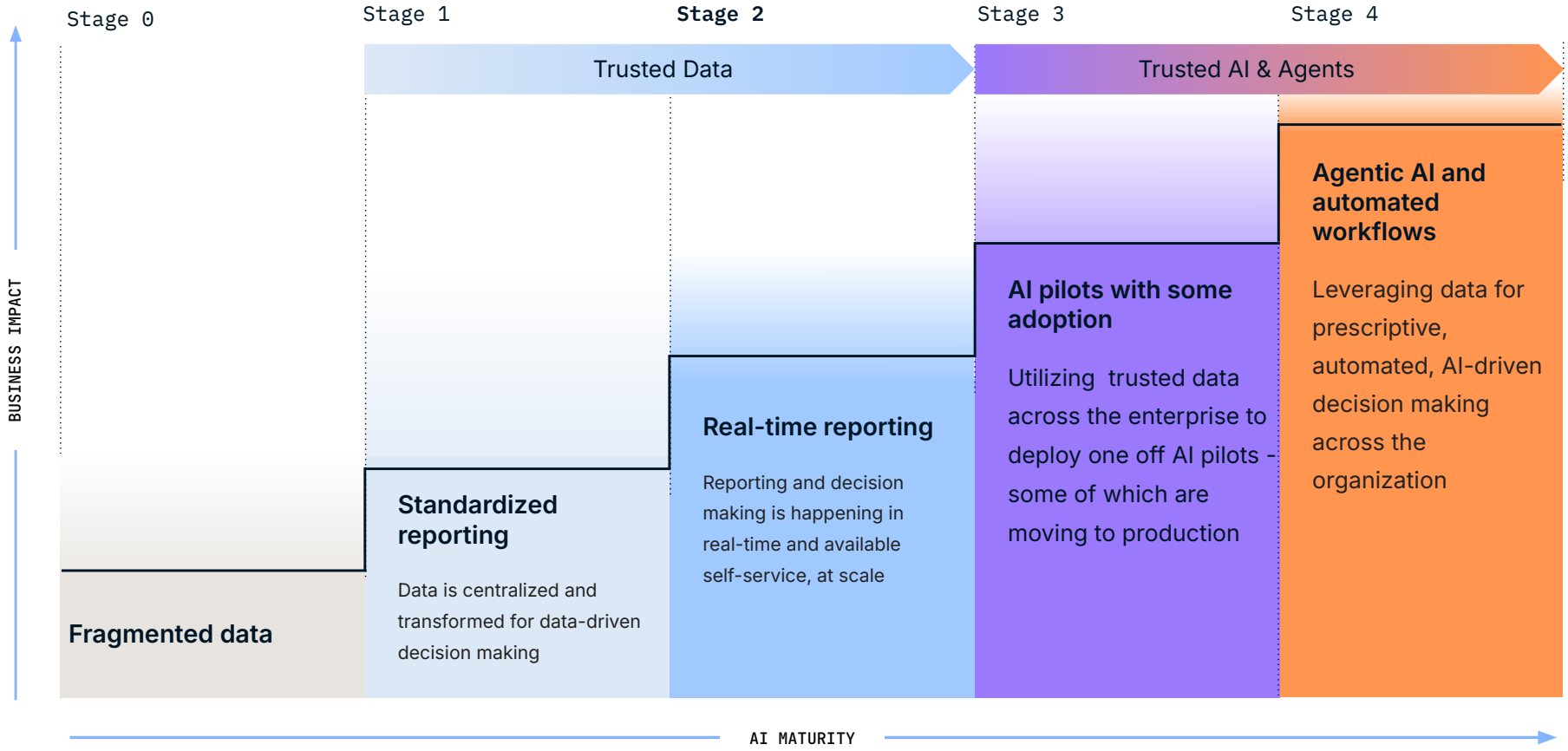
Costs explode

- Repeated retrievals
- Unoptimized modeling
- Linked storage and compute



Results in unreliable agents. No ROI. Runaway costs.

We know organizations are at different stages



A trusted data infrastructure

Builds the foundation for analytics, operations, and agentic AI

Flexible by design

Portability increases business agility and decreases future rearchitecture costs

Fresh and accurate

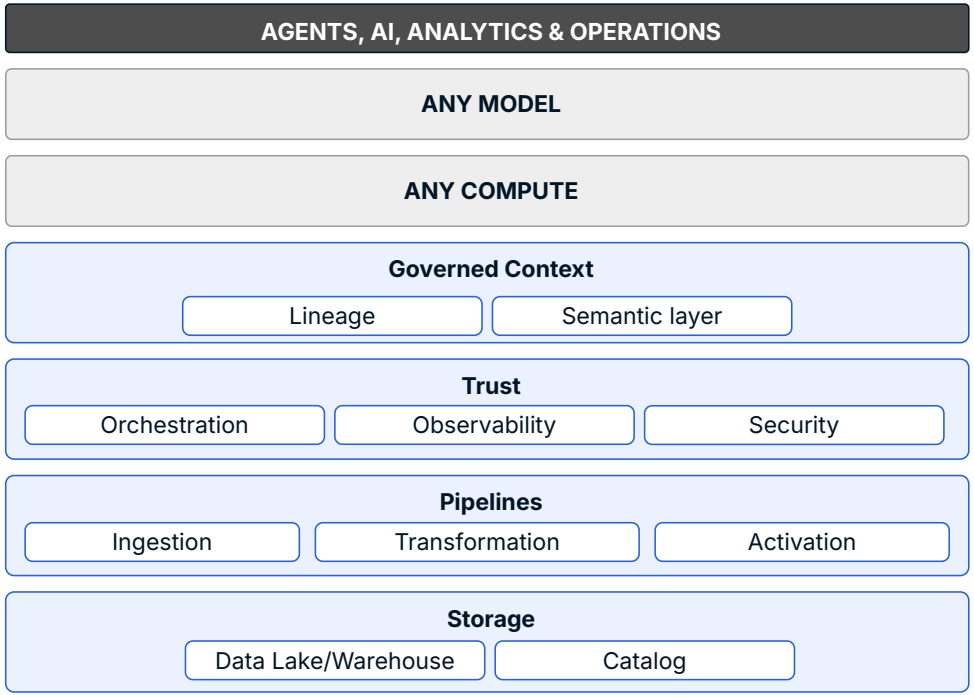
Business decisions happen faster with less data delays or roadblocks

Trusted by default

Greater innovation occurs when every business user trusts every business answer

Efficient at scale

Unit costs deflate as volume scales



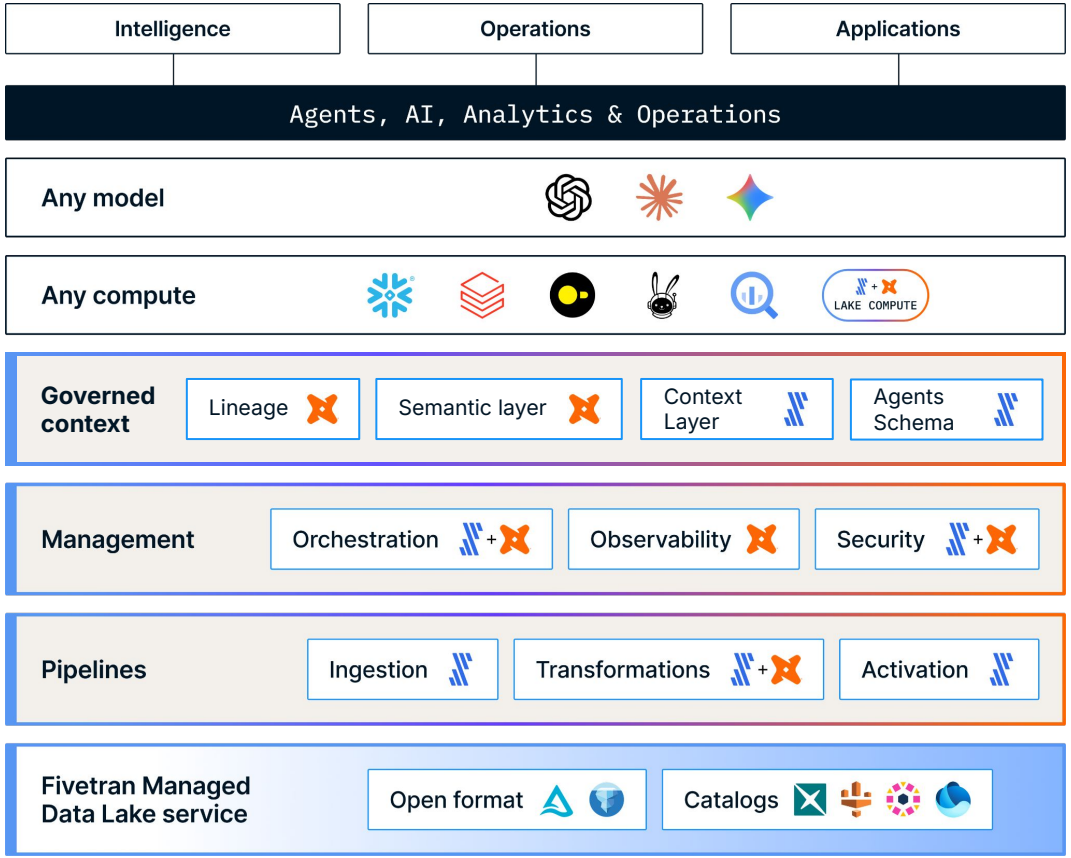
Fivetran + dbt Labs are your trusted data infrastructure

Flexible storage + compute
Any cloud, any compute, any LLM

Reliable data movement
Centralizing all of your data

Rich trust and context layer
Traceable information for every output

Scalable infrastructure
For any workload or use case



The ROI of trusted data infrastructure

Make money
Faster agent deploy

Move from pilots to production faster

Reduce risk
Built-in trust

Decrease hallucinations and data leaks

Save money
Lower AI costs

Reduce compute + token waste

Prepare for the future
Portable data

Adapt to innovation without rebuilding

46%

reduction in data warehouse costs*

Waipapa Taumata Rau University of Auckland

OUR SNOWFLAKE JOURNEY

From Data Chaos to Connected Intelligence

How we are gradually building the capability — crawl, walk, run, and the sprint ahead.

CRAWL	WALK	RUN	SPRINT
Survive	*Stabilise*	*Enable*	*Anticipate*
		WE ARE HERE	

Indrajeet Singh · Associate Director, Data & Analytics Engineering

Built on Snowflake

Our data strategy serves Taumata Teitei

Taumata Teitei — our Vision 2030 and Strategic Plan 2028 — sets the ambition. Our data strategy turns that vision into trusted information that decisions can be built on.

TAUMATA TEITEI

Vision 2030 · Strategic Plan 2028

Honouring Te Tiriti o Waitangi, partnering for impact, and building a future-ready Aotearoa New Zealand.

- Research & innovation
- Learning & teaching
- Equity & belonging
- Sustainability
- Partnerships

Our Data Strategy

"Data reflects the vibrant life of our University — utilising data, enabled by data, driven by data."

- **Make data available.** Bring institutional data into one trusted platform.
- **Define data clearly.** Shared business definitions, owned and certified.
- **Govern access by design.** Security and ethics built into the layer, not added later.
- **Build a data-fluent culture.** Equip every team to use trusted data with confidence.

Strategy first, then technology. Snowflake is how — Taumata Teitei is why.

Waipapa Taumata Rau University of Auckland

Key facts and figures 2025

University of Auckland Annual Report 2025 · figures as at 31 December 2025

No.1 ranked New Zealand university

Top 100 university in the world

Top 30 in the world for sustainability

Top 50 in the world in 5 subjects

STUDENTS

49,901

student headcount

38,142 Local

11,759 Overseas

EFTS

39,183

equivalent full-time students

29,392 undergrad · 9,792 postgrad

+7.1% on 2024

FULL-TIME EQUIVALENT STAFF



6,250

Total FTE

3,796 professional

2,454 academic

RESEARCH

5,622

research publications

70.4% in top-20% journals

Adjusted net surplus

$46.5m 2.9% of revenue

International research connections

3,856 international co-authored works

Degrees completed

12,438 completed in 2025

Source: University of Auckland Annual Report 2025, key facts and figures (pp. 10–11) and Chancellor's review (p. 6).

Crawl. Walk. Run. Sprint.

Each stage built a new capability the next one depends on. We're not finished — right now, we're finding our run, and the sprint is what we're building toward.



THEN

CRAWL

Capability: survive

Chaos & firefighting

Fragile pipelines, manual fixes, conflicting numbers. The capability we recognised we lacked: trust.



BUILT

WALK

Capability: stabilise

A governed foundation

One platform, layered modelling, single source of truth — built once and inherited everywhere.



WE ARE HERE

RUN

Capability: enable

Answers in everyone's hands

TUI live, a feedback loop, real use cases shipping. Still building — gaining pace.



AHEAD

SPRINT

Capability: anticipate

Answers before they're asked

Forecasting, what-if and ML-served signals — the capability we're building toward next.

The foundation is the reason every later stage is even possible.

Why some datasets need special treatment

These data products carry real institutional stakes — financial, reputational and legal.

EFTS

Student funding data

Enrolment counts submitted to TEC determine government funding allocations.

- $ TEC funding
- + TEC compliance
- @ Reputation
- ~ Enrolment planning

Research

Research outputs

PBRF performance, contestable grants and research strategy depend on accurate data.

- $ Research funding
- + PBRF reporting
- @ Rankings & profile
- ~ Strategy decisions

Staff

Workforce data

Headcount, FTE and payroll data underpin statutory reporting and HR decisions.

- $ Payroll accuracy
- + Statutory reporting
- @ Employment law
- ~ Workforce planning

⚠ Without a central model — what goes wrong

Conflicting numbers

Different teams answer the same question differently. Leadership loses confidence in every figure presented.

Compliance exposure

Misaligned data submitted to TEC, PBRF or auditors creates direct financial and legal risk.

Reputational damage

Contradictory reports to Council, external bodies or government undermine institutional credibility and trust.

Duplicated cost & effort

Multiple teams rebuild the same logic independently, wasting budget, time and analytical capability.

Modernise without disruption

A staged approach kept operations stable while the governed Snowflake foundation grew alongside the legacy platform.

THE BALANCING ACT

Modernise fast
AI, real-time insight and self-service analytics

Stay stable
Protect operations, compliance and auditability

Scale for future
Build a platform ready for growth and new AI use cases

Balance point: evolutionary modernisation + governed innovation

THE MIGRATION PATTERN

1. **Start with high-value domains** — Move priority reporting first — where trust and business value matter most.
2. **Abstract access** — Put a semantic / access layer in front so users are not tied to old structures.
3. **Build in parallel** — Create governed Snowflake models while existing reporting continues to operate.
4. **Redirect and retire** — Validate, switch traffic, and decommission legacy logic domain by domain.

Future-ready by design — Rules, ownership, metadata, access control, lineage and validation are built into the new layer — not recreated later in every report.

Waipapa Taumata Rau University of Auckland

STAGE 02 · WALK · STABILISE

One platform, one path for every dataset

Fivetran replicates source systems into Snowflake every 30 minutes. From there, the same data becomes more usable, more aligned and more trusted as it flows through governed layers.

SOURCE SYSTEMS

- Workday · HR & Payroll
- CS9 · Student System
- Finance System
- aPay · Payroll

Fivetran
every 30 min

RAW

Raw replication

Landed direct from source — close to original structure. Built for traceability, not yet business-ready.

DW

Curated enterprise logic

Cleansed and conformed. Shared business rules and joins; refreshed overnight as a reusable foundation.

PM

Presentation models

Audience-ready datasets with consistent definitions and metadata — optimised for broad, trusted reuse.

WHAT THIS GAVE US: *one consistent ingestion path, layered modelling, and a foundation we could actually build on. Build once.*

Waipapa Taumata Rau University of Auckland

STAGE 02 · WALK · STABILISE

One source of truth — every tool inherits it

Conversational analytics, decision intelligence, reporting and machine learning — every experience draws from the same governed semantic layer. No re-derivation. No drift.

FOUR EXPERIENCES — ONE FOUNDATION

TUI
Conversational analytics

Connected Decision Intelligence
KPI signals · what-if

Reporting & BI
Dashboards & official reports

Machine Learning
Predictions & served signals

Snowflake Semantic Layer

University-approved business logic, applied once and served consistently to every experience.

- Business rules
- Governed definitions
- Row-level security
- Equity suppression
- Snapshot logic
- COUNT(DISTINCT)

Snowflake Institutional Datasets Students · Staff · Research · Finance · Equity and other trusted domains

Same data, same meaning, every time — across every tool.

Waipapa Taumata Rau University of Auckland

STAGE 02 · WALK · STABILISE

Shared rules for shared data

The hardest part wasn't technical — it was alignment. We agreed, with our stakeholders, how data should be created and maintained.

1



Reuse certified sources first

If a trusted enterprise asset already exists, reuse it — so definitions and logic stay aligned across every team and tool.

2



Make the semantics explicit

If a new model is genuinely needed, capture purpose, definitions, owners, metadata and quality rules up front — with stakeholders, not a closed group.

3



Publish once, for broad reuse

Release only once it's well-defined and governed. The shared outcome: everyone interprets the same thing the same way.



Critical products get stronger governance. EFTS funding, research outputs and workforce data carry financial, statutory and reputational stakes — they're treated differently from ordinary extracts.

Build once.
Define clearly.
Trust everywhere.

Waipapa Taumata Rau University of Auckland

STAGE 03 · RUN · ENABLE

Academic Heads, now in TUI

TUI — Te Whaiwhakaaro UoA Insights — lets Heads of School ask in plain language and get trusted answers, instead of chasing numbers across systems.

BEFORE

Heads of School need different data sources for one answer — time lost to admin instead of leadership.

- Enrolment
- Finance
- Research
- HR
- Supervision
- Course

WITH TUI — JUST ASK, IN SNOWFLAKE INTELLIGENCE

Real questions Heads ask, in plain language — answered from trusted data:

STUDENTS
"Show me first-year retention for the last three years."

STAFF / HR
"What's the total leave balance across everyone who reports to me?"

RESEARCH
"Research income by Faculty for the last three years."

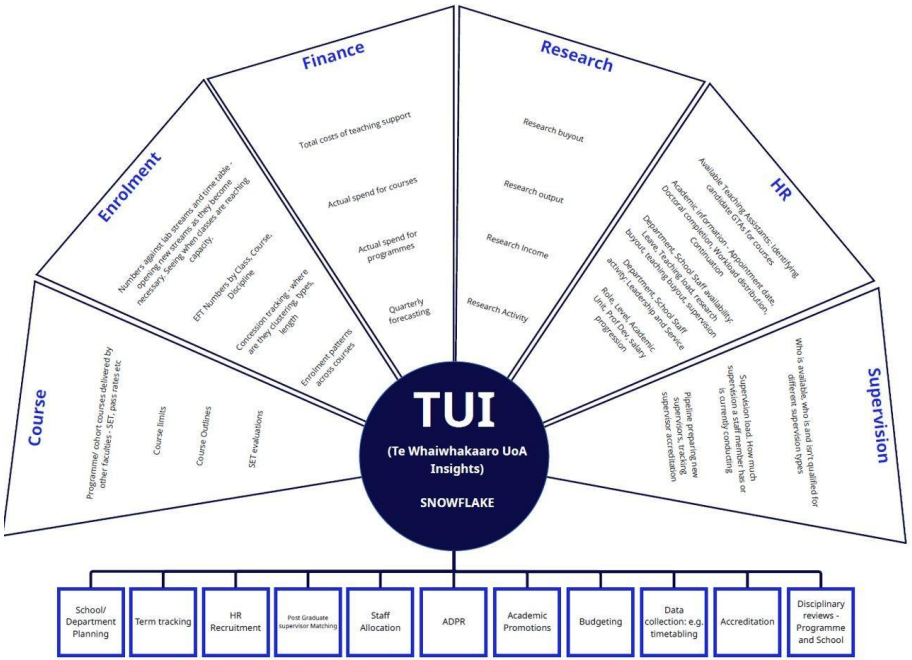
One place. Same trusted numbers everyone uses.

The win: Heads spend less time wrangling data and more time leading learning, teaching and research — and their support staff move from compiling numbers to advising on them.

STAGE 03 · RUN · ENABLE

One brain for the university data

TUI pulls many areas of University data into one place — and powers the everyday decisions academic leaders make across all of them.



How it all comes together

From trusted data to insights, AI and impact

Bring data in, clean & conform it → *Shape it into reusable data products* → *Publish one fast, secure copy* → *Give it shared business meaning* → *Put it to work, for people and AI agents*

BUSINESS VALUE

- One source of truth
- Consistent definitions
- Trusted metrics
- Faster insights
- Better decisions

1. DATA FOUNDATION (CURATED)

Source Systems

- Student Information System
- Learning Management System
- Finance System
- CRM & Advancement
- Files & Documents
- External Data Sources

Raw / Landing Layer

- Ingested via Fivetran / APIs / File Loads
- Immutable, history retained
- No business logic applied

Cleansed & Conformed Layer

- Standardised, validated, and de-duplicated data
- Surrogate keys assigned
- Conformed across domains

2. DATA PRODUCT BUILD (ENGINEERING & ANALYST WORKSPACE)

Built by Data & Analytics Engineering

Access restricted to approved analyst group

Dimension Tables

e.g. Dim Student, Dim Programme, Dim Faculty, Dim Course, Dim Date (Snapshot Date), Dim Status

- Business-friendly attributes and hierarchies

Fact Tables

- fact_efts_snapshot
- fact_fte_snapshot

Measures included (examples):

- EFTS measures: gross EFTS, funded EFTS, domestic EFTS, international EFTS
- FTE measures: occupied FTE, budgeted FTE, funded FTE, vacant FTE

Data Product Views

- Business logic, joins, filters, and calculations applied
- Reusable, consistent, and performance-optimised
- Defines the business model

3. PUBLISHED LAYER (DYNAMIC TABLE)

Dynamic Table

(Replica of View)

- Replicates the Data Product View (1:1)
- Incrementally refreshed automatically
- Persistent storage for performance & reliability
- Reduces compute cost compared to re-querying views
- Public interface for consumption – no direct access to underlying views, facts or dimensions
- **RBAC, RLS and ABAC apply at this layer and flow downstream.**

Refresh Strategy

- Incremental refresh
- Schedule based on data latency needs
- Monitor & alert on failure

4. SEMANTIC LAYER (SEMANTIC MODEL)

Semantic Model

(Built on Dynamic Table)

- Business definitions, measures, relationships and hierarchies
- Calculated measures, KPIs and business logic
- Row-level security (RLS) applied
- Certified model for enterprise consumption

Governed Access

- Role-based access (RBAC)
- Row-level security (RLS)
- Attribute-based access control (ABAC)
- Managed by data governance principles

5. CONSUMPTION & INTEGRATION

Reporting

Trend: fewer built reports, more AI-generated

Reports & Artifacts in AI tools

Most reporting: on demand in Snowflake, NebulaOne etc.

Strategic Dashboards (Power BI)

A few, certified: tracking the University's strategic direction

AI / Conversational

- Natural language queries
- AI agents

Data Applications

- Custom apps
- Portals
- Embedded analytics

Data Sharing

- Secure sharing
- APIs / Exports

Everyone is a Data User

People and AI agents alike — and a platform that starts the conversation

Every person is a data user

- Students, academics, professional staff and leaders all use data daily
- Data literacy for all: governed self-service pitched to each role
- Analysts go deeper via the approved consumption layer

Non-human users: a growing role

- AI agents and copilots (e.g. TUI) query and act on behalf of people
- Automated workflows, services and integrations consume data directly
- First-class users: own identity, access rules and audit trail

Data that talks to us

- Proactive alerts on anomalies, thresholds and emerging risks
- Personalised insight digests, explained in plain language
- Agents that start the conversation and suggest next actions

Read reports → Ask questions → Data talks to us

GOVERNANCE (ALWAYS APPLIED)

Access & Security

- Roles & privileges (RBAC)
- Row access policies (RLS)
- Attribute-based access control (ABAC)
- Column masking

Data Quality

- Data quality rules
- Thresholds
- Monitoring & alerts

Metadata & Lineage

- Business descriptions
- Data lineage
- Change history

Audit & Compliance

- Usage audit
- Policy enforcement
- Regulatory compliance

Governed access ensures people and agents see the right data, in the right way, for the right purpose.

ANALYST ACCESS ONLY

Direct access permitted only for:

- ✓ Data discovery and analysis
- ✓ Data quality checks
- ✓ Prototyping data products and setting build requirements
- ✓ Business logic validation

END USER ACCESS (ALL CONSUMPTION)

All reporting, AI tools and agents, integration and analytics must use:

Dynamic Table + Semantic Model

Never connect directly to Views, Fact Tables or Dimension Tables.

GLOSSARY

Dynamic Table	Auto-refreshed stored copy of the data product view used for performance and cost efficiency.
Semantic Model	Business-friendly layer with trusted metrics, relationships and logic.
Snapshot Date	The date the data represents.
Artifact	AI-generated report, chart or summary made on demand in an AI tool.
Strategic Dashboard	Small, certified Power BI set tracking strategic direction.
RBAC · RLS · ABAC	Role-based access · Row-level security · Attribute-based access control.

LEGEND

- Data Flow
- Access (Restricted)
- Consumption (Governed)

KEY PRINCIPLES

- **Trust** — Curated data, consistent definitions
- **Performance** — Optimised delivery, lower cost
- **Security** — Role-based access, privacy by design
- **Reuse** — Reusable data products
- **Scalability** — Built to grow with the university

The feedback loop is the governance engine

Every TUI answer can be challenged. When a user says "that shouldn't be using that dataset," they're not a problem to manage — they're telling us exactly where meaning hasn't been agreed yet.

A real example

"Why does my domestic total differ from the Council pack? I get 14,250 — the pack shows 14,118. Both say week-3 snapshot."

The cause: two different definitions of "domestic." One is funding-source based; the Council pack is residency-based. Neither is wrong — they answer different questions.

The feedback surfaced a definition gap no audit had caught.

WHAT EVERY FLAG GIVES US

Alignment on meaning

We see exactly where two teams interpret the same word differently — and fix the definition once.

Real governance conversations

Feedback starts the data-governance discussion with evidence, not opinion — on the cases that actually matter.

Priorities set by real need

We see what people genuinely ask and struggle to answer — so we build against demand on the ground, not guesswork.

Waipapa Taumata Rau University of Auckland

STAGE 03 · RUN · ENABLE

People follow a question, not a tool

Same governed foundation, same definitions. A single concern about engagement can move through all four experiences — and nothing has to be reconciled along the way.

1

Predict the signal

Machine Learning

An engagement-risk model flags a faculty whose pattern is starting to soften — served back as a governed measure.

>>

2

Track & model

Connected Decision Intelligence

The KPI tile reflects it first. “If we lift engagement 5%, what does retention do?” — what-if lives here.

>>

3

See the detail

Reporting & BI

Open the dashboards behind the KPI — drivers broken down by faculty, programme and week, on the same definitions.

>>

4

Drill lower

TUI

Conversational follow-through: “Which 200-level courses are driving the drop?” Asked in plain language. Same trusted numbers.

One thought. Four experiences. Zero reconciliation.

Transcripts: from vendor to in-house

We couldn't process international transcripts ourselves — they were read by hand, which slowed every application. A solution used to mean an outside vendor. Now with capability in Snowflake, we do it ourselves.

BEFORE — NO WAY TO DO IT OURSELVES

- ✕ **Read by hand, every time**
  Transcripts were checked manually — slow and people-heavy.
- ✕ **It slowed applications down**
  Every manual check added time to a student's application.
- ✕ **A vendor was the only fix**
  To automate it, we'd have had to pay an external vendor.

NOW — DONE INSIDE THE UNIVERSITY

- ✓ **Our data already sits on Snowflake**
  The platform has capability to extract data from transcripts where they live.
- ✓ **The platform reads them, and improves**
  It pulls the detail out automatically — and keeps getting better.
- ✓ **We build it our way**
  In-house, faster, cheaper — and we choose how to use it.

We turned work we'd have paid a vendor for into something we own, shape and keep improving.

Where we're heading next

COVID taught us how fast the questions can change. We answered then — heroically, on scattered data. The foundation is how we answer faster next time, and start building before the question lands.

THEN · speed under pressure

Engagement signals, contact tracking and online-attendance views stood up fast — on scattered data, with heroic manual effort, racing the clock.

NOW · speed by design

The same kind of signal now sits on one governed foundation. What took weeks of scramble can be assembled in days — from data we already trust.

Next: answering the questions before they're even asked

See what's coming

Forecast enrolments, funding and workload before the year starts.

Spot risks early

Flag the students and areas that need attention, automatically.

Test decisions safely

Trade intake, funding and intervention against the trusted baseline.

Waipapa Taumata Rau University of Auckland

WHERE WE ARE

Crawled. Walked. Now running.

Still building — and the sprint is what we're building toward.

1

Evolve, don't explode

Modernise without disruption. Deliver value in months — grow the new alongside the old.

2

Transparency is governance

Let users challenge the answer. Their feedback aligns meaning and sets real priorities.

3

The foundation is the strategy

A governed, shared platform makes every next capability — AI included — faster and more trusted.

Indrajeet Singh · Data & Analytics Engineering · Waipapa Taumata Rau, University of Auckland

Ngā mihi
Thank you »



Thank you

Appendix

YOUR DATA FOUNDATION FOR TRUSTED AI AND ANALYTICS

MOVE data from 750+ sources

High-throughput, low-latency CDC with automatic schema and API change handling across your data sources

+ Connector SDK connect to any source with Python

MANAGE 30+ destinations

- Cloud data warehouses
- Databases
- Data lakes
- Streaming platforms

MANAGED DATA LAKE

Open table formats: Apache Iceberg™ & Delta Lake

Catalog & governance support: including Fivetran managed Apache Polaris™

TRANSFORM

- dbt Wizard
- Modular SQL transformations
- Intelligent orchestration
- Semantic layer
- Data lineage
- Data contracts

CONTEXT

- Parse files
- Search index
- Catalog datasets
- Ontology discovery

AGENTS SCHEMA

an open-source standard giving agents a standard place to look up context

ACTIVATE

200+ apps

Reverse ETL end-to-end pipelines

No code required

Agent Context MCP

Connect data and context to any AI solution

OUTCOMES

- Data-driven operations
- AI & Agents
- Real-time analytics

ANY COMPUTE

LAKE COMPUTE

PLATFORM FOUNDATION

BUILD & EXTEND
with SDKs, APIs, agent skills and dbt Wizard

SECURITY & PRIVACY
with full compliance, encryption, and multiple deployment offerings

RELIABLE & AUTOMATED
with 99.9% uptime across 7M+ daily syncs

Your data infrastructure needs to change

Fivetran

is the industry-leading data movement platform.

dbt Labs

is the industry standard for data transformation.

Together, we enabled the modern data stack:



Modern data stack

- Solved siloed data
- Set up the standards for data transformation
- Enabled always-on analytics and reporting
- Embraced by thousands of companies worldwide

What's changed?

- Data powers AI agents, automation, and real-time systems
- Workloads are more dynamic, high-volume, and diverse
- Infrastructure must evolve continuously, not incrementally

Trusted by global organizations

8k+ PB of data moved, monthly

3.8B models built, monthly

Maintained by **900+** engineers

Recognition

Forbes Cloud 100 2025

Snowflake Data Integration winner 2025

Google Cloud Data & Analytics winner 2026

8,000+ Fivetran customers

100,000+ data teams using dbt

1,500+ joint customers

AUTODESK, OpenAI, nab, LVMH, CEMEX, docusign, Raiffeisen Bank, Lufthansa, Pfizer, snowflake, CONDÉ NAST, GEICO, zoom, jetBlue, databricks, SIEMENS, Nintendo, salesforce, General Mills, Blackstone, COSTCO WHOLESALE, Tyson Foods, Hilton, AstraZeneca

Humans and agents are different kinds of workers

	Humans	Agents	Implications
Context	Bring it intuitively	Require context explicitly	**Trust** needs to be codified
Persistence	Operate intermittently	Operate constantly	100 agents = 1000x compute + token **cost**
Error handling	Use judgement to catch errors	Scale mistakes without hesitation	**Scale** of impact - good or bad - is amplified

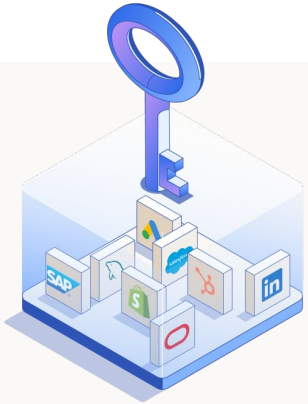
Humans compensate for bad context. Agents amplify it.

Agents need governed context and data - cost-efficiently.

Do you trust your data infrastructure to power agents?

Is data governed, accurate, and traceable?





Is it portable, and built on open standards?

Can it handle agent costs, at scale?



Fivetran and dbt are already integrated to deliver trusted data, today

