

## INDUSTRY TALK SERIES

# Building Effective Data Teams: Streaming, ETL, and LLM-Powered Reporting

VINCENT CHU  
APRIL 30, 2025



# Hello!

**Vincent Chu**  
SFU Computing Science '06

EA / Apple / Visier  
Head of Engineering for Later.com  
Head of Engineering / Product at AgencyAnalytics



# 4 topics today

## Building teams

Do you need a large team to be effective?

Can Gen AI solve everything?

## ETL

Do you need ETL, can't you just dump everything into a bucket for LLM?

## Streaming

What if you want to do analysis really fast?

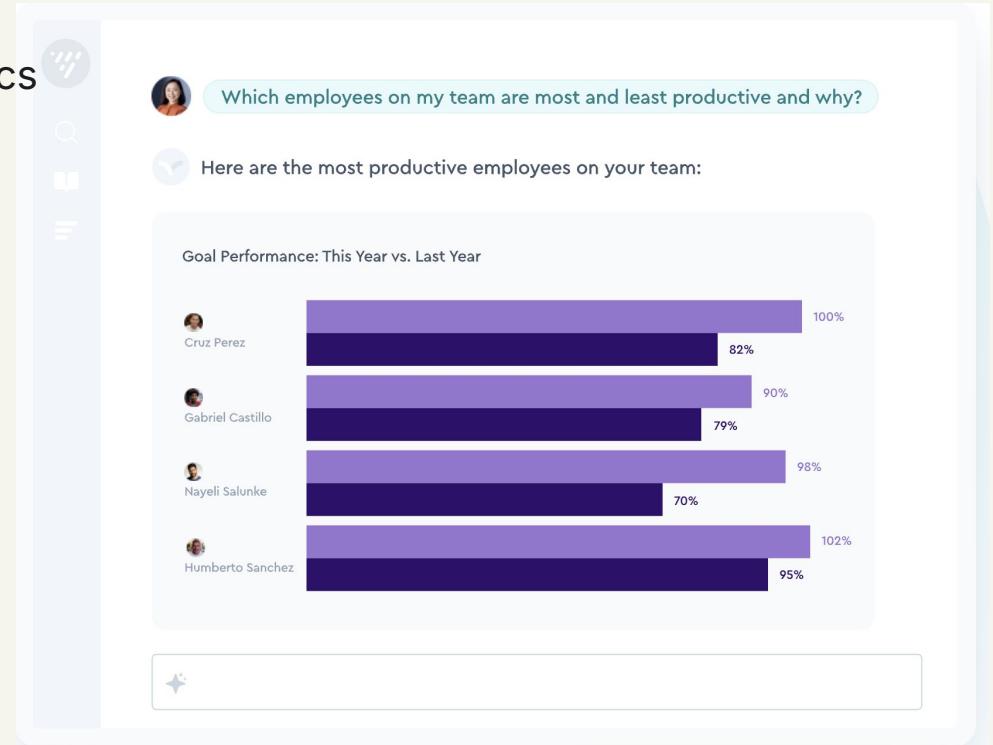
## Reporting with AI

Is Gen AI going to hallucinate?

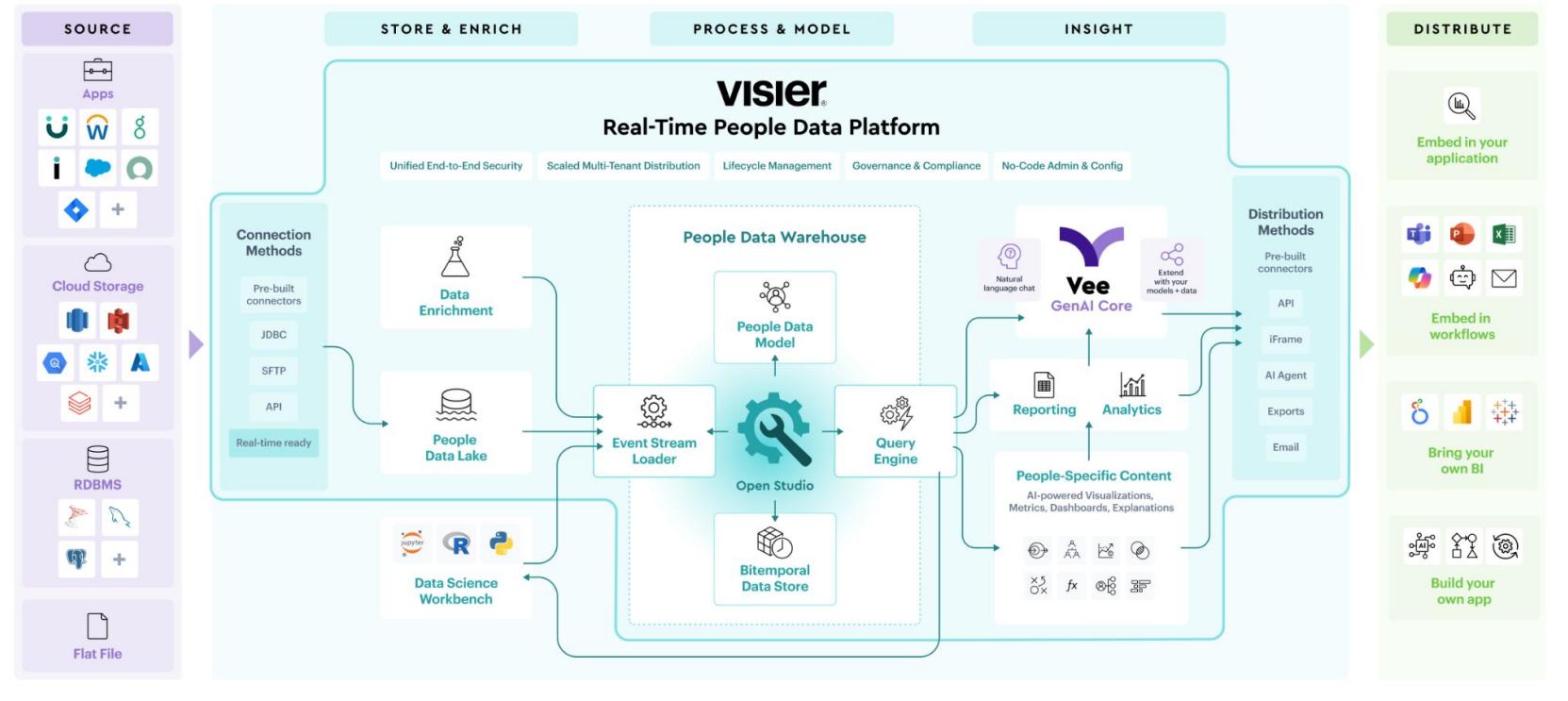
# What does Visier do?

- Founded in 2010 by pioneers in Analytics
- Deep technology  
(not just mashing things together)
- Pioneer HR analytics.  
Now known as “People Analytics”

What are People Analytics?



# Modern data and insight architecture



# 4 topics today

## Building teams

Do you need a large team to be effective?

Can Gen AI solve everything?

## ETL

Do you need ETL, can't you just dump everything into a bucket for LLM?

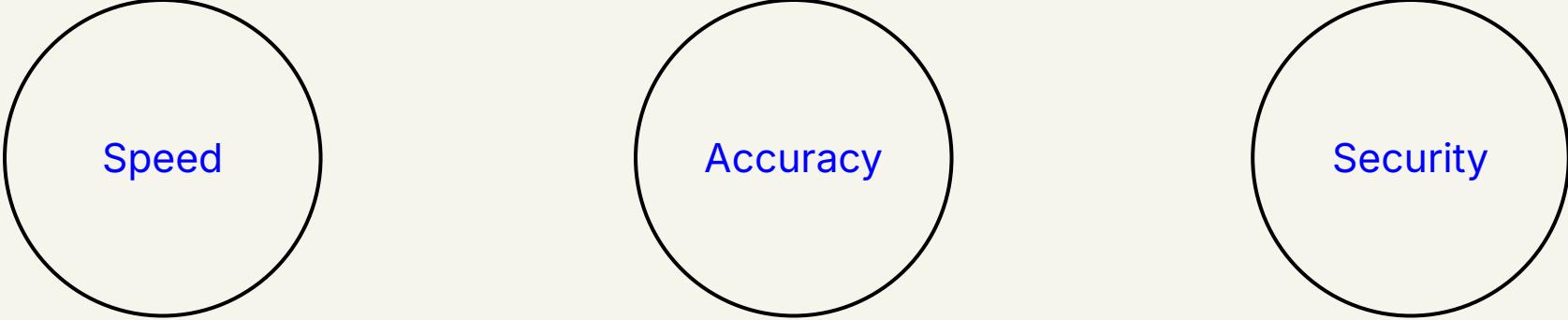
## Streaming

What if you want to do analysis really fast?

## Reporting with AI

Is Gen AI going to hallucinate?

# “AI” problem is still a data problem

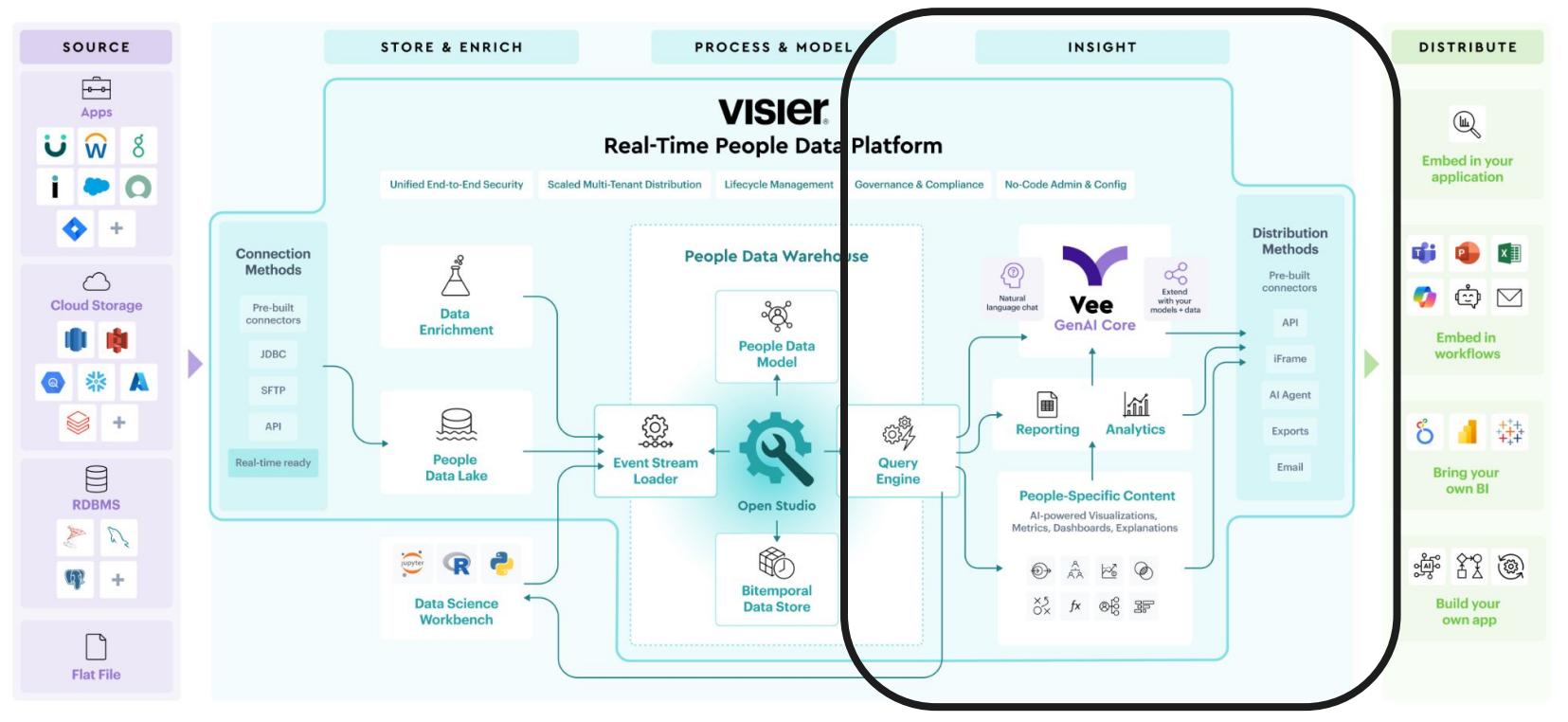


Speed

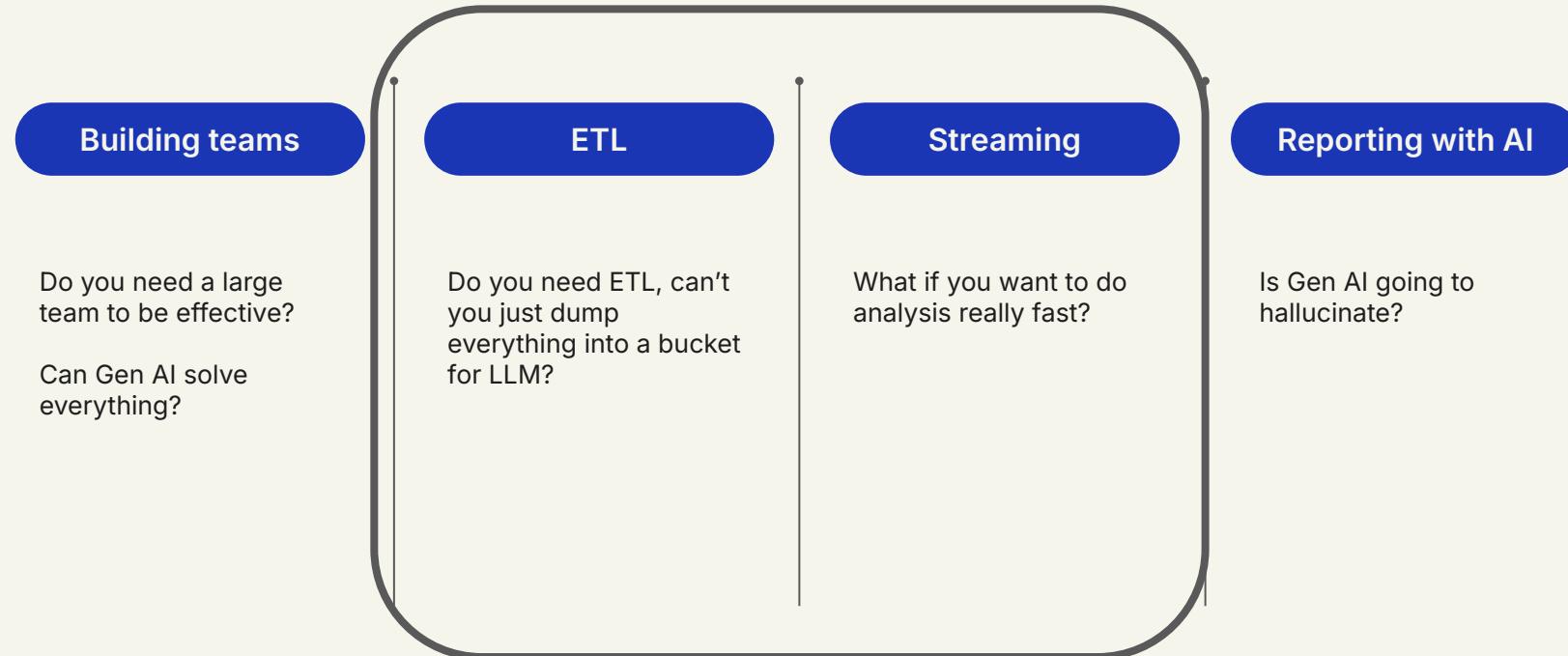
Accuracy

Security

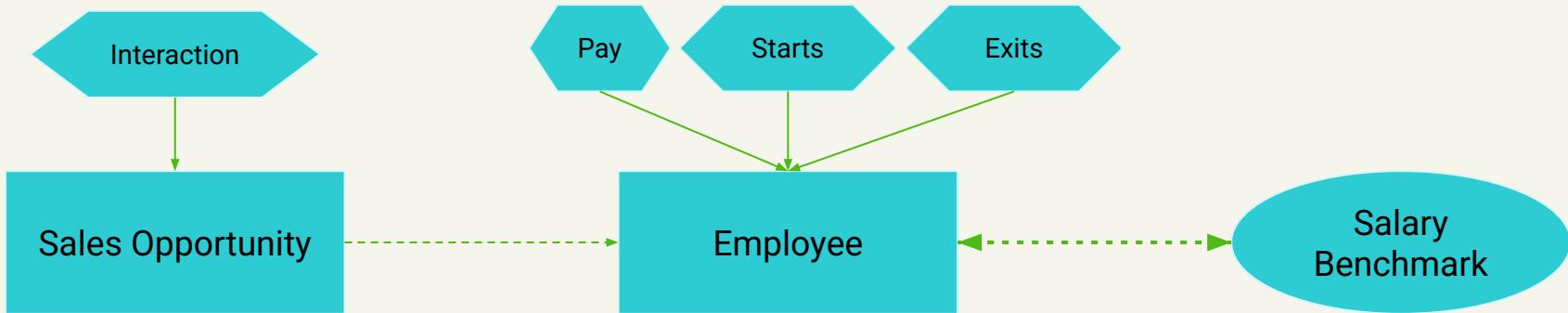
# “AI” complements structured queries



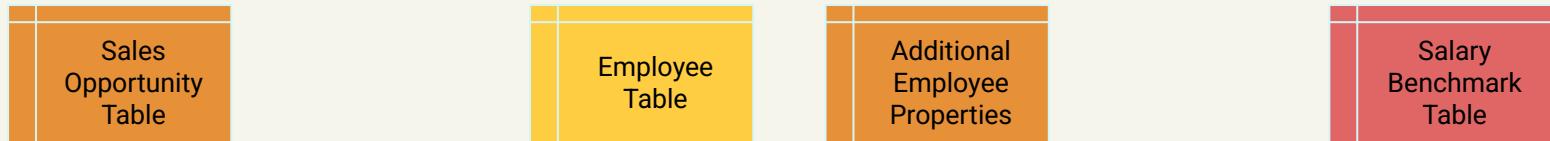
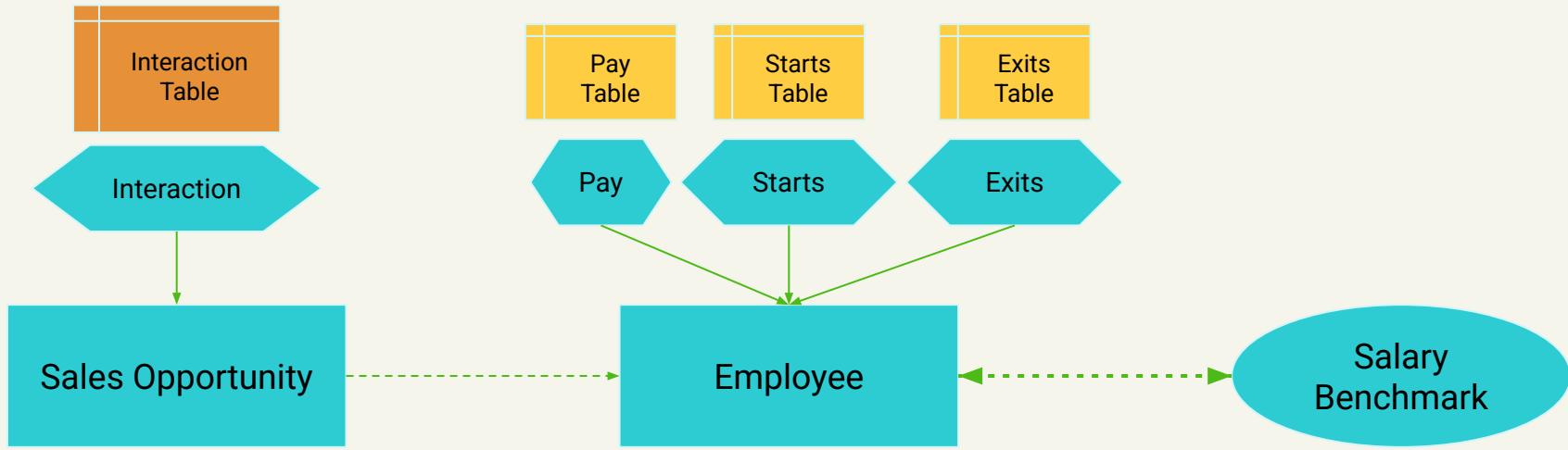
# So... that's it? Are we done?



# Data Modelling



# Data Modelling



# Data Modelling

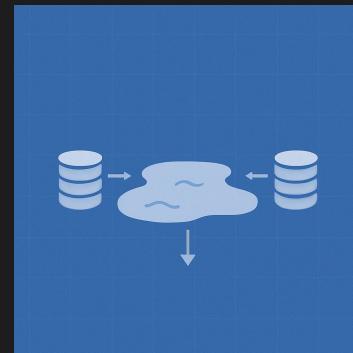


# Data warehouse / Data Lake / Data Lakehouse



## Data warehouse

Structured, high-quality data optimized for fast querying and reporting.



## Data lake

Centralized storage for raw, unstructured, or semi-structured data at scale.



## Data Lakehouse

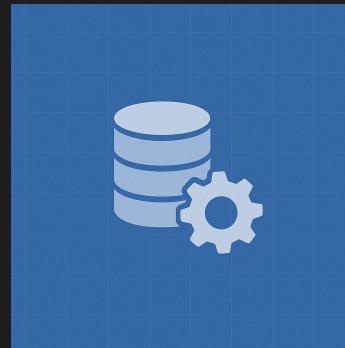
Combines the reliability of warehouses with the flexibility of lakes for unified analytics.

# OLAP / OLTP



OLAP

Fast, transactional processing for day-to-day operations like inserts, updates, and deletes.



OLTP

High-performance querying and analysis on large volumes of historical data.

# Other benefits of VisierDB

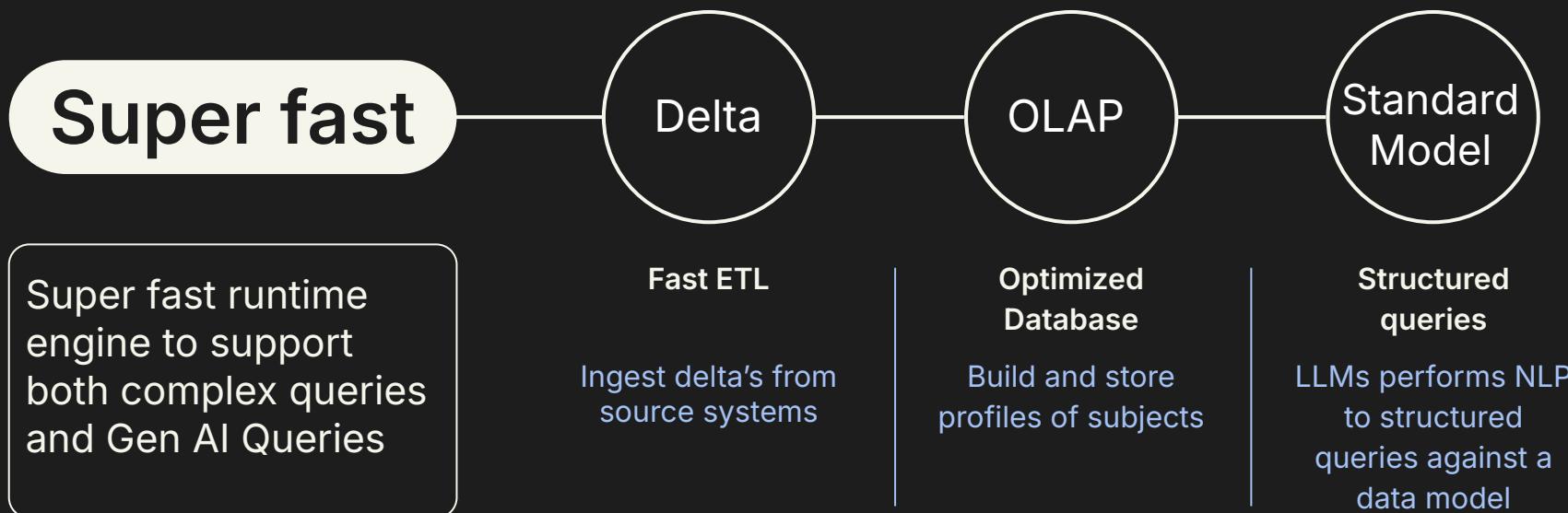


Time  
Machine

Secured  
once by  
Subjects

Combine  
data from  
many  
sources

# So how do you make analytics super fast?



# Key takeaway about the data layer

# So... that's it? Are we done?

## Building teams

Do you need a large team to be effective?

Can Gen AI solve everything?

## ETL

Do you need ETL, can't you just dump everything into a bucket for LLM?

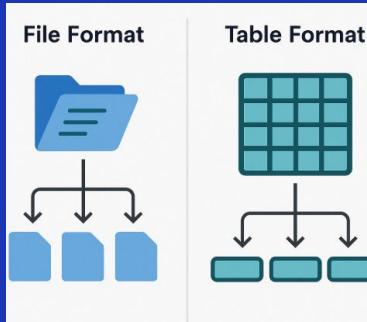
## Streaming

What if you want to do analysis really fast?

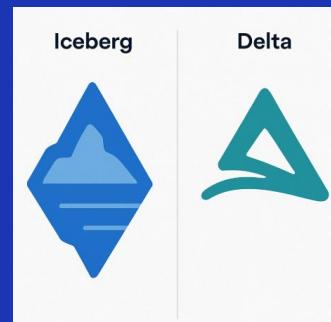
## Reporting with AI

Is Gen AI going to hallucinate?

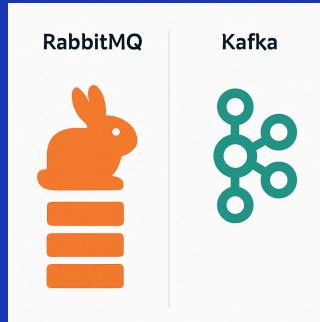
# Few thoughts on data streaming



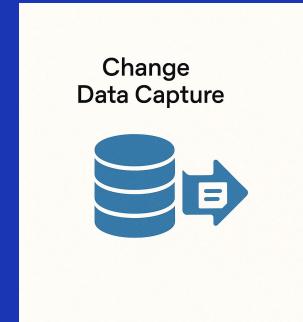
File vs table format



Iceberg and  
"Delta Lake"



Message queues



Change data capture

"Apache Iceberg: The  
Definitive Guide"

Kafka:  
The Definitive Guide

# Putting everything together

## Building teams

Do you need a large team to be effective?

Can Gen AI solve everything?

## ETL

Do you need ETL, can't you just dump everything into a bucket for LLM?

## Streaming

What if you want to do analysis really fast?

## Reporting with AI

Is Gen AI going to hallucinate?

Thank you!

## Building Effective Data Teams: Streaming, ETL, and LLM-Powered Reporting



You can find me on LinkedIn:

<https://www.linkedin.com/in/vincentchu/>

We are also hiring for co-ops to start in August!