Streamhouse: a valós idejű, AI-ready vállalat technológiai alapja

Huszti Dániel
daniel.huszti1@ibm.com

Kern Bence
bence.kern@ibm.com

Promise

Personalized Health Care

Frictionless Banking & Financial Services

All Your Business, Always Current

Ultrafast Logistics

Personalized Shopping & Content

Virtual Assistants for Work & Personal Life

Reality

32%

of production workloads use agentic AI

DATA STREAMING REPORT

63%

of organizations do not have, or are unsure they have, the right data for AI

GARTNER

72%

report insufficient real-time data infrastructure blocking AI at scale

DATA STREAMING REPORT

95%

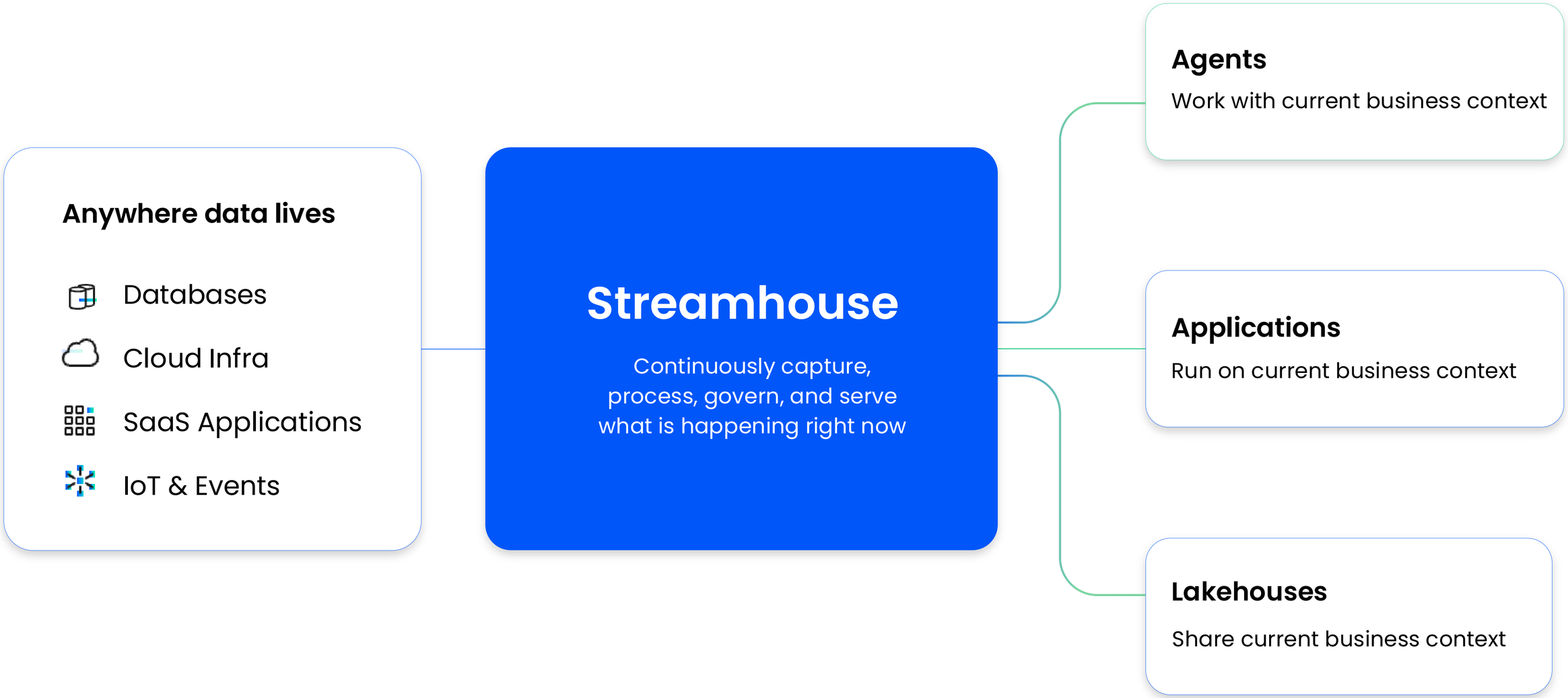
failure rate for enterprise AI solutions

MIT GENAI DIVIDE

Enterprise data architecture must evolve

Warehouse	→	Lakehouse	+	?
Internal analysis		*Internal* modeling		Run the business

The Streamhouse data architecture to power your agents, applications, analytics and current view of the business



Case study of a global FMCG industry leader

THE EVERYDAY PROBLEM

Everything today is in the stream. **That is not the same as an answer.**

A stream is a diary. It is very good at recording what happened, in order. It is very bad at handing you one line out of it.

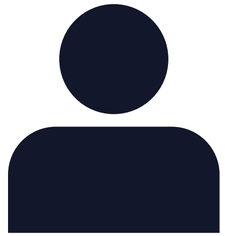
TODAY'S SHIPMENT LOG



read all of it

84,000 rows, every query

“Has shipment CH-88421 arrived yet?”



A planner, a dashboard, an AI agent.
All asking about one shipment.

Nobody wants the whole day. They want one line, and they want it now.

WHAT PEOPLE ACTUALLY ASK

Three questions. **They need three different things.**



“Has this one arrived?”

One shipment. One answer.

Hundreds of times a minute, by apps, screens and agents.

Point Lookup

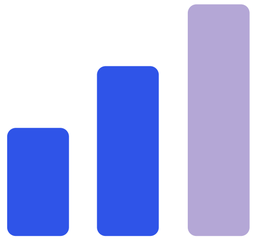


“What happened in the last two hours?”

A slice of the day, in order.

Every time a dashboard refreshes itself.

Range Scan



“How did the week go, including the last minute?”

Archive read + now, in one answer.

Analyzing trends by not missing out the latest data.

Union Read

Give shortest answer, while the business is running. Streamhouse is about all three.

WHY THIS WAS HARD

So everyone bought **a second system** to hold the answer

Not because they wanted to. Because the stream had no way to hand back one line.

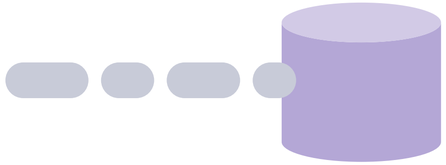
1



The stream knows everything

Every update from every truck, machine and app, in the order it happened.

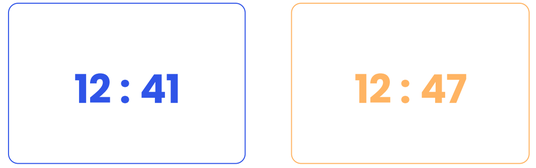
2



So we copy it out

Into a database, a cache or an analytics store. Something that can filter.

3



same question, two answers

Now there are two “truths”

Two systems to run, two bills, and two versions of the truth that drift apart.

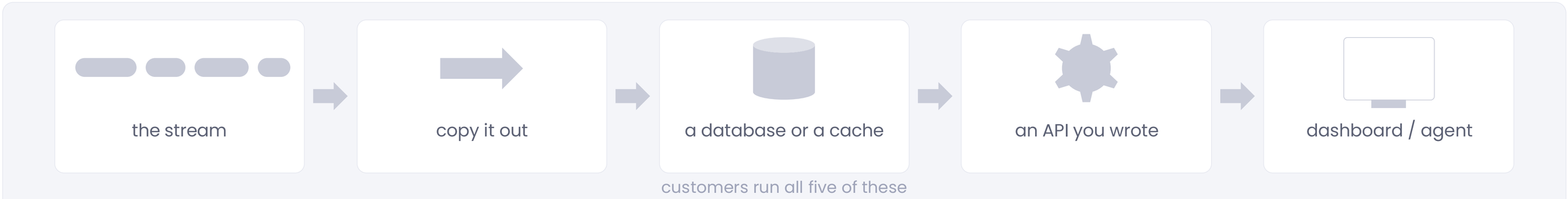
Every customer we know has built this. It was the only way to serve, and it is the complexity we are here to remove.

WHAT WE ARE ADDING

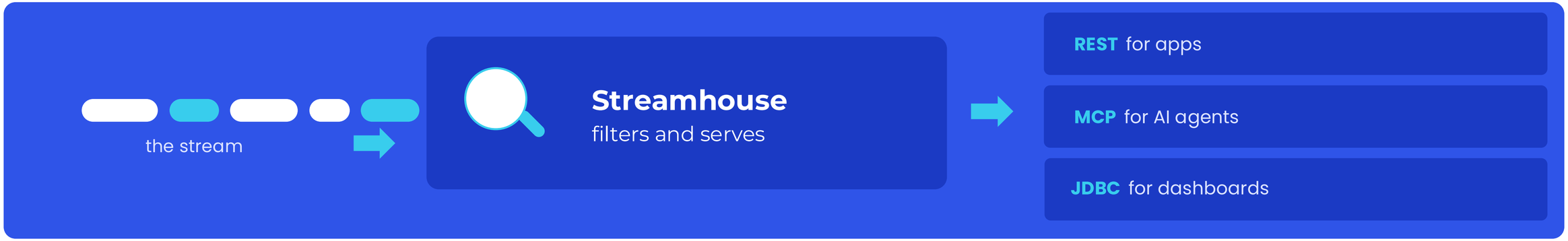
We are putting **the filter inside the stream and serve like a polyglot**

The stream stays the stream. It just learns how to answer a question about one thing.

TODAY



WITH THE STREAMHOUSE



LARGE EUROPEAN FMCG COMPANY - CARGO USE CASE

Why change? Why change NOW?

PRESENT STATE & CHALLENGES

Shipment visibility is slowed by different systems, forcing planners to search historical data for a current status.

METRICS (CHALLENGES)

- Revenue at risk from delayed deliveries
- Premium freight and shipment recovery costs
- Inventory carrying cost tied up in transit



CORE TRANSFORMATIONAL GOAL

Real-time shipment visibility replacing fragmented database lookups.

FUTURE VISION & DESIRED OUTCOME

Streamhouse maintains the latest shipment state and serves direct point lookups and historical range queries, reducing reliance on a separate serving database.

METRICS (SUCCESS)

- Revenue protected through earlier intervention
- Reduction in expedited freight and recovery costs
- Reduction in inventory days and working capital tied up in transit

“Where is this specific shipment?”





LARGE EUROPEAN FMCG COMPANY – CARGO USE CASE

“Where is this specific shipment?”





One shipment.
One answer.
Now.

WHAT THEY RUN TODAY

- a database holding the last ten updates per shipment
- a ETL pipeline keeping that database fed
- a small service in front of it so the agent can ask

WITH THE STREAMHOUSE

One lookup, straight from the platform.

Decoupling, no point to point anymore
Direct serve dashboards, no need for external DB
Point Lookups + Range Queries + Union Reads

Why change? Why change NOW?

Present State & Challenges

Current challenges include fragmented factory infrastructure, unreliable connectivity and limited access to factory insights for analytics.

Metrics (challenges):

- Cost of operating and maintaining factory data integrations per site
- Revenue or production value at risk from unplanned downtime
- Scrap, rework and quality-related costs

Future Vision & Desired Outcome

A resilient, real-time factory data backbone that lowers costs, improves production visibility and scales analytics across factories.

Metrics (success):

- Reduction in factory integration and platform operating costs
- Reduction in unplanned downtime and increase in OEE
- Reduction in scrap, rework and quality incidents

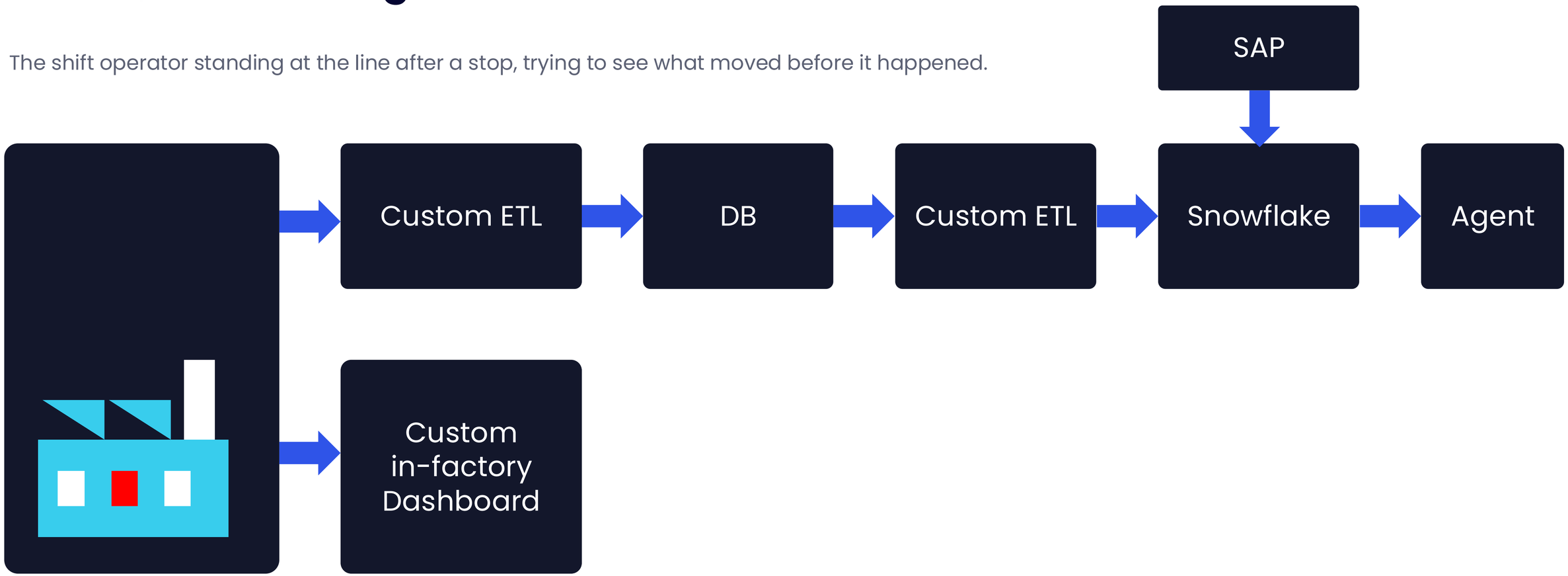


Real-Time Factory Connectivity & Operational Analytics Infrastructure

CONNECTED FACTORY

“Line 4, **what changed in the last two hours?**”

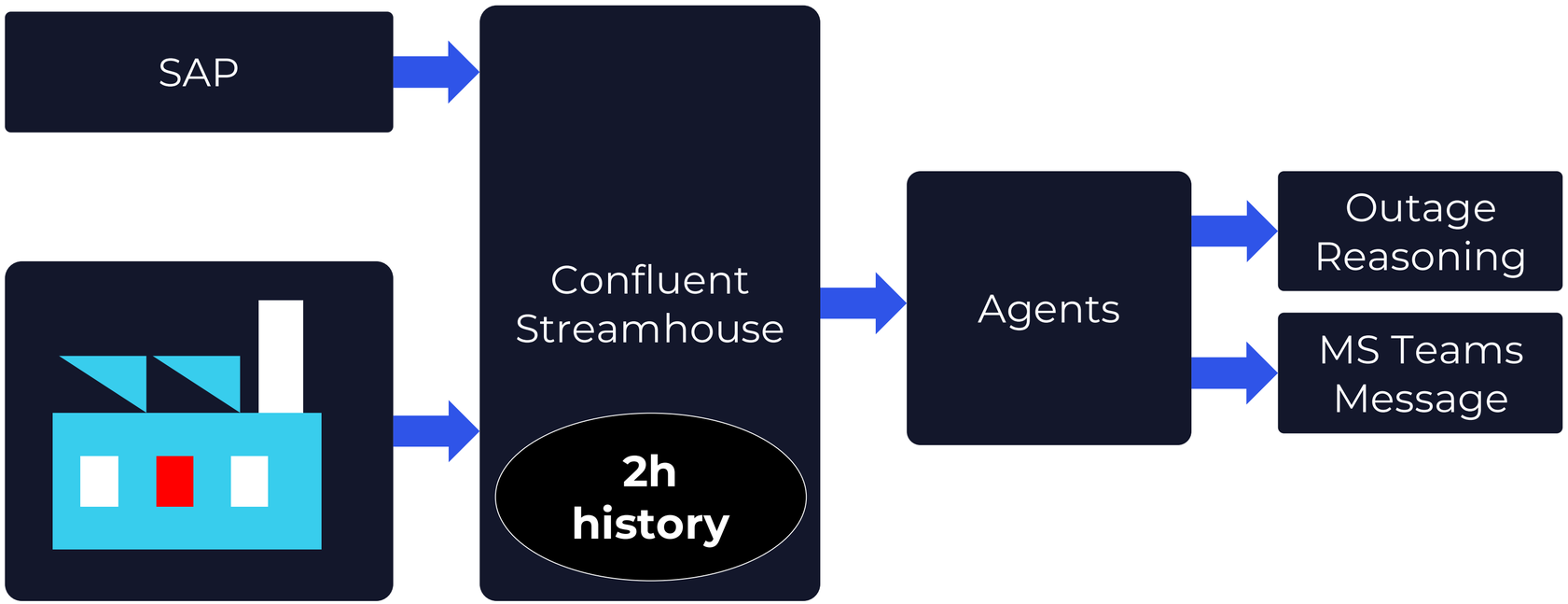
The shift operator standing at the line after a stop, trying to see what moved before it happened.



CONNECTED FACTORY

“Line 4, **what changed in the last two hours?**”

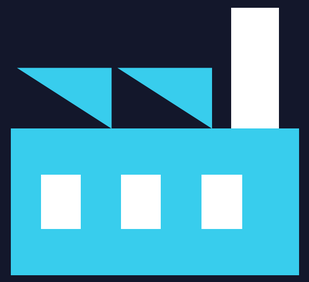
The shift operator standing at the line after a stop, trying to see what moved before it happened.



CONNECTED FACTORY

“Line 4, **what changed in the last two hours?**”

The shift operator standing at the line after a stop, trying to see what moved before it happened.



A slice of the day, in order.

WHAT THEY RUN TODAY

- custom apps for dashboarding factory info for operators
- custom scripts for extracting factory info to cloud
- database on Azure with custom apps to aggregate data from various factories
- custom apps to distribute results to various downstream systems

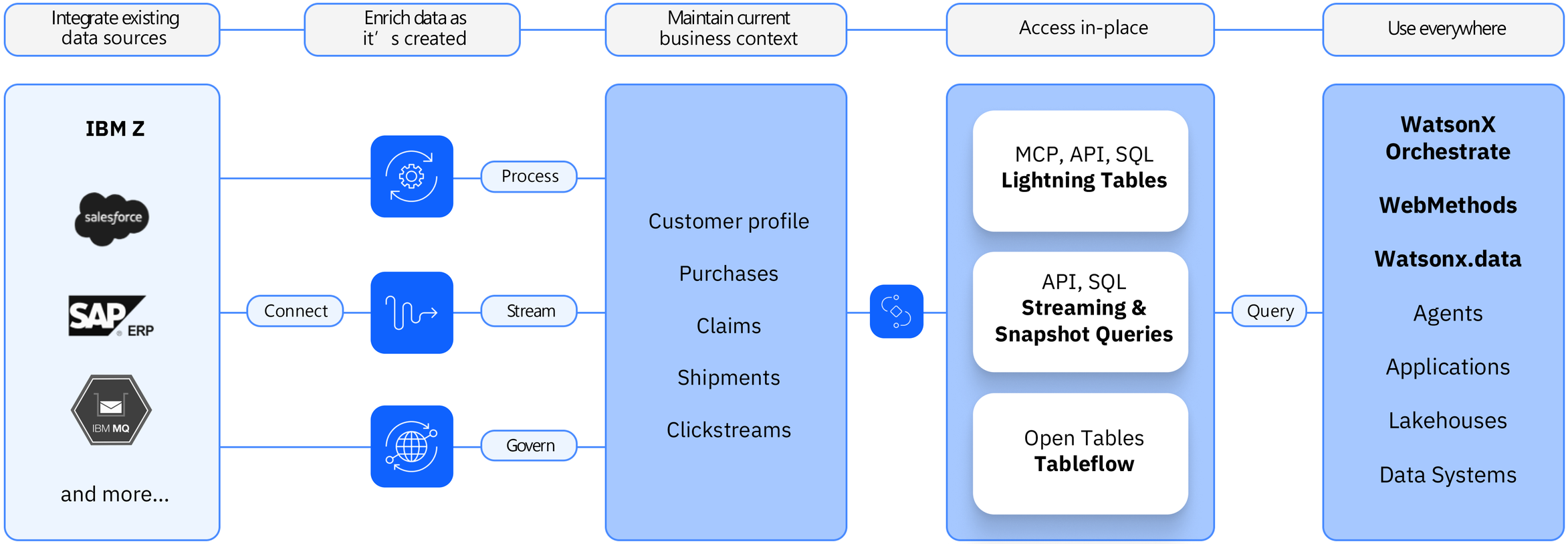
WITH THE STREAMHOUSE



The last two hours served directly to the screen.

No need of custom ETL
No analytical engine dependency for agents
Serve specific timeranges only

Confluent **powers agents, applications, analytics, and lakehouses** with all your business, now



From reactive assistants to proactive, event-driven agents

1 The evolution of customer-facing AI maturity

01 Rules-Based Chatbot

Scripted responses

Key differentiator
Logic

02 GenAI Knowledge Assistant

Knowledge-based answers

Key differentiator
Knowledge

03 Context-Aware Assistant

Personalized guidance

Key differentiator
Context

04 Autonomous Agent

Executes tasks

Key differentiator
Action

05 Proactive Agent

Responds to events

Key differentiator
Signals

2 How event-driven agents work

Business event
A meaningful change occurs

→

Context assembled
Live data provides current state

→

Agent triggered
Right agent or workflow responds

→

Governed action
Approved actions execute, with human oversight

→

Continuous loop
Outcomes generate new events

Live business events trigger intelligent, governed action — enabling agents to continuously respond, act, and adapt.

Maturity model source: Gartner, 839386 (5 Levels of Customer-Facing AI Maturity), adapted.

Largest enterprises and newest digital natives have already started implementing Streamhouse

BMW GROUP

Protects vehicle production by **predicting equipment failures** and **orchestrating autonomous parts delivery** across 30+ plants.

SIEMENS Healthineers

Improves equipment uptime and factory quality by **detecting MRI/CT anomalies in real-time** across 8M+ daily events.

together.ai

Measures **real-time usage billing** across their inference stack for 450K+ developers, and **observability pipelines**.

indeed

Powers **booking, payments,** and **search** for 2.5K+ microservices, where a 1-second delay measurably reduces bookings.

Citizens Bank®

Speeds lending by 40% and cuts fraud false positives 15% through real-time **alerts**, **notifications**, and **customer insights.**

Rocket Mortgage

Keeps **mortgage origination**, **lead management**, **servicing**, and **trading decisions** moving across their business.

Data Product and Data Governance

Data product

Reusable, self-contained package

- **Support** diverse business use cases
- **Combines** data, metadata, semantics and templates
- **Components** such as Datasets, dashboards, reports, ML models, pre-built queries, data pipelines

Enable everyone

- Discoverable, interoperable, actionable
- **Everyone** can extract meaningful value from data trapped within the enterprise
- Business users, data analysts, data scientists, data stewards, engineers

Product-thinking approach

- Product development principles
- **Understanding** user needs, high-value features, iterating by feedback
- **Data as a product** to solve specific problem



Key characteristics of a data product



Discoverable

Easily discover and find the right data product for the use case.



Understandable

Include clear metadata, be structured to the specific business domains, interpret and apply information effectively.



Interoperable

Integrate seamlessly, deliver consistent insights across.



Shareable

Embrace new digital tools and methods without sacrificing human empathy, creativity and sense of belonging.



Secure

Have access controls and security, only authorized users by maintaining compliance.



Reusable

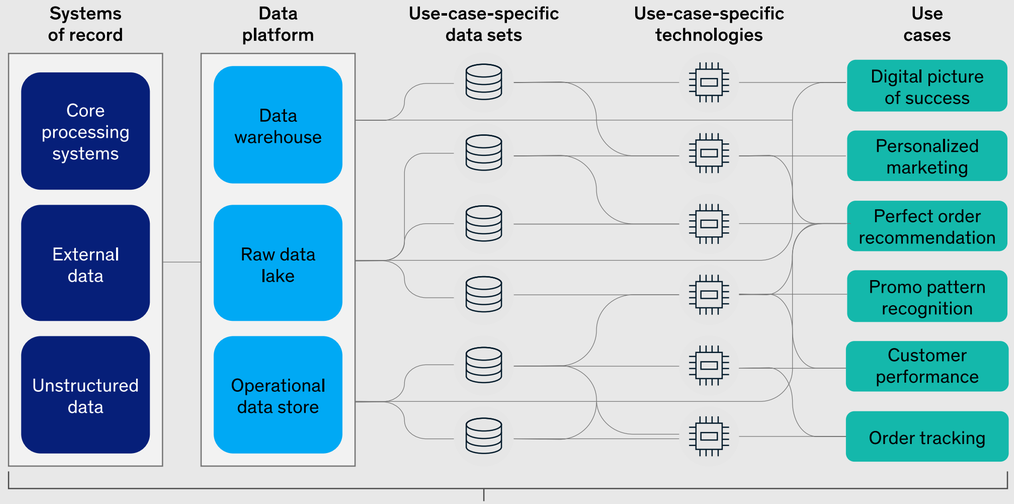
Built from modular components to be repurposed, reducing redundancy,

Use cases

A CPG company set up data products to detangle and simplify data use and unlock significant efficiencies in scaling.

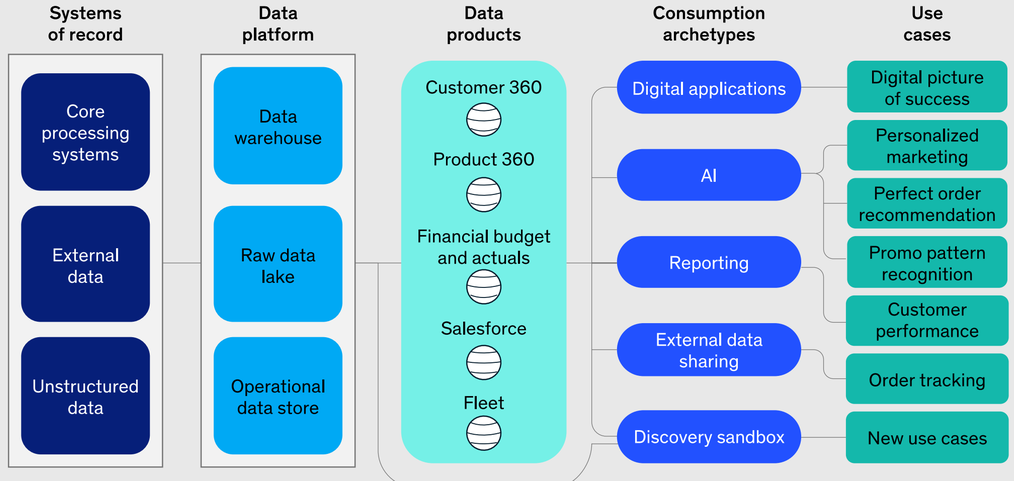
Illustrative CPG example

Before setting up data products



Data for each domain, such as the customer, is inefficiently reworked for every use case.

After setting up data products



Banking sector

- Single customer data product
- Powers 60 diverse use cases
- Multiple channels from real-time credit risk scoring to AI chatbots
- $60 million additional annual revenue

Consumer packaged goods firm

- Data products to streamline data usage, improving efficiency and scalability
- Deployment over 50 cross-functional teams to implement data-driven solutions

Introducing

watsonx.data intelligence

Feel confident in the quality of your data and the insights it generates with IBM's data intelligence solution for structured and unstructured* data

Data sharing

Central marketplace for data product distribution and consumption

Data lineage

Track and understand the journey of data from creation to consumption

Data quality

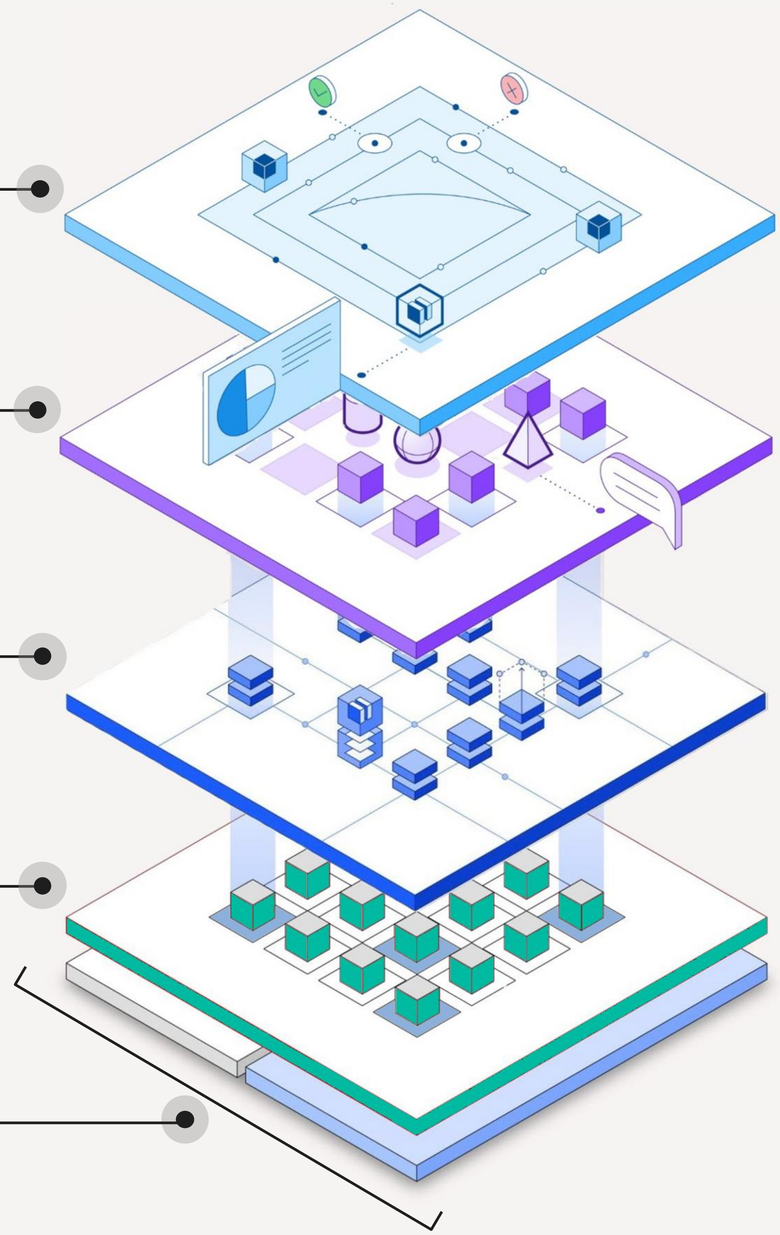
Automatically assess and monitor data quality and remediate quality issues

Data governance

Central enterprise metadata catalog that activates data for AI and analytics

Hybrid cloud platform

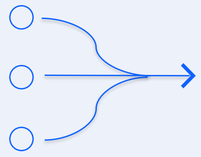
Run anywhere



* Unstructured data capabilities available only in IBM Cloud Toronto Data Center

Get end-to-end **data governance** of your entire enterprise data estate

1 Data Estate



Real-Time Data

Confluent Kafka
& Flink



Data At-Rest

Lakehouse, data
warehouses, etc.

2 Unified Data Governance and Lineage

Watsonx.data intelligence

- Enterprise-wide visibility
- Spanning data at-rest & real time
- Drives better business decisions

3 Trusted, Governed Data Estate
for real-time data + data at-rest



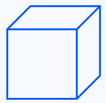
Lineage & auditability
End-to-end traceability



Compliance & security
Regulatory Controls and Data Protection

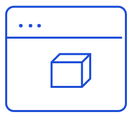


Business semantics
Standardized definitions and metrics



And More....

Business Outcomes



Compliance reporting



Impact analysis

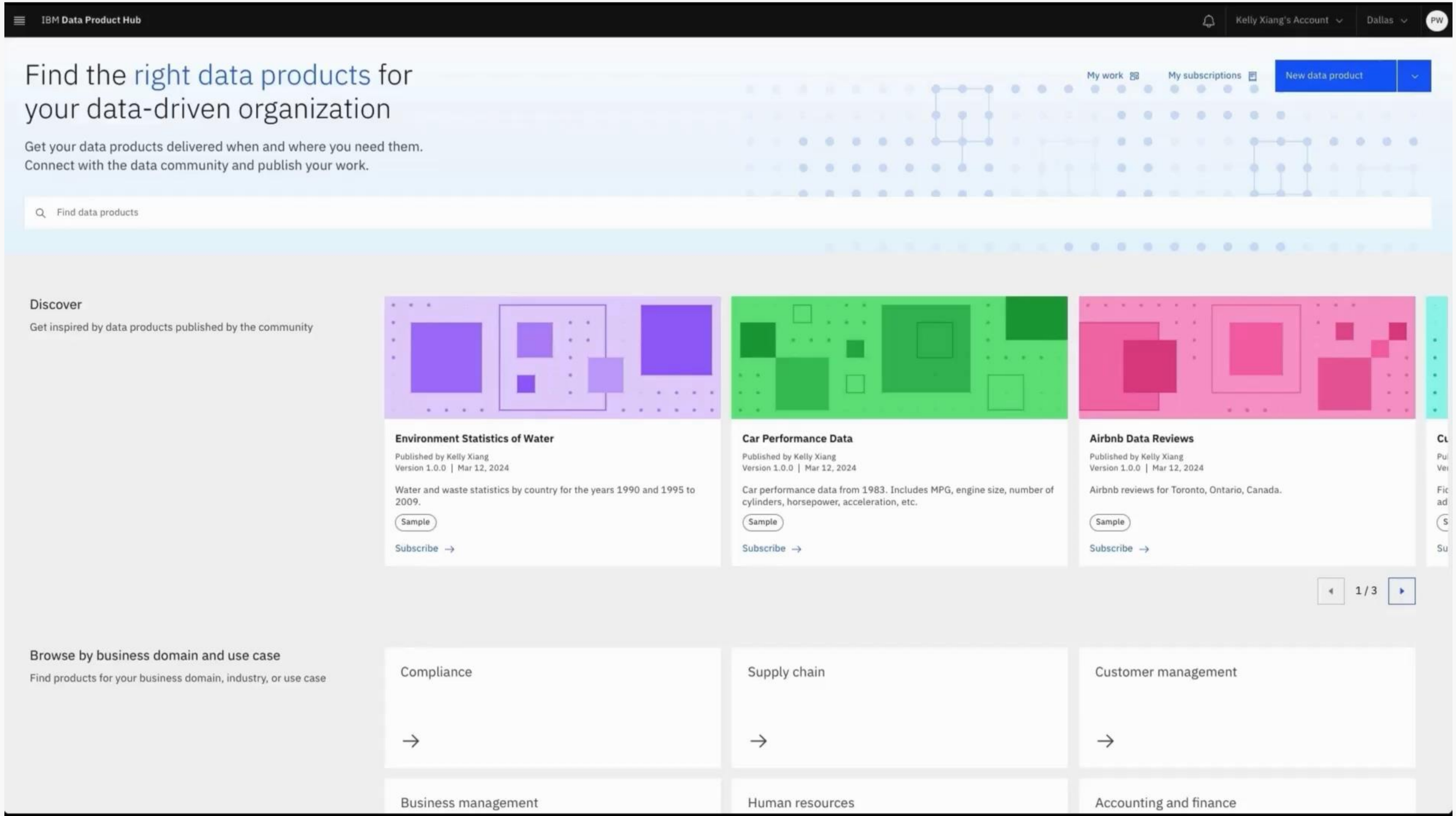


Root cause analysis



Historical Analytics

Data Product Hub overview



Data Product Hub for producers

Alleviating data silos
and ensuring data quality

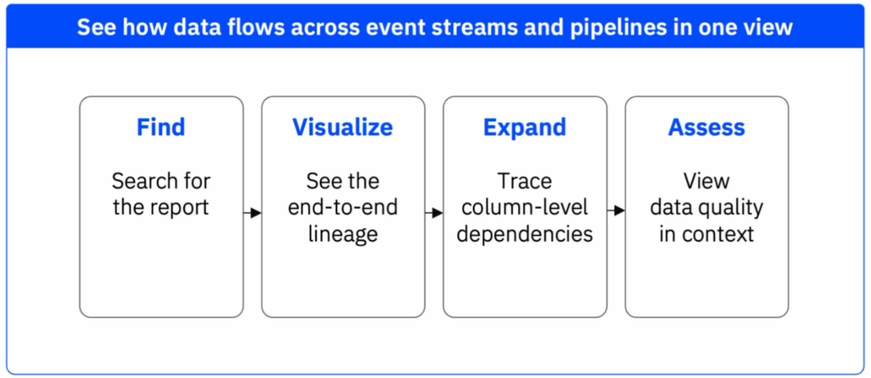
Data Product Hub for consumers

Discovering and accessing data
quickly

Data lineage for Confluent

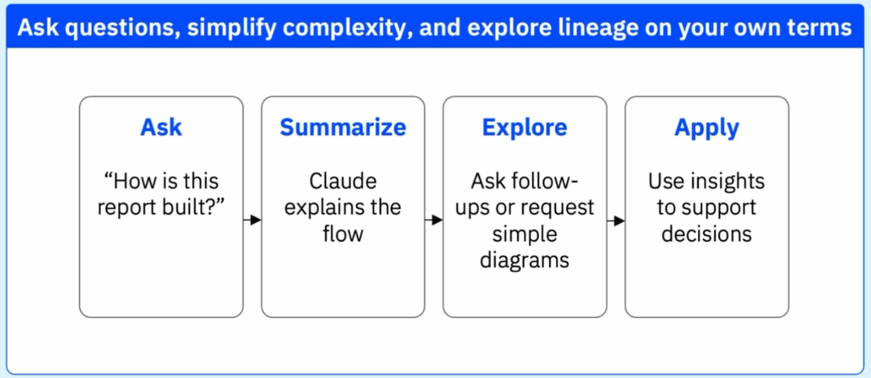
In this demo

Part 1: Trace a Report Across Event-Driven and Pipeline-Based Systems



Complete visibility into cross-technology data flows and dependencies

Part 2: Navigate Lineage Without the Graph



Conversational access to complex lineage and metadata

How to get started

1

Pick a use case

Start with one high-value, real-time business scenario.

2

Build real-time flows

Stream events from source systems with Confluent.

3

Shape data products

Package streams as discoverable, reusable products.

4

Govern and scale

Add lineage and ownership, then extend to more teams.

IBM watsonx.data intelligence key advantages

- Enterprise-grade security, trust and governance
- Hybrid cloud flexibility: run where your data lives
- AI and data portfolio backed by global expertise

Confluent key advantages

- Fully managed Apache Kafka and Flink
- Real-time data in motion across all your systems
- Pre-built connectors and built-in stream governance



Listen on Apple Podcasts

Listen on Spotify Podcasts

