



# Building Real-time Next Generation AI Applications with Confluent



Ryan Voon  
Solutions Engineer, Confluent  
[rvoon@confluent.io](mailto:rvoon@confluent.io)

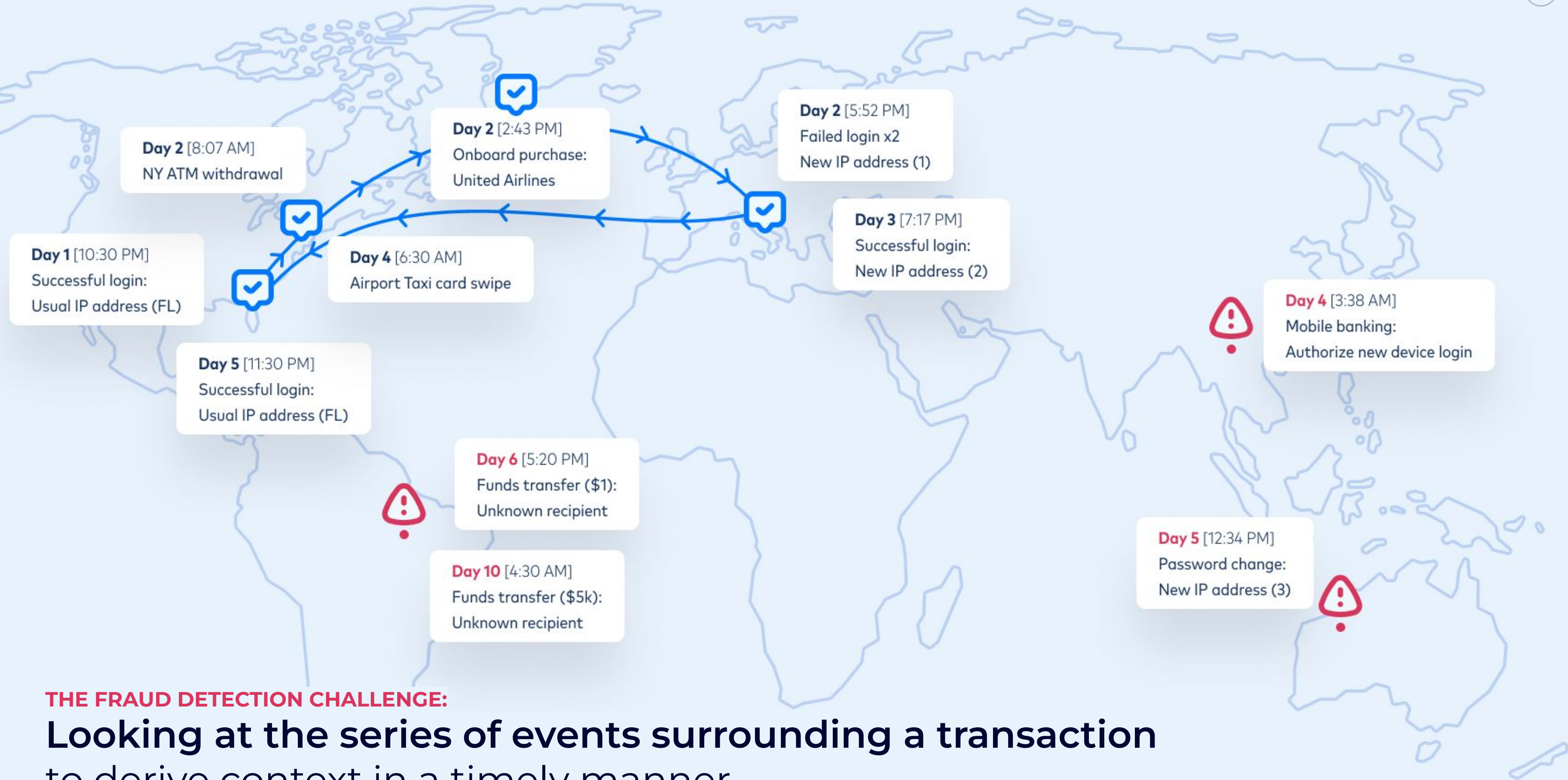


“Our latest research estimates that generative AI could **add the equivalent of \$2.6 trillion to \$4.4 trillion annually** across the 63 use cases we analyzed.”

Source: [Economic Potential of Generative AI](#), McKinsey







**THE FRAUD DETECTION CHALLENGE:**

**Looking at the series of events surrounding a transaction**  
to derive context in a timely manner

Without context,  
trustworthiness  
or real-time data  
applicability,  
LLMs can't drive  
meaningful value

What is the source of  
this information? Is this  
trustworthy? Is it fresh  
and accurate?

How do you securely augment  
customer data with real-time  
data and process them on the fly  
to provide meaningful insights?

What is the status of  
my flight **to New York**?

It is currently delayed by 2 hours and  
expected to depart at 5 pm GMT.

Is there another flight available  
to the **same city** that will depart  
and arrive sooner? What are the  
seating options and cost?

The next available flight to New York  
**with United departs later but will  
arrive faster than your current** flight.

**The only available seats in this flight  
are first class window seats and  
costs \$1,500.**

Can your GenAI  
assistant remember  
data from an earlier  
conversation?



## The reality of today's data integration strategy

A giant mess of monolithic point-to-point connections  
with data fidelity and governance challenges



# Your legacy data infrastructure is unfit for the real-time AI revolution

## Missing Real-time Contextual Data

Batch-based data processing creates stale data and integration challenges, making it difficult to build a real-time, trustworthy knowledge base

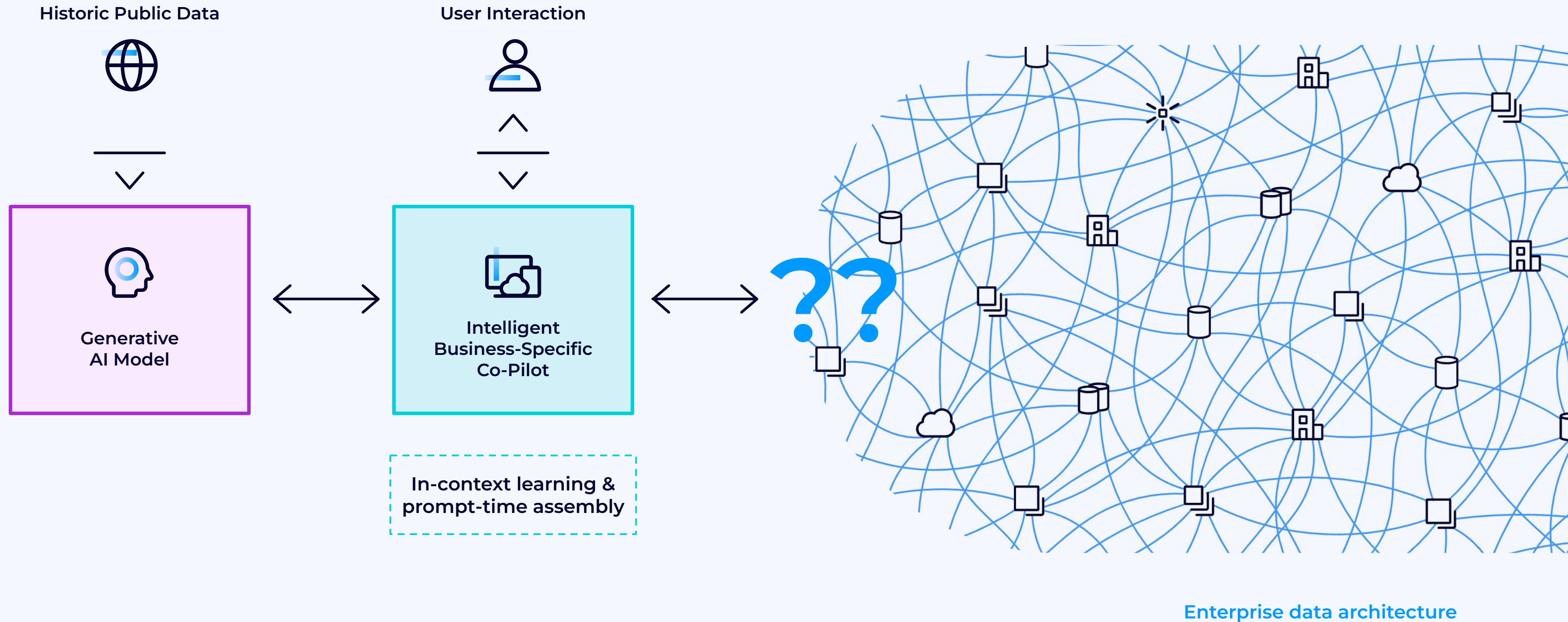
## Innovation Friction and Lock-in

A tightly coupled architecture increases the risk of lock in and slows the pace of innovation when new AI apps and models come to market

## Weak Security and Governance

As AI data needs emerge and number of point-to-point integrations increase, data access, lineage, and security challenges escalate

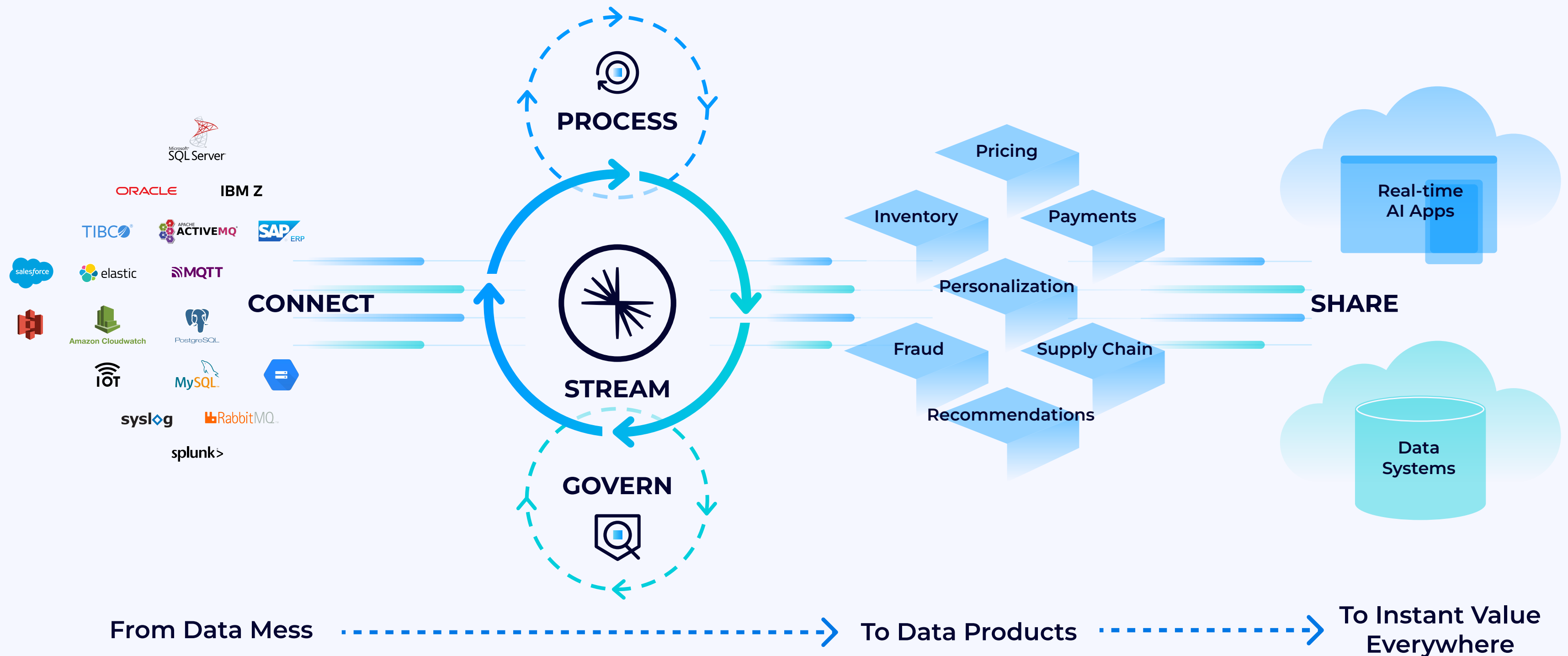
# Traditional enterprise data architecture is a GenAI innovation bottleneck



# A new paradigm to set your data in motion:

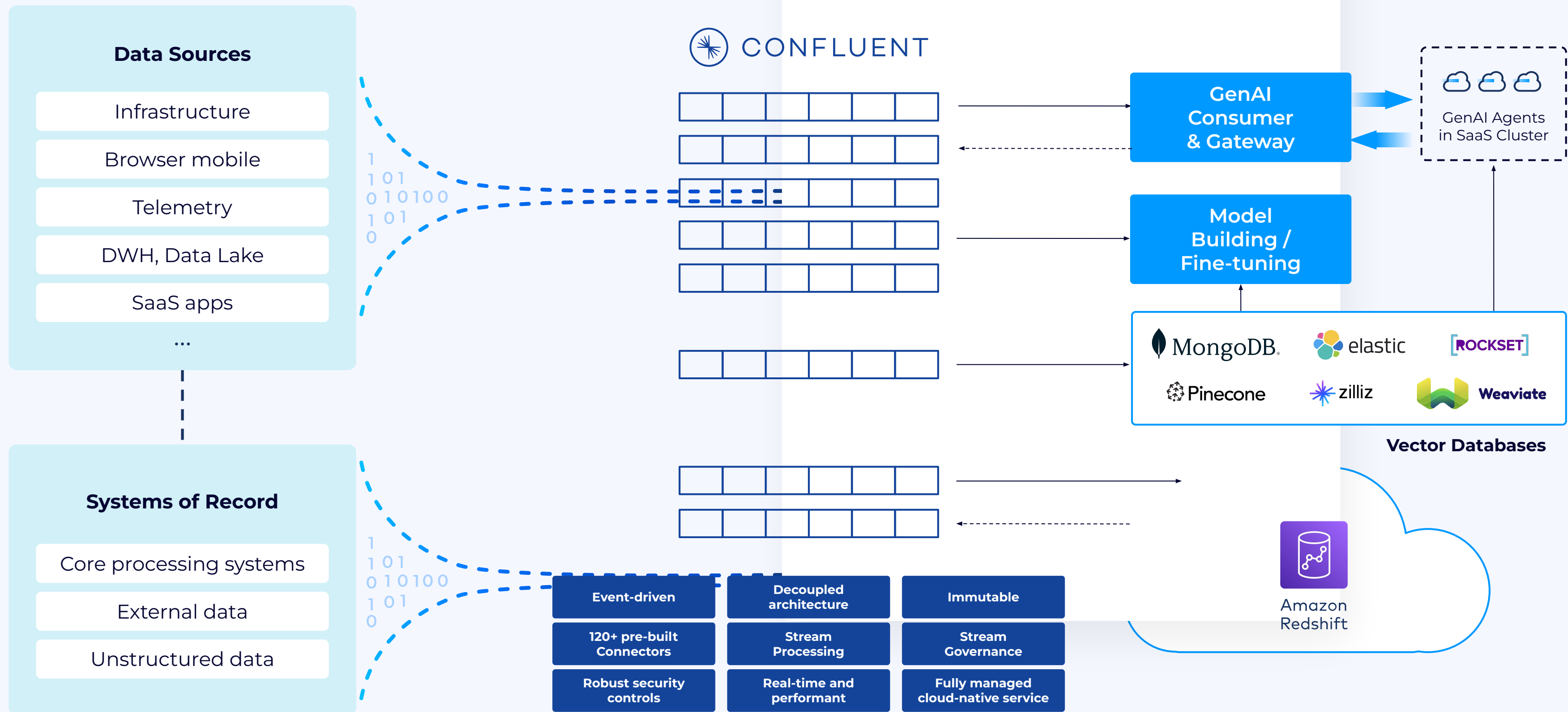
## Confluent Data Streaming Platform

Build a real-time, contextual and trustworthy knowledge base for your AI applications





# How



Create a real-time knowledge base

Bring real-time context at query time

Build governed, secured, and trusted AI

Experiment, scale and innovate faster



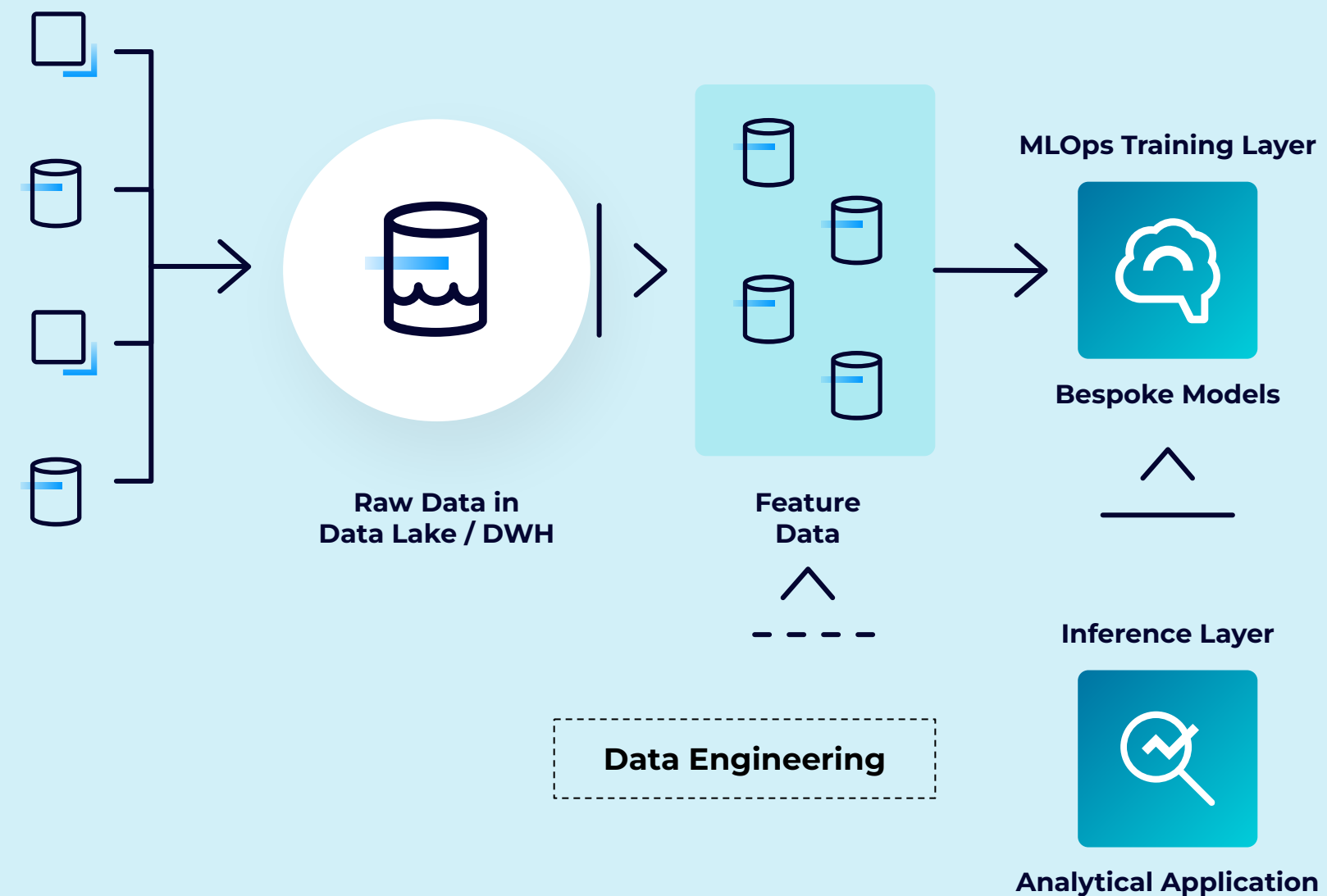
# Confluent for Real-time GenAI Applications Deep Dive



# What's different about Generative AI

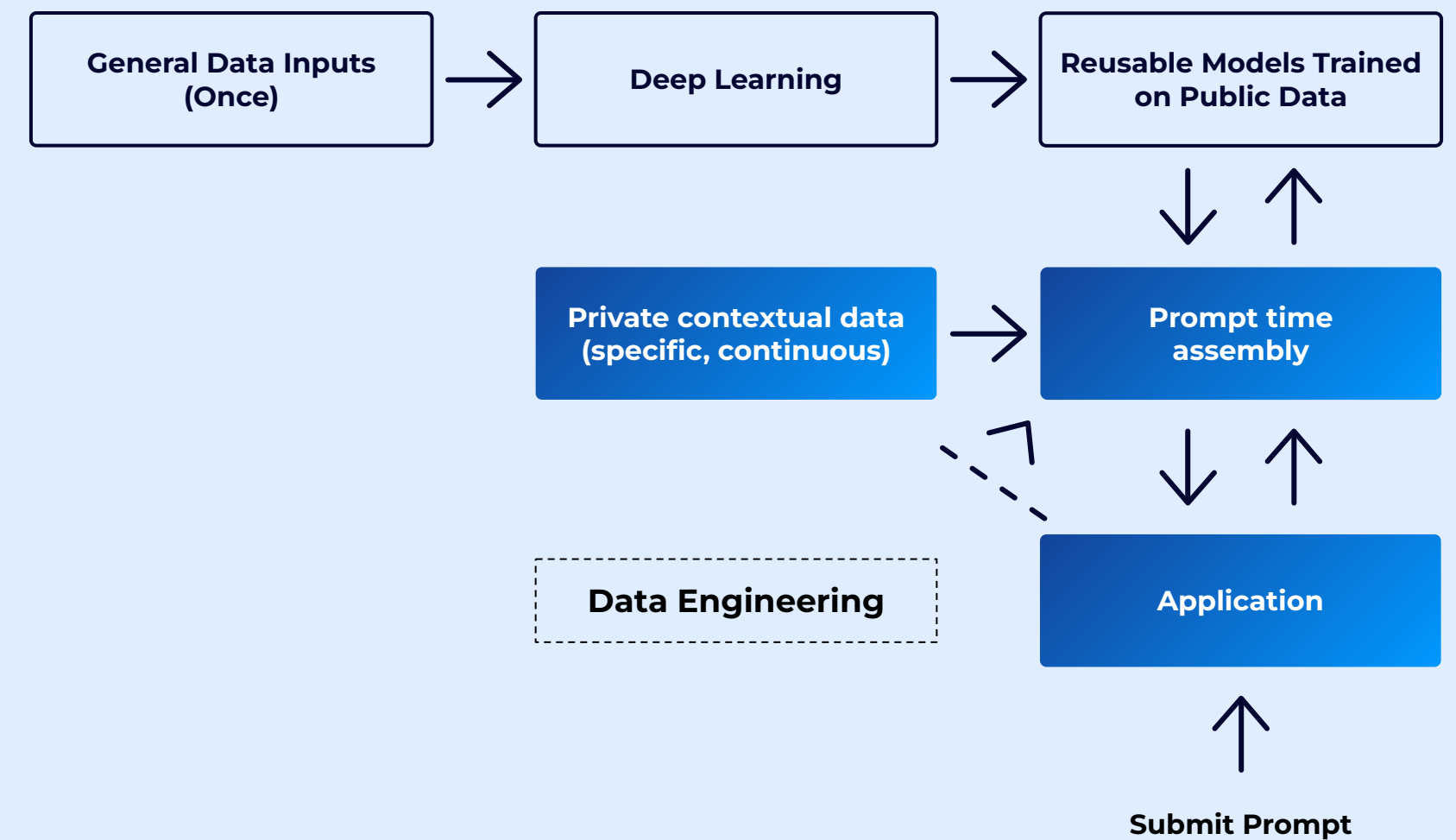
## Predictive Artificial Intelligence

Makes predictions. Purpose-built models. Offline training.



## Generative Artificial Intelligence

Creates content. Reusable models. In-context learning.



# LLM-enabled applications have four steps



## Data Augmentation



Prepare data for a real-time knowledge base and contextualization in LLM queries

## Inference



Programmatically connecting relevant information with each prompt

## Workflows



Parsing natural language to synthesize necessary information and apply contextual reasoning on-the-fly

## Post-Processing

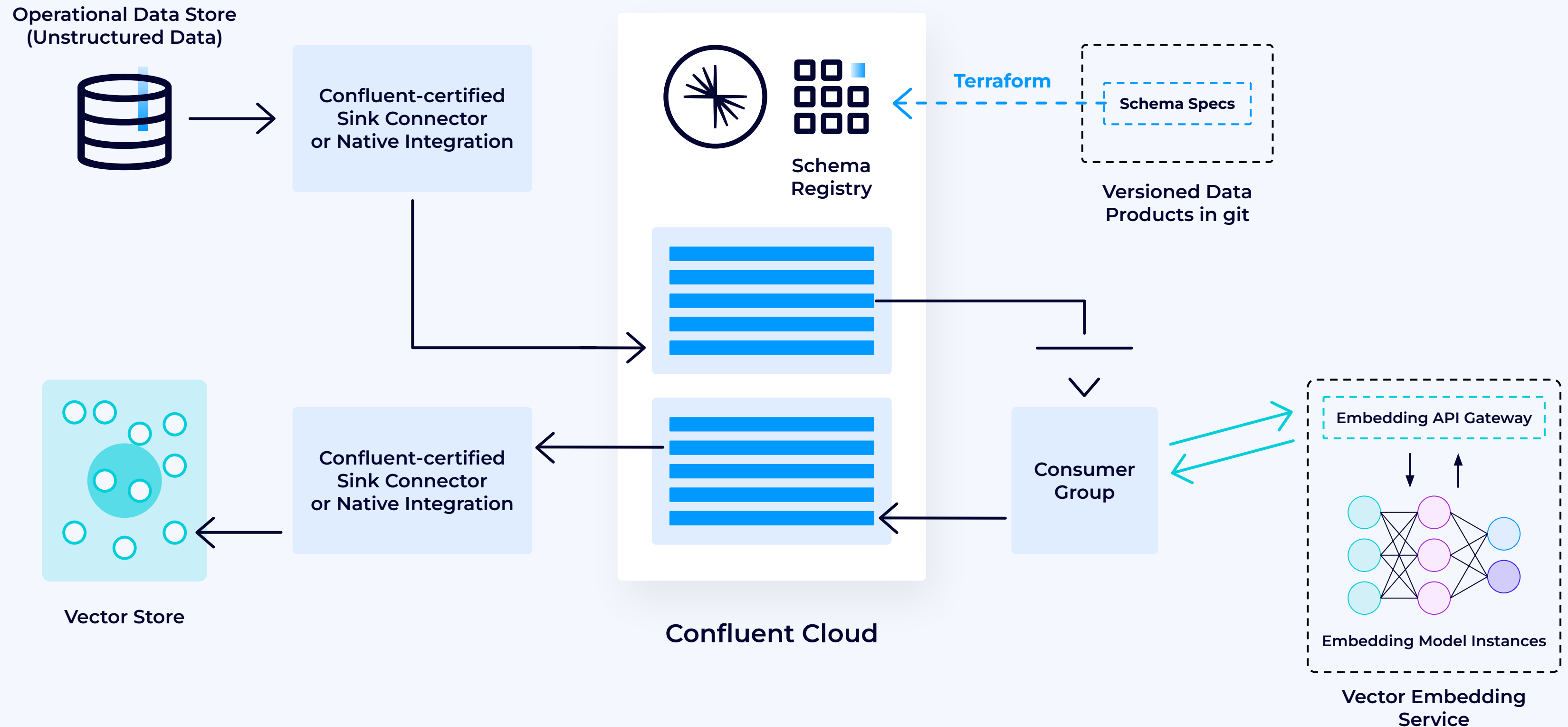


Validates outputs and enforces business logic to detect hallucinations and ensure trustworthy responses



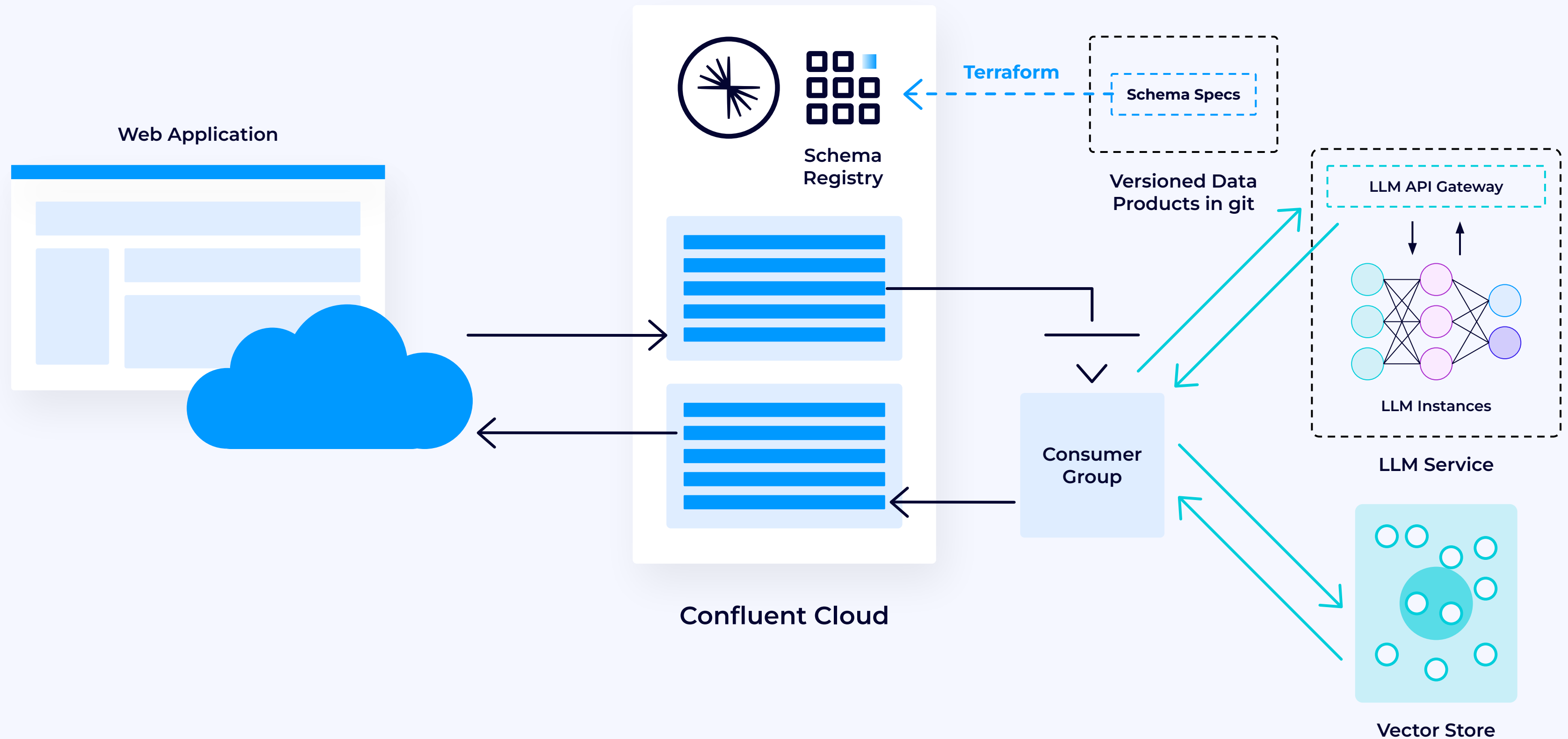
# How Confluent Works: Data Augmentation

Prepare data for a real-time knowledge base and contextualization in LLM queries



# How Confluent Works: Inference

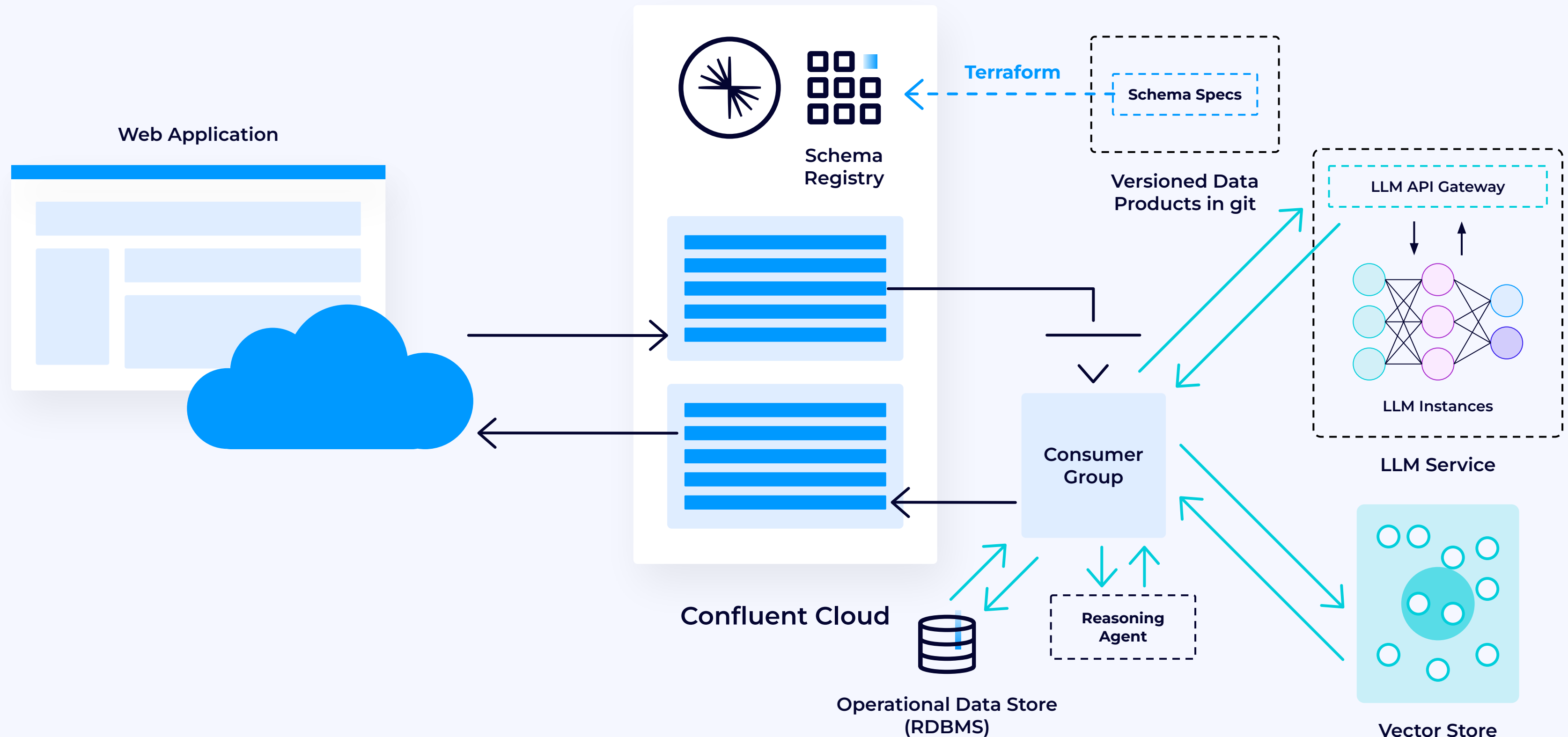
Programmatically connecting relevant information with each prompt





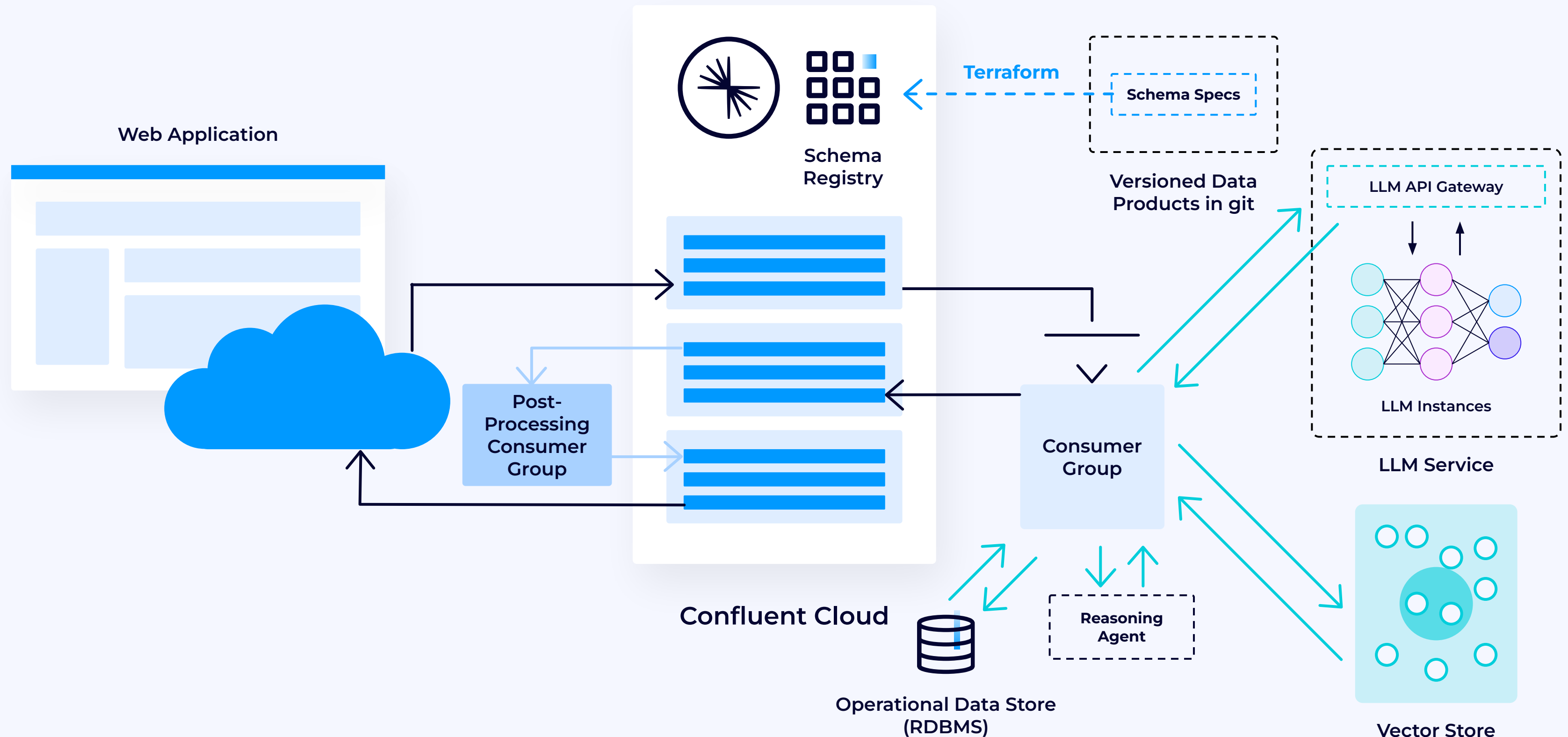
# How Confluent Works: **Workflows**

Parsing natural language to synthesize necessary information and apply contextual reasoning



# How Confluent Works: Post-Processing

Enforce Business Logic and Compliance Requirements with LLM Outputs







# Popular Real-time GenAI Use Cases

Extract enterprise value from all your data in real-time to power GenAI experiences for your customers



Semantic  
Search



Customer  
Service



Content Discovery /  
Recommendation

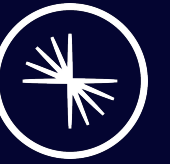


Agents (Task  
Automation)



# Getting Started

# Build the right data architecture for future



FULLY MANAGED CLOUD SERVICE



## Confluent Cloud

Apache Kafka Re-Engineered for the Cloud

Available on the leading public clouds



SELF-MANAGED SOFTWARE



## Confluent Platform

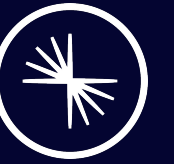
The Enterprise Distribution of Apache Kafka

Deploy on-premises or in your private cloud





# Leverage the industry's only hybrid & multi-cloud solution for Apache Kafka



## Private Cloud

Deploy on premises with  
**Confluent Platform**

## Hybrid Cloud

Build a persistent bridge  
from datacenter to cloud  
with **Cluster Linking**

## Public/Multi-Cloud

Leverage a fully managed  
service with **Confluent  
Cloud**

# How to get started

Subscribe to Confluent Cloud in and start with a free \$400 (to be used within 30 days)

[confluent.io/confluent-cloud/tryfree/](https://confluent.io/confluent-cloud/tryfree/)



# Confluent

