

# Beam 3.0 - a vision for the future

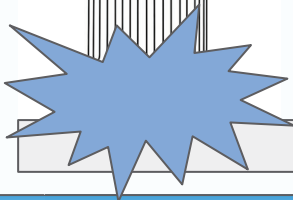
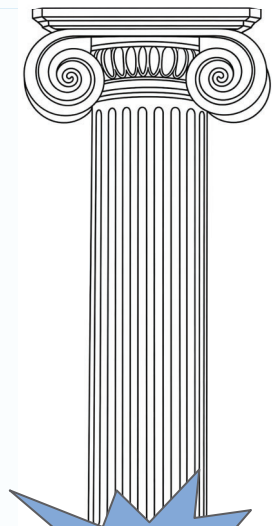
## 3 Pillars of Beam 3.0

How do I even  
write my data  
processing?

Beam's Trailblazing Foundation

## 3 Pillars of Beam 3.0

Access for All



Beam's Trailblazing Foundation



# Access For All: Expansive IO Ecosystem



debezium



redis



cassandra



Parquet



PULSAR



influxdb



elasticsearch



00011110001100  
11000010100110  
01100011000010  
1110000011011  
11000000101101  
01100000101100  
101110011010101  
10101010100011010  
011100000011010

Apache™  
Tika



APACHE  
HBASE



mongoDB®



kafka



Cloud SQL



Datastream



Datastore



Data Catalog



Vertex AI



Cloud Spanner



Cloud Bigtable



Firestore



Healthcare API



BigQuery



Pub/Sub



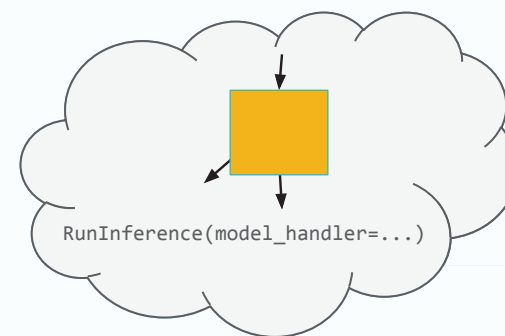
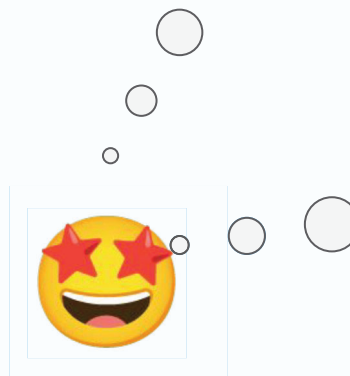
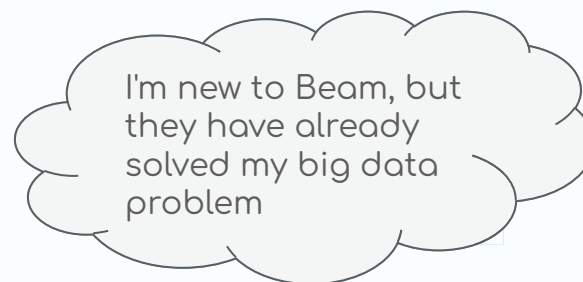
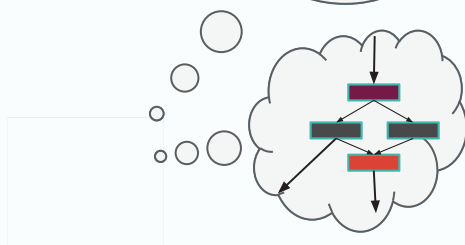
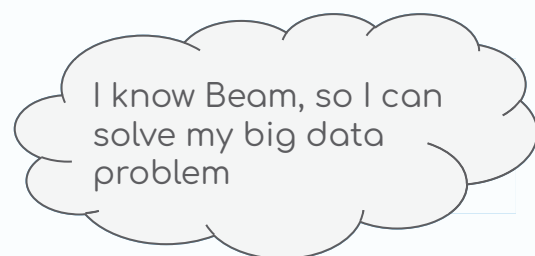
Cloud Storage

ICEBERG





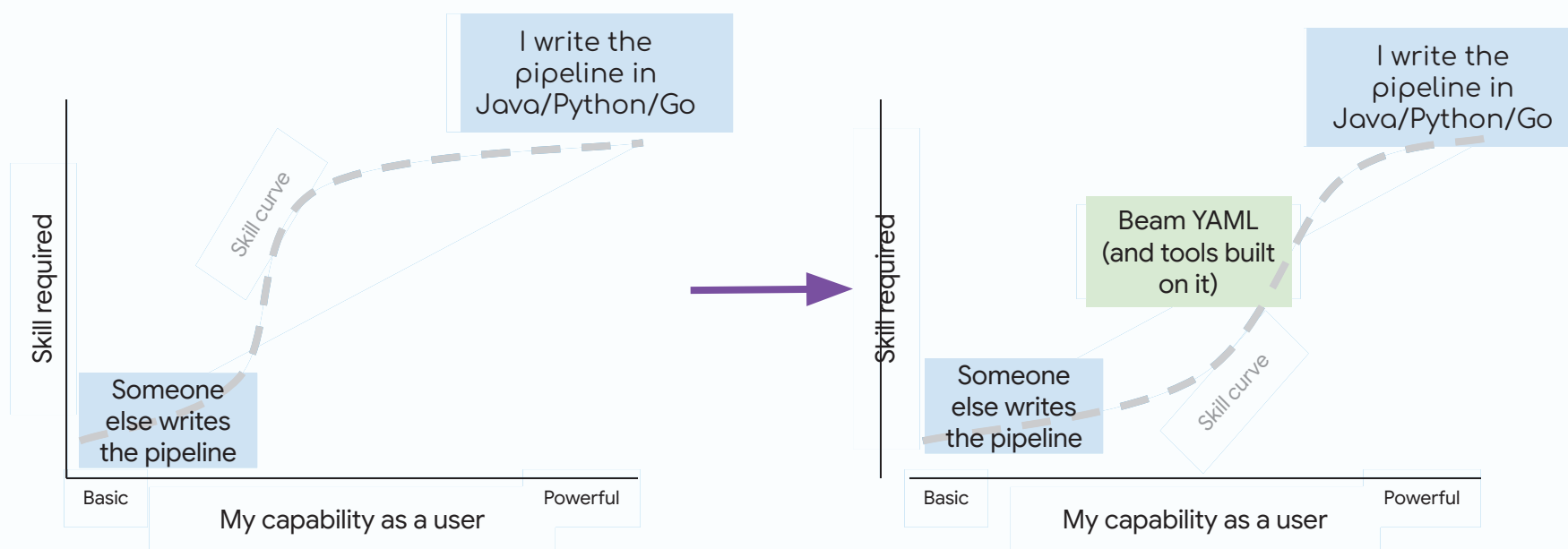
## Access For All: Turnkey transforms



`RunInference(model_handler=...)`

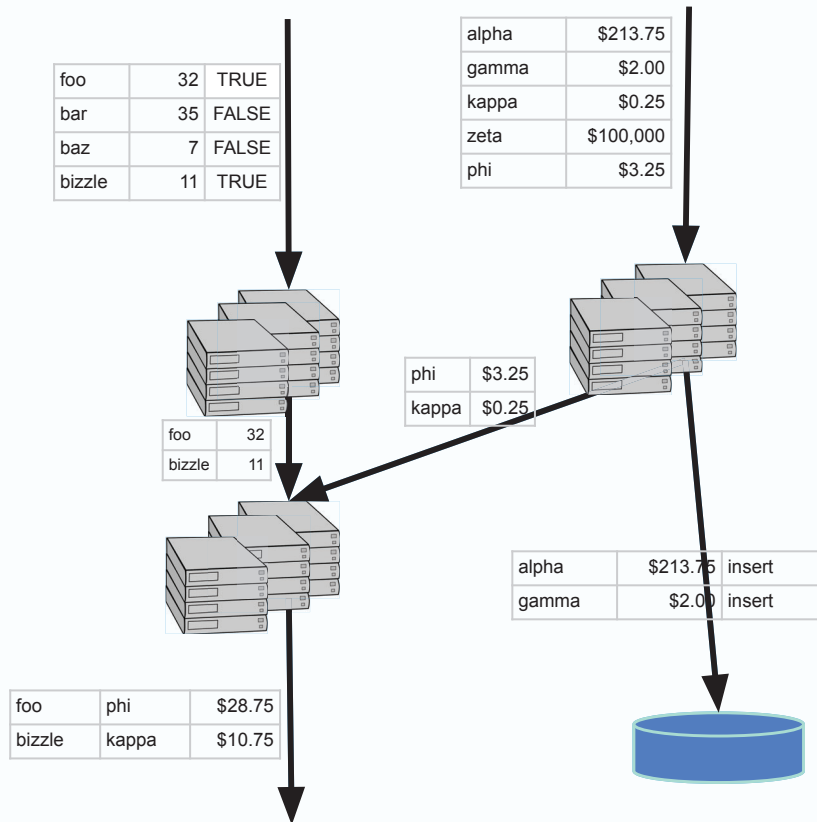
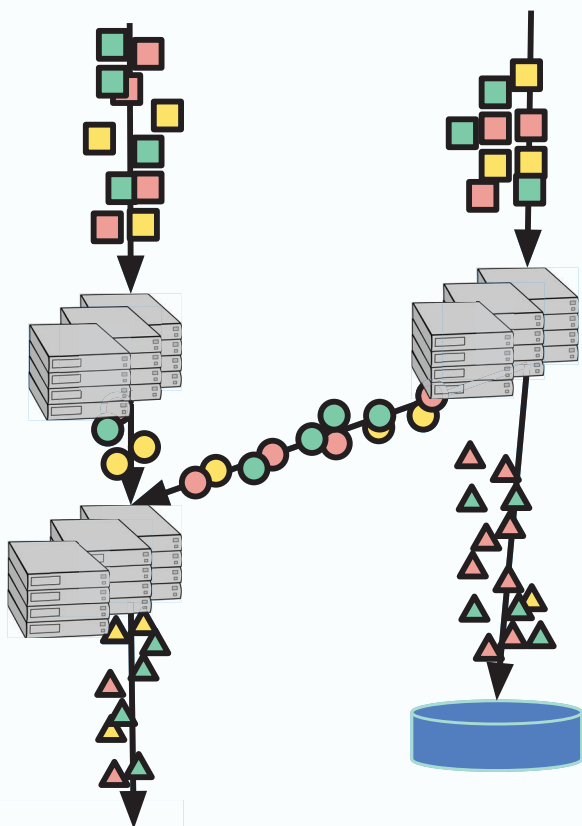


# Access For All: Beam YAML



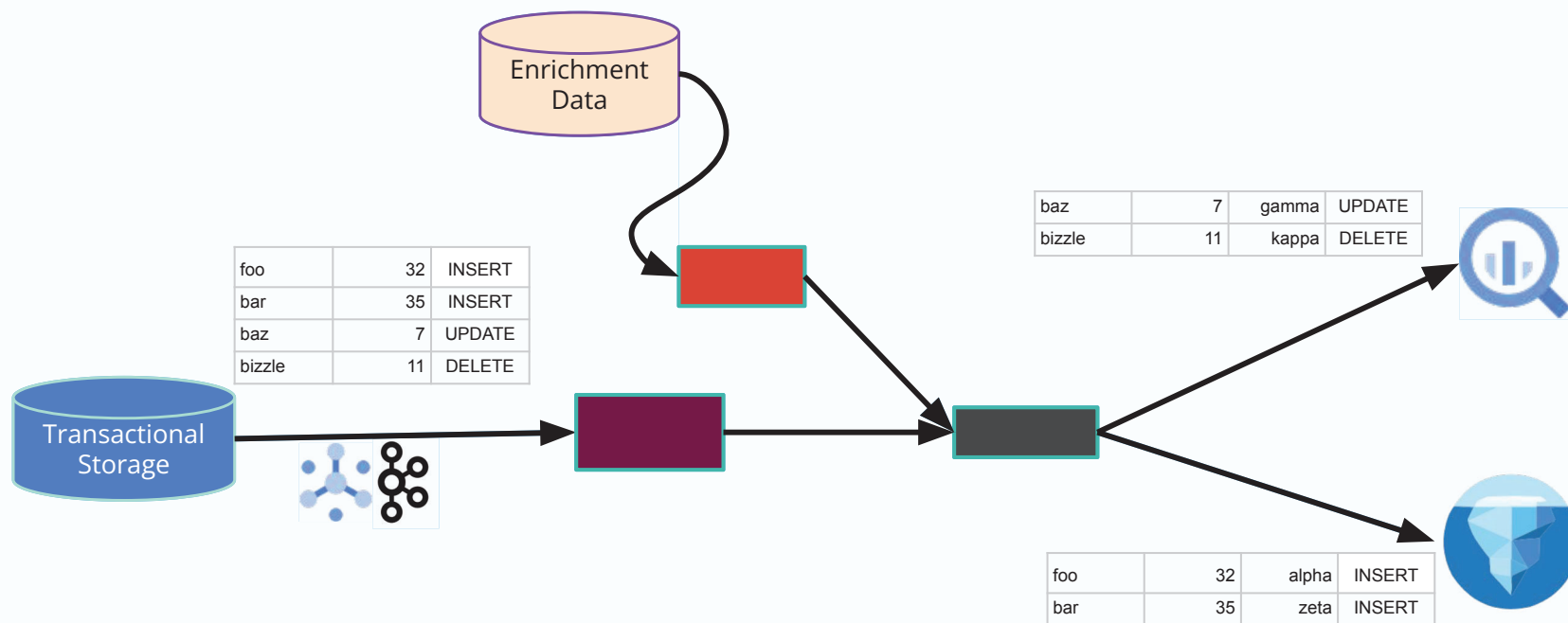


# Access For All: Structured Data (and transforms)





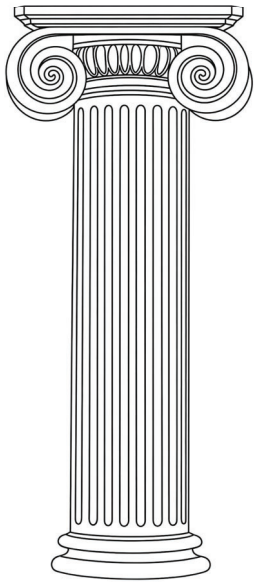
# Access For All: CDC / Lakehouse





## 3 Pillars of Beam 3.0

Access for All

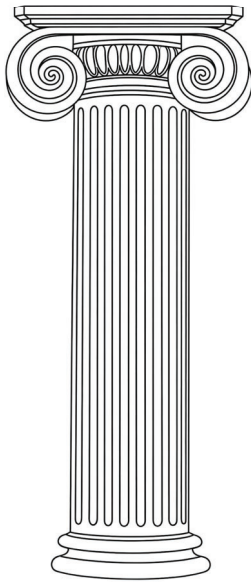


How do I  
maintain a  
production  
pipeline?

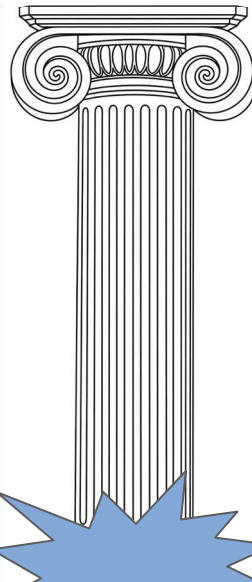
Beam's Trailblazing Foundation

## 3 Pillars of Beam 3.0

Access for All



"Just works" ops

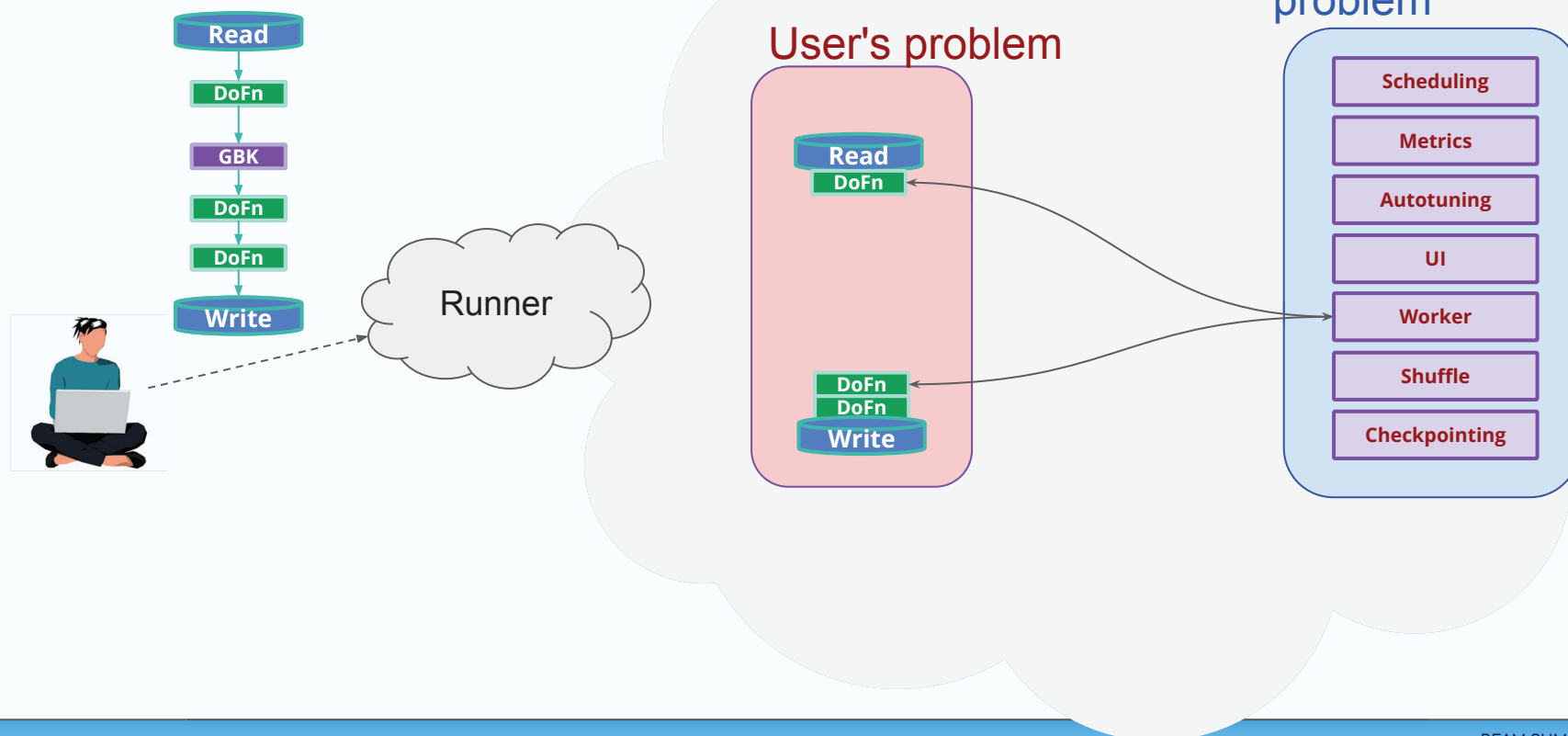


Beam's Training Foundation



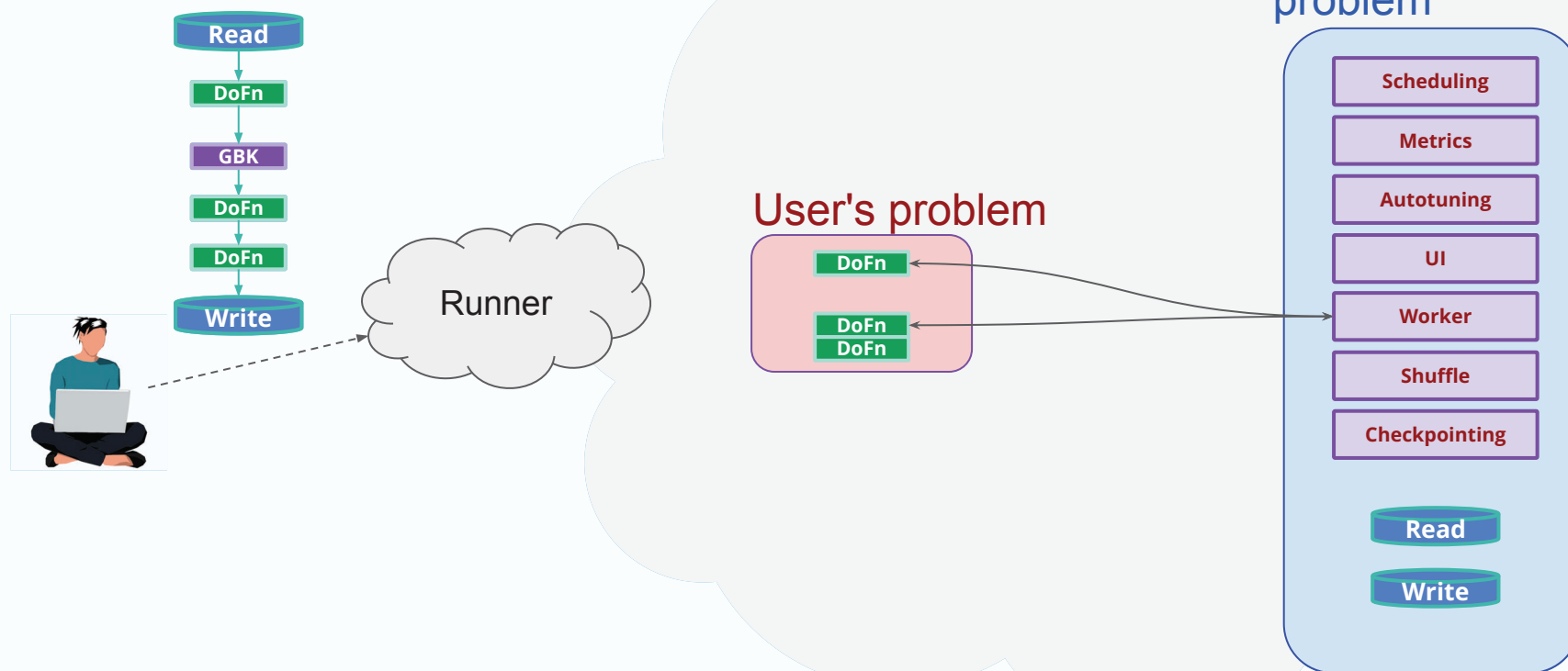


# "Just works" ops: Managed Transforms



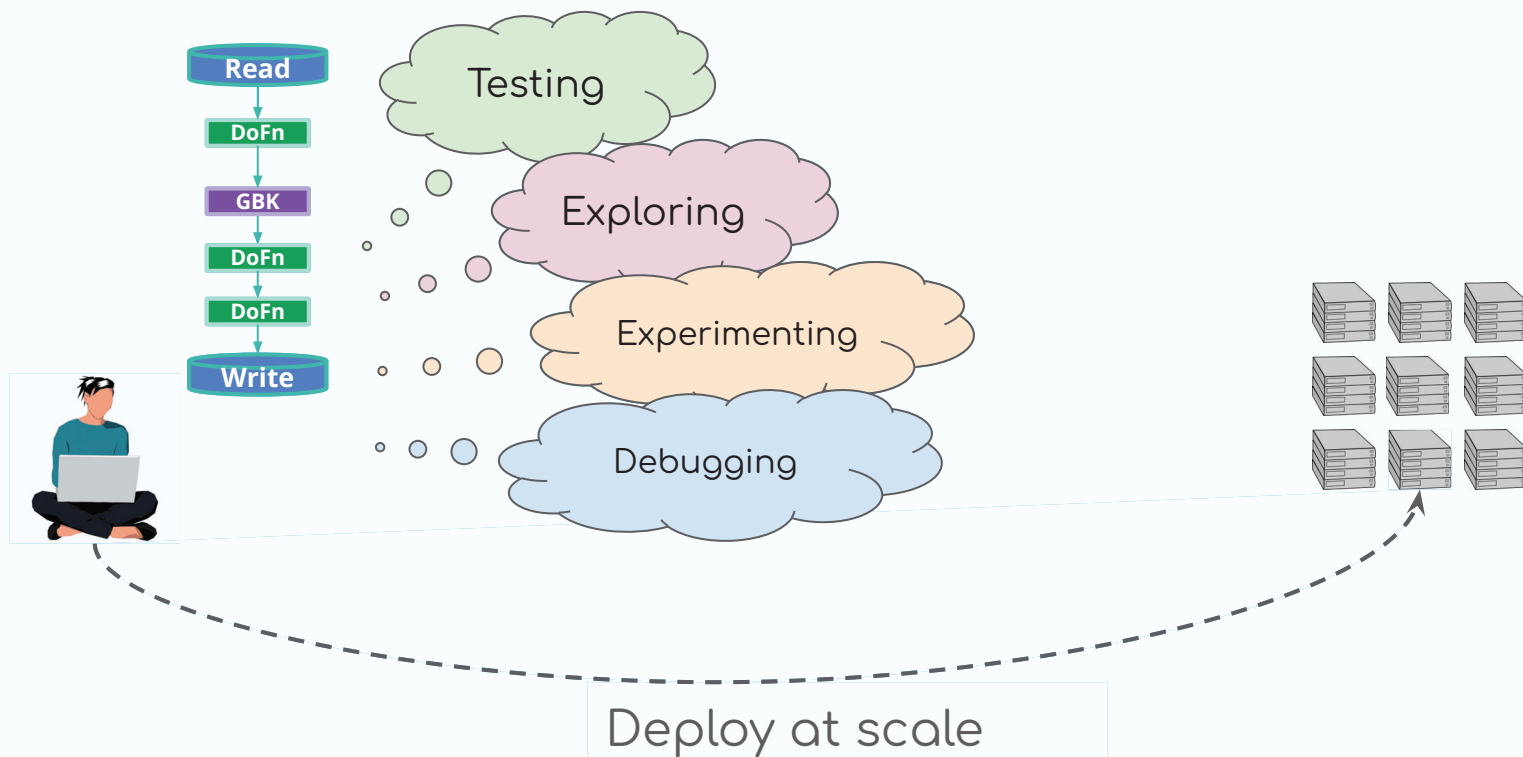


# "Just works" ops: Managed Transforms



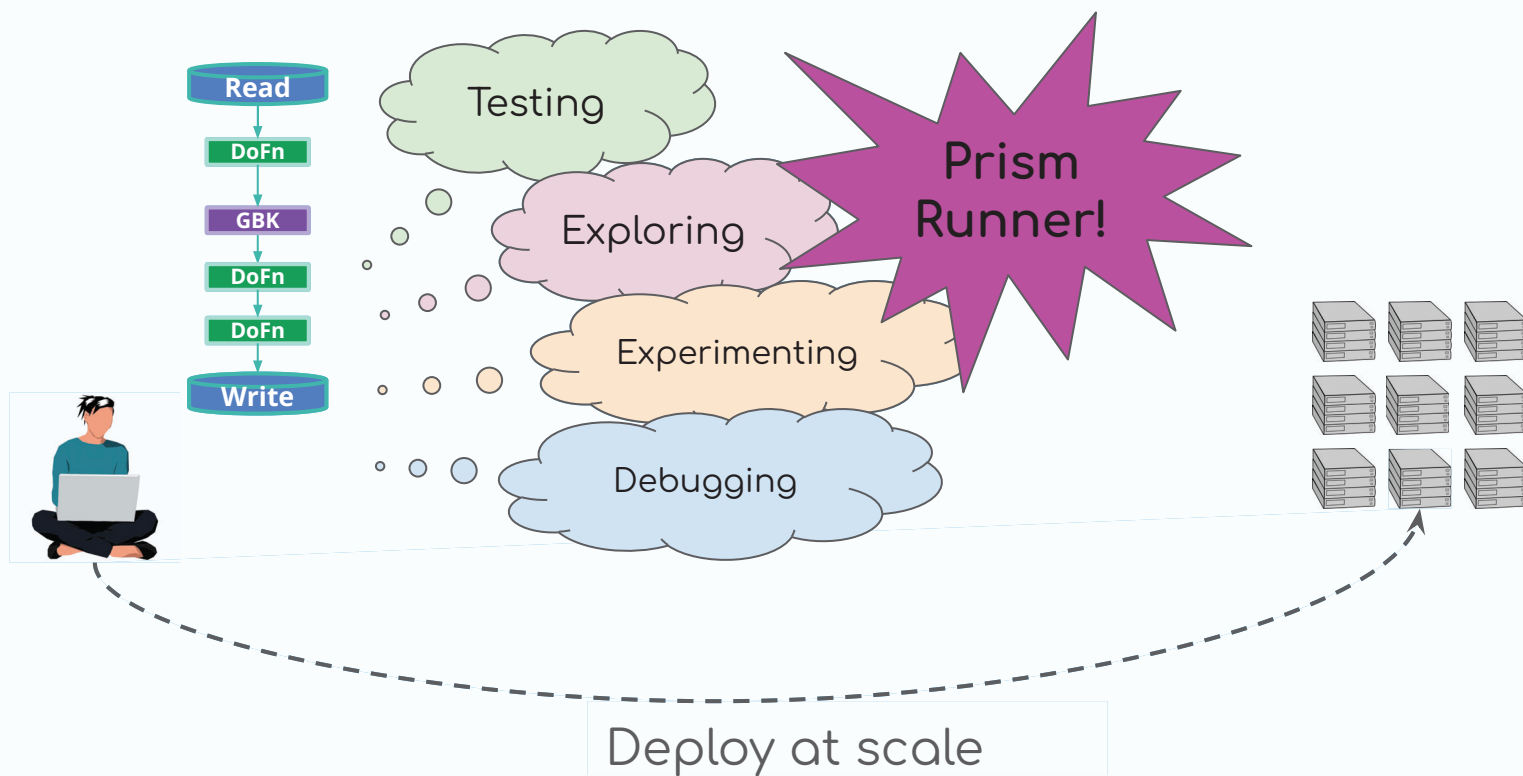


# "Just works" ops: Fully Featured Local Experience



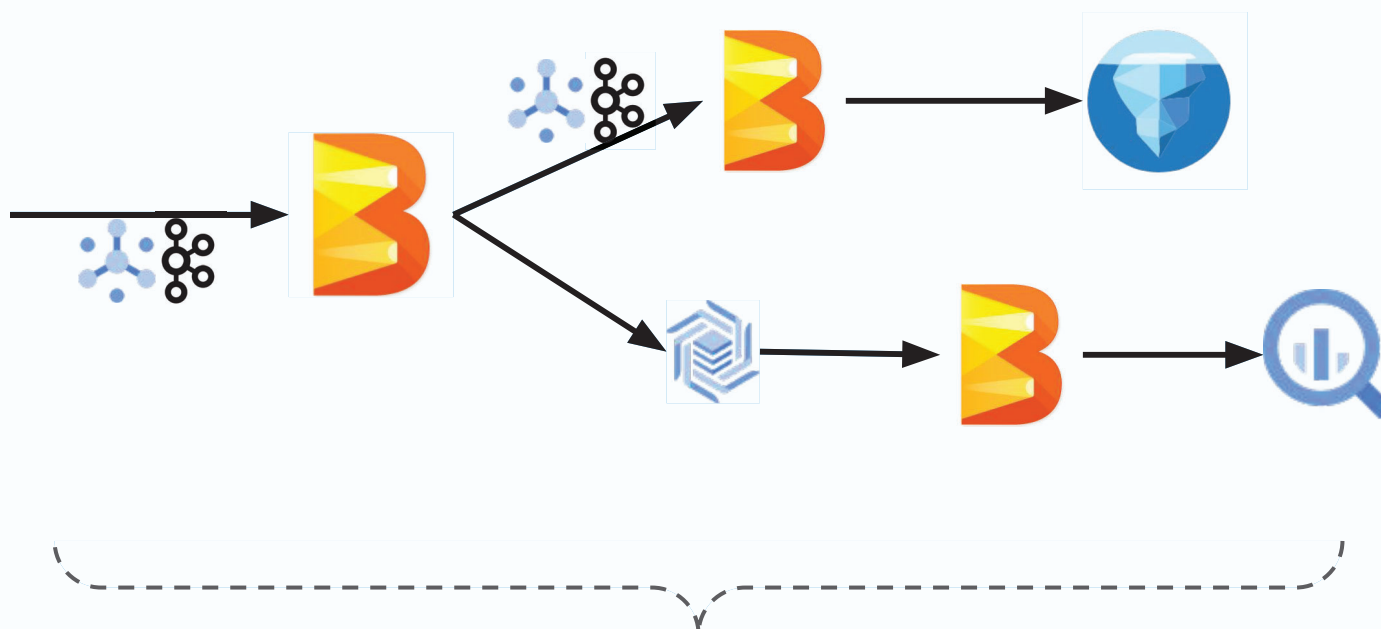


# "Just works" ops: Fully Featured Local Experience





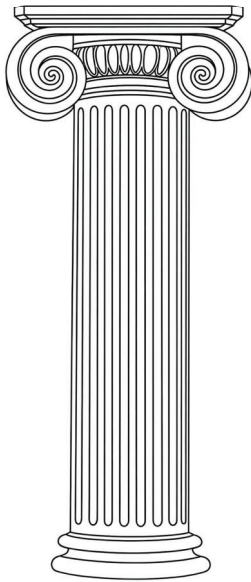
## "Just works" ops: whole app tracing



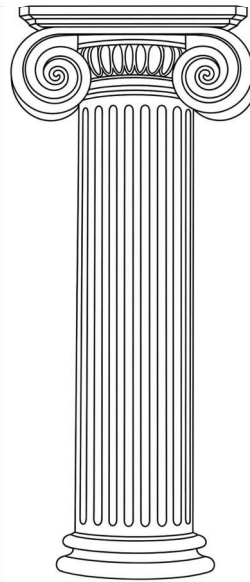
what's the latency? where's the error?

## 3 Pillars of Beam 3.0

Access for All



"Just works" ops



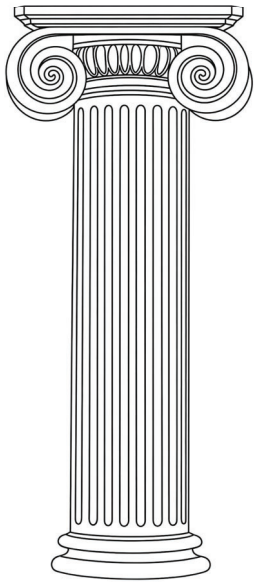
How do I  
AI/ML/LLM?

Beam's Trailblazing Foundation

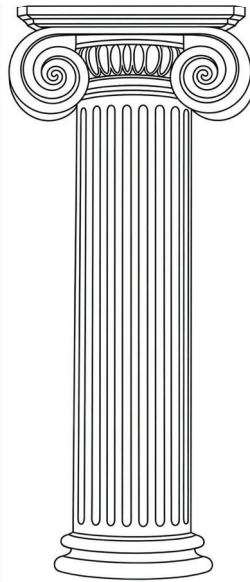


## 3 Pillars of Beam 3.0

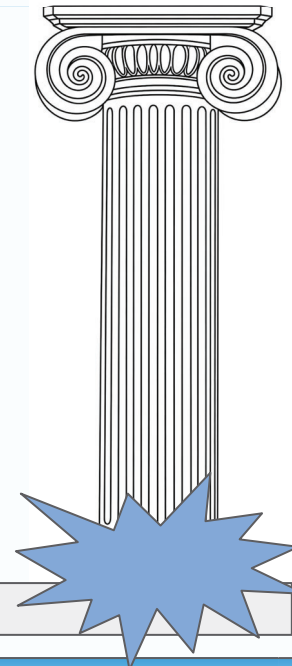
Access for All



"Just works" ops



First class ML



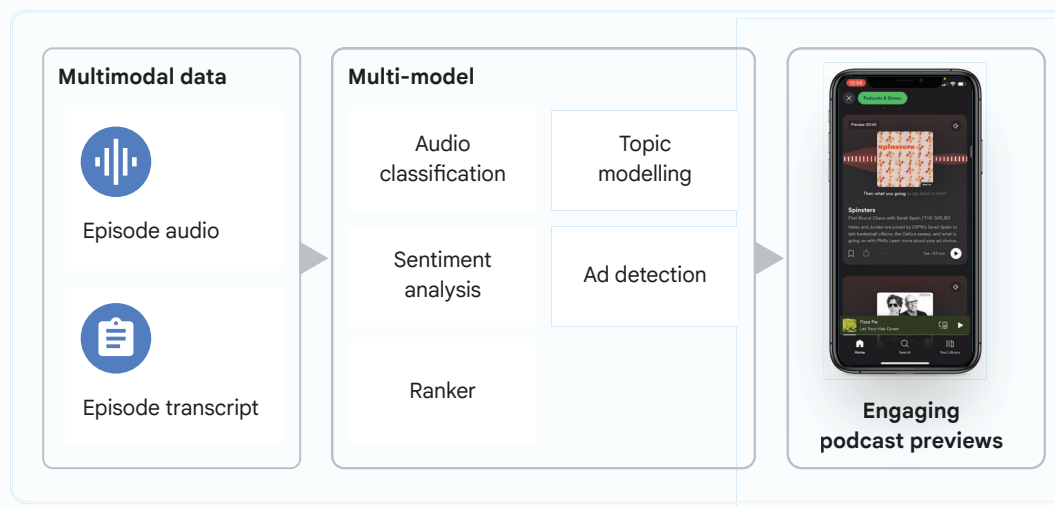
Beam's Trailblazing Foundation



## First class ML: Model Serving



- Real time and bulk inference
- Efficient, easy to use

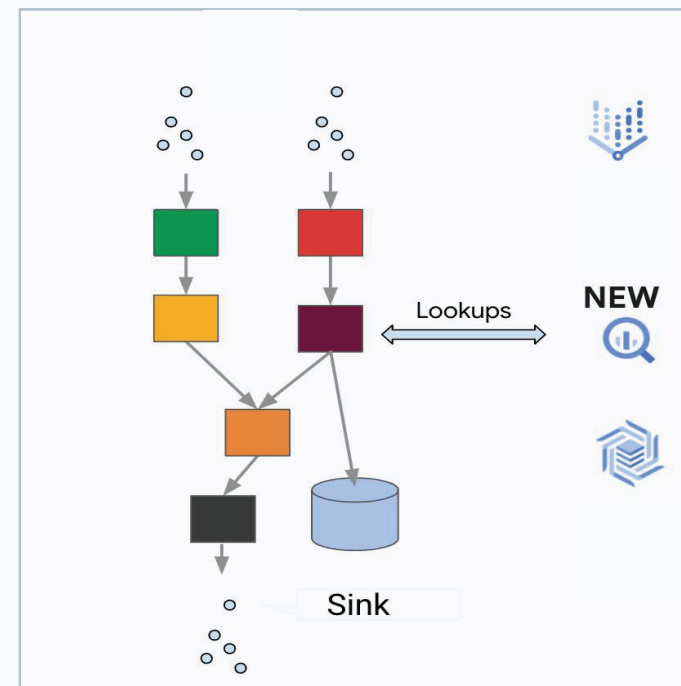




## First class ML: Feature Generation



- Low code feature enrichment
- Easy primitives for building features
- Tight integration with the ecosystem

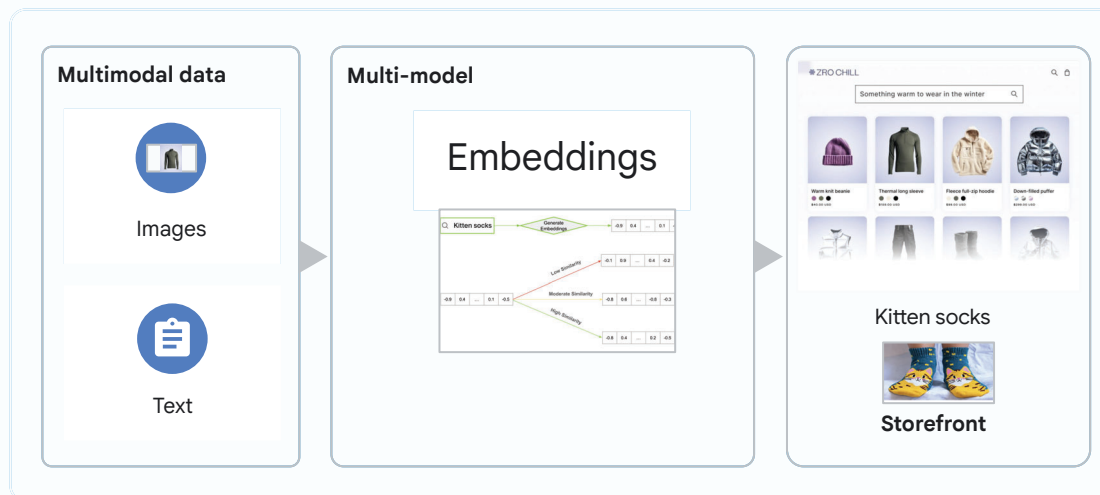




# First class ML: Embedding Generation



- Generate embeddings for search or RAG
- Low code, easy to integrate with vector DBs

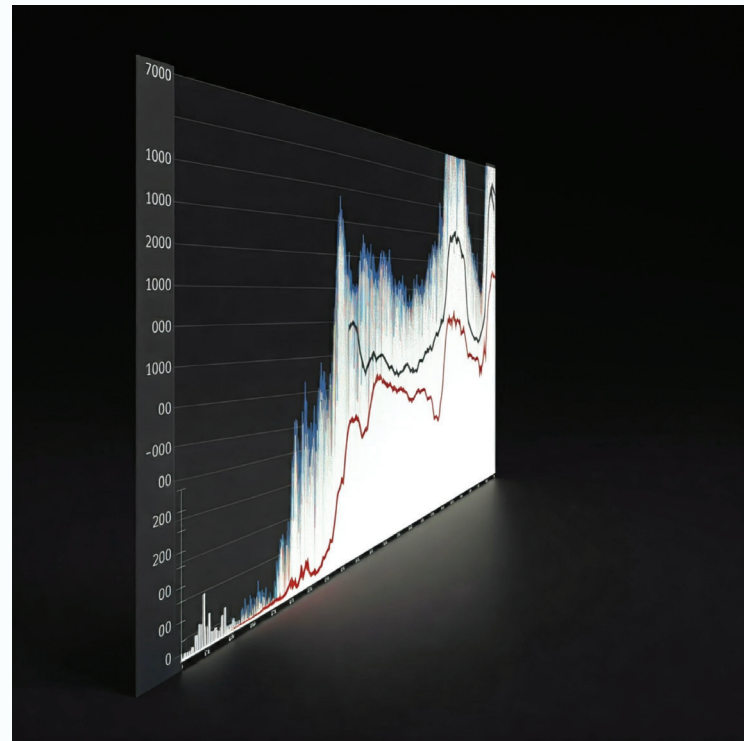




## First class ML: Anomaly Detection



- Low code
- Streaming
- Online





## First class ML: New Capabilities



- Model Evaluation
- Training
- Efficiency Improvements
- Whatever latest/greatest ML capabilities come next

 PyTorch



ONNX

*dmlc*  
**XGBoost**





## First class ML: Talks Today



Efficient LLM-Based Data Transformations Transformations via Multiple LoRA Adapters	Konstantin Buschmeier	10:30
Enhancing Data Quality for AI Success	Vijay Shekhawat	11:00
Introduction to the Apache Beam RAG package	Claude van der Merwe	11:30
Data Quality in ML Pipelines	Pritam Dodeja	11:30
Scalable Prompt Optimization in Apache Beam LLM Workflows	Tomi Ajakaiye	12:30



## First class ML: Talks Yesterday

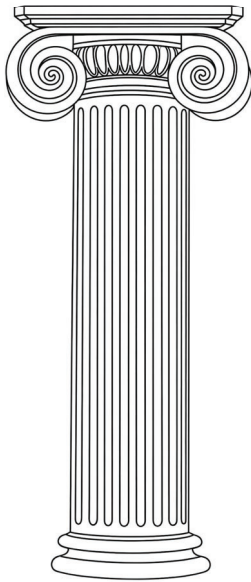


Scaling Real-Time Feature Generation Platform @Lyft	Rakesh Kumar
From taming energy market data to hyperparameter hunting at scale: leveraging Apache Beam & BigQuery	Nicholas Bonfanti & Matteo Pacciani
Real-time Threat Detection at Box with Apache Beam	Abhishek Mishra, Mark Chin & Elango Prasad
Leveraging LLMs for Agentic Workflow Orchestration in Apache Beam YAML Pipelines	Charles Adetiloye
Remote LLM Inference with Apache Beam: Practical Guide with Gemini and Gemma on Vertex AI	Taka Shinagawa
Real-Time Predictive Modeling with MLServer, MLFlow, and Apache Beam	Devon Peticolas & Jeswanth Yadagani
Superpowering Agents with Apache Beam	Konstantin Buschmeier & Jasper Van den Bossche
Talk to your pipeline: how to use AI to create dynamic transforms in streaming	Israel Herraiz
Simplified Streaming Anomaly Detection with Apache Beam's Latest Transform	Shuning Huang
How Beam serves models with vLLM	Danny McCormick

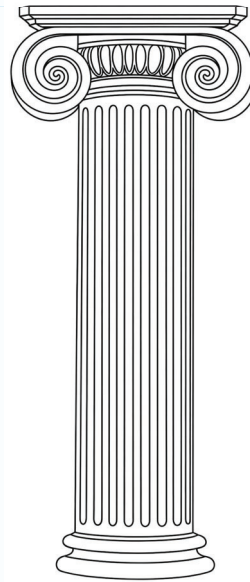


## 3 Pillars of Beam 3.0

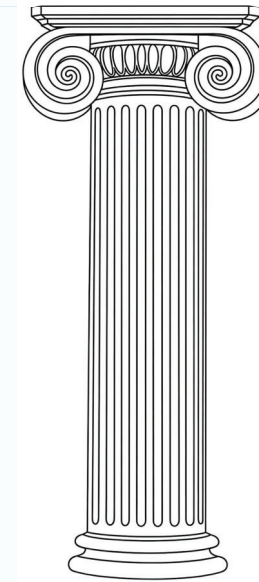
Access for All



"Just works" ops



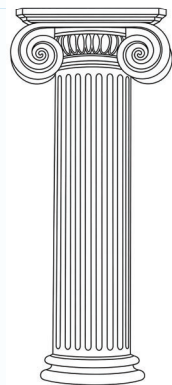
First class ML



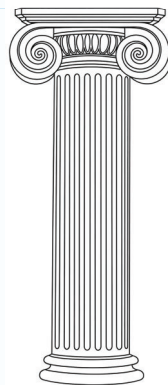
Beam's Trailblazing Foundation

Most important part: the foundation

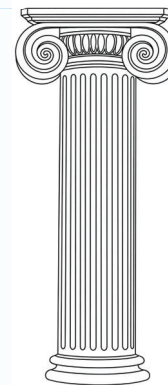
Access for All



"Just Works" ops



First class ML



Beam's Trailblazing Foundation

Beam Community

Where will you take Beam 3.0?

Kenneth Knowles, Danny McCormick

# QUESTIONS?

damccorm  
kennknowles

**BEAM**  
SUMMIT  
NYC 2025