Data Council **2024**

From Twilio to Propel: Building Real-Time Customer-Facing Analytics at Scale

Nico Acosta - Co-founder and CEO at Propel
@acossta

Propel

I'm Nico Acosta

- Computer Scientist.

- Co-founder and CEO at Propel.

- I was employee 65 at Twilio. Spent a decade building API, data, and AI products.

- AWS Global expansion.

- Co-founder of Masivian.

Part 1

Building customer-facing analytics at Twilio

✔ Comms Platform as a Service

✔ SMS, Voice, and Email APIs

✔ Public company. IPO'd in 2016

✔ > 3 TB of data per day

✔ > 4 billion raw events per day

Part 2

Lessons learned and building Propel

✔ Analytics Platform as a Service

✔ Customer-facing analytics

✔ APIs, SQL, Embedded & Portals

✔ Built to handle insane scale

✔ Completely Serverless

Part 3

Live coding: Real-time customer-facing analytics in 5 mins

✔ Unify data from multiple sources

✔ Transform in real-time

✔ Data-Serving APIs and SQL

✔ Build UI with React Components

✔ Multi-tenant access controls.

Customer-Facing
Analytics

>

Embedded
Dashboards

Customer-Facing Analytics

Customer-facing analytics are insights provided directly to you, <u>the customer</u>, about the service or product you are using.

What do the best Customer-Facing Analytics experiences have in common?

$$$

Customers are willing to pay for them

What does it take to build an analytics product that customers <u>are willing to pay for?</u>

Sub-second query latencies

Sub-minute data freshness

Consume data multiple ways

Customer-Facing Analytics at Twilio

- Processing over **3 TB** per day
- Over **5 billion** raw events per day
- Every call emitted an event every 10 seconds
- All processed in real time
- Powering mission critical applications

Mission Critical Use Cases @ Twilio

- **Voice Insights:** Real time insights for debugging.

- **Fraud detection:** Direct and indirect fraud.

- **Carrier routing:** Quality and cost optimization.

A lot of things can (and do) go wrong
When operating communications
at a global scale

Voice Insights at Twilio

High-value customers had unique reporting requirements.

Simple, in-product dashboards didn't cut it

Ad-hoc and recurring data requests.

They were a poor customer experience and labor intensive to fulfil

High-value customers needed to consume data their way.

Via embedded dashboards in the Console, SQL, BI Tools, or APIs.
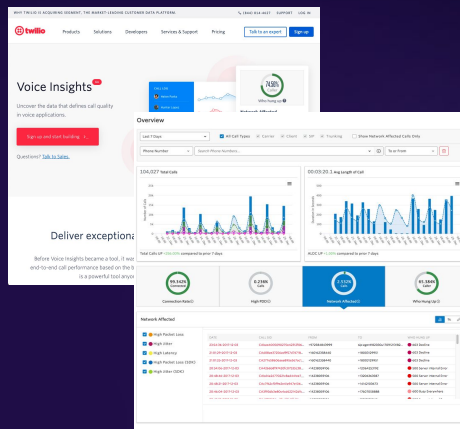
UI   API   Streams

Serve data with sub-second query latencies

Transform and enrich in real time

Unify data across multiple data sources

Gave high-value customers the confidence to scale on the platform

Scaled the voice product from $10M ARR to $100ARR in the next 3 years.

Freed-up teams to focus on the core job.

Eliminated all ad-hoc data requests.

Became one of the fastest revenue-generating products

60% net-dollar expansion on adopters

5 billion events per day	⟷	Sub-second query latencies
Lots of dimensions and time zones	⟷	Sub-minute data freshness
Really expensive to run	⟷	Profit center

Productizing Analytics at scale:
The Fundamental Trade-off

*At scale, you can have
2 of the 3,
but not all 3!!*

Cost-effectiveness
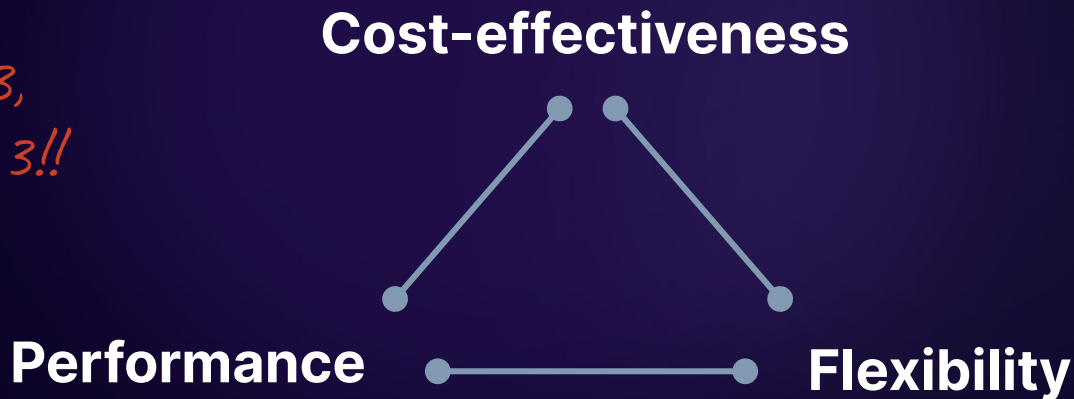
*It's a product
decision first*

Performance

Flexibility

Productizing Analytics at scale:
The Fundamental Trade-off
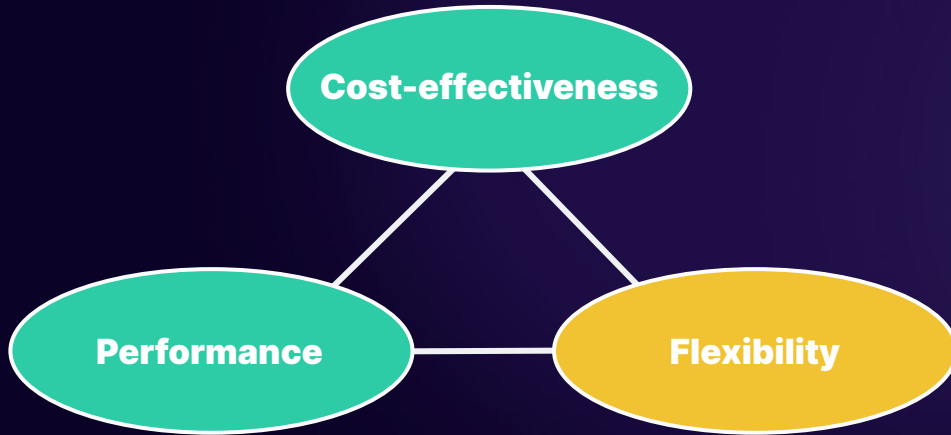
Decisions we made (not all of them ideal):

- Aggregate in UTC
- Key Performance Metrics:
  - Duration, pdd, and jitter.
- Aggregations:
  - Min, Max, Avg, 95th Percentile, 99th Percentile
- Dimensions:
  - Predefined set of dimensions

Twilio Voice Insights
Architecture
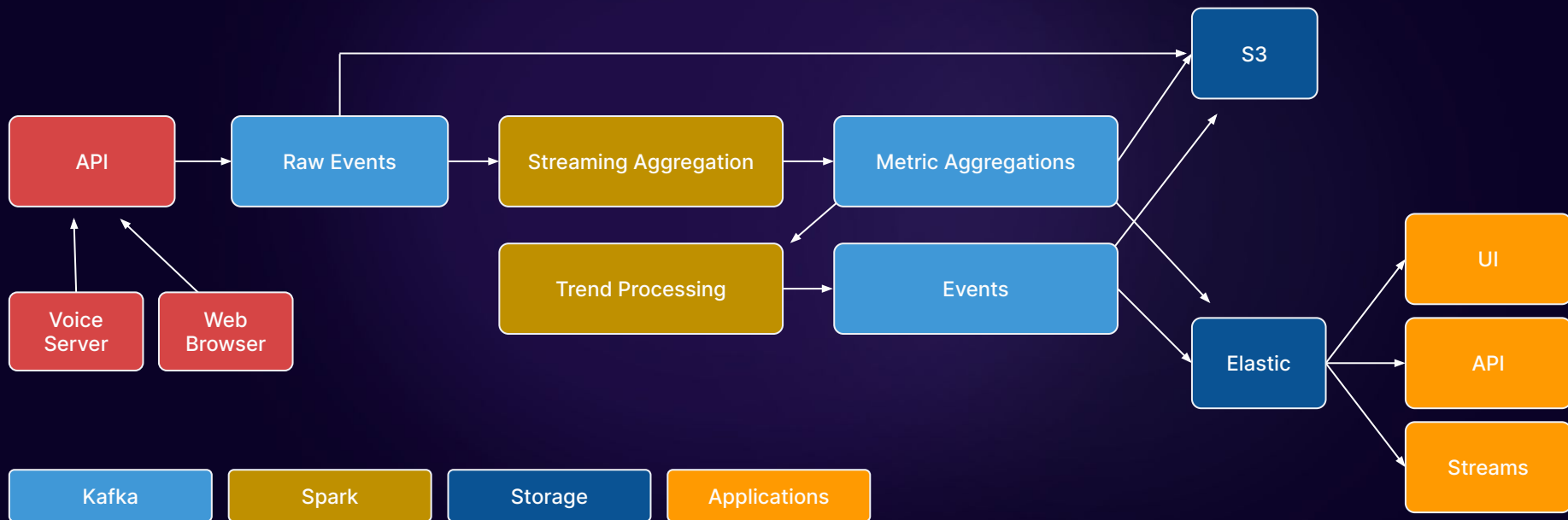
Twilio Voice Insights
Data Contracts

Data quality is the responsibility of the producer

```
message CallEvent {
  // when the event occured utc
  optional uint64 event_time
  // when we received said event utc ***
  optional uint64 logged_event_time
  // unique identifier for said event - 
  optional string event_id
  // encapsulates a logical end to end r
  optional string route_id
  // used to store multiple types in thi
  optional CallEventType event_type
  // when event type is signaling_event 
  optional SignalingEvent signaling_even
  // use with the event_type and logical
  optional Dimensions event_dimensions
}
```

- Time:
  - Source timestamp can't be always trusted.
  - ALWAYS log the time you receive the event

- ID is idempotent , i.e. the dedupe key

- Extensible to new metrics

- Extensible to new dimensions

Twilio Voice Insights
Data Contracts

Data quality is the responsibility of the producer

```
message CallEvent {
  // when the event occured utc
  optional uint64 event_time
  // when we received said event utc ***
  optional uint64 logged_event_time
  // unique identifier for said event -
  optional string event_id
  // encapsulates a logical end to end r
  optional string route_id
  // used to store multiple types in thi
  optional CallEventType event_type
  // when event type is signaling_event
  optional SignalingEvent signaling_even
  // use with the event_type and logical
  optional Dimensions event_dimensions
}
```

```
message SignalingEvent {
  optional string name                    = 1;
  optional SignalingEventType event_type  = 2;
  optional PddEvent pdd                    = 3;
  optional CallStateEvent call_state       = 4;
}

enum SignalingEventType {
  unknown_event_type  = 0;
  pdd                 = 1;
  call_state          = 2;
}

message PddEvent {
  optional float pdd = 1;
}
```

```
message MetricAggregation {
  optional string metric            = 1;
  optional uint64 window_start      = 2;
  optional uint64 window_end        = 3;
  optional string window_interval   = 4;
  optional uint32 samples           = 5;
  optional Stats stats              = 6;
  optional Histogram histogram      = 7;
  optional Dimensions dimensions    = 8;
  optional string dimension_hash    = 10;
}
```

From a Product to a Platform

Lessons learn at **Twilio** and how we applied them to **Propel**

Most SaaS Companies Spend a Lot of Effort & Get Little Value

Engineering Effort

Low

High

Simple, In-product Dashboard

Slow queries

Data not fresh enough

More filters

Unify multiple data sources

Data transformations

Wrestling UX customizations

?

Low

High

Customer Value

Propel
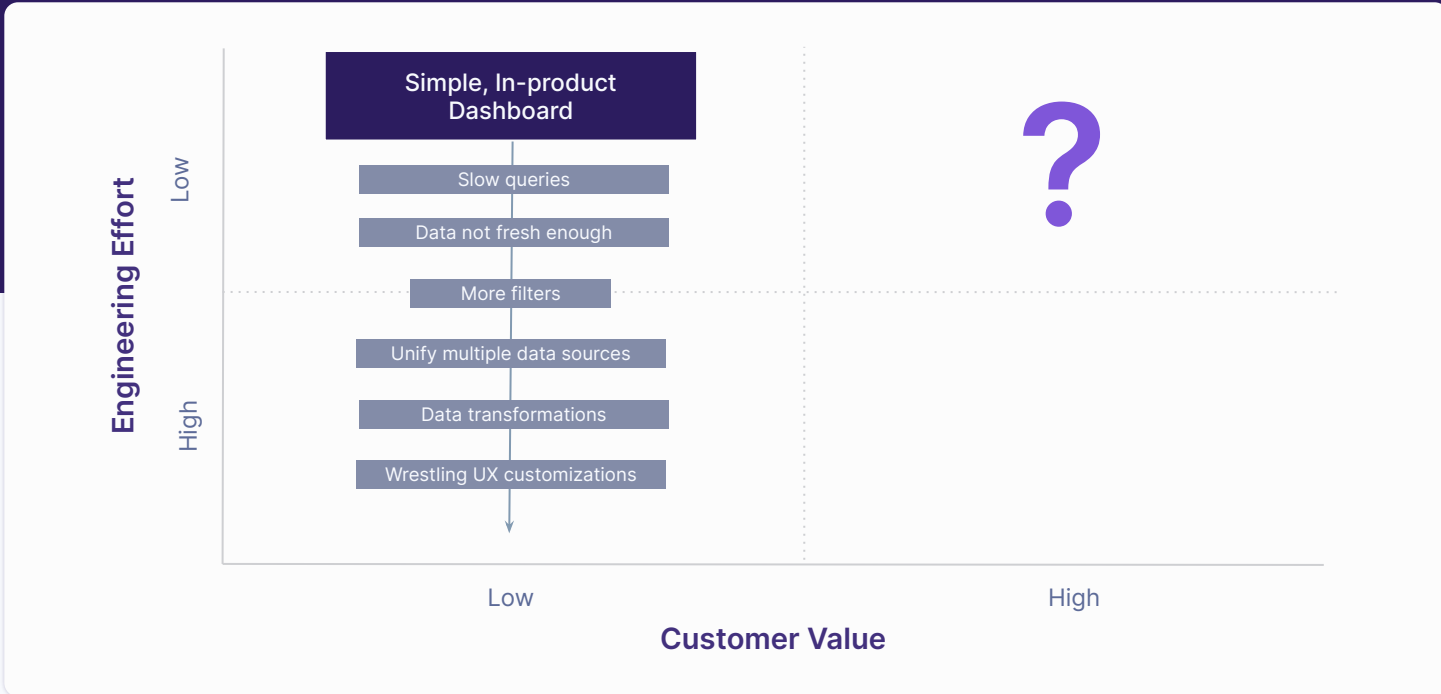
Customer-Facing Analytics Roadmap

The highest-value, low effort products are not always obvious.

	Who It Serves	Engineering Effort	Revenue Opportunity
1. SQL Connector	High-value customers	Low	High
2. White Label Reporting Portal	High-value customers	Low	High
3. Analytics APIs	High-value customers	Medium	High
4. Embedded Dashboards	Long-tail customers	Medium	Low
5. Custom-built In-product Dashboards	Long-tail customers	High	Low

From an Analytics Product at Twilio
To a Platform at Propel

Twilio

API-Only Ingestion

Real-Time Spark Streaming Aggregations (Event data)

Predefined Dimensions and aggregations

Consume via Dashboard, API, and Streams

Propel

Ingest from multiple data sources

Real-time SQL-based transformations (event and entity data)

Customer-provided dimensions and aggregations

Consume via Dashboard, API, SQL, or Reporting Portal

Propel

Propel
Architecture

Propel
Architecture

Data Warehouses

Webhooks

Streaming

Databases

Ingestion Service

React UI Kit

Access Policies

Data-Serving APIs

Metric Definitions

Data Pools

Materialized Views

ClickHouse

Access Controls

API

SQL

Data Warehouses

Report Center

Embedded Dashboards

APIs and UI Components
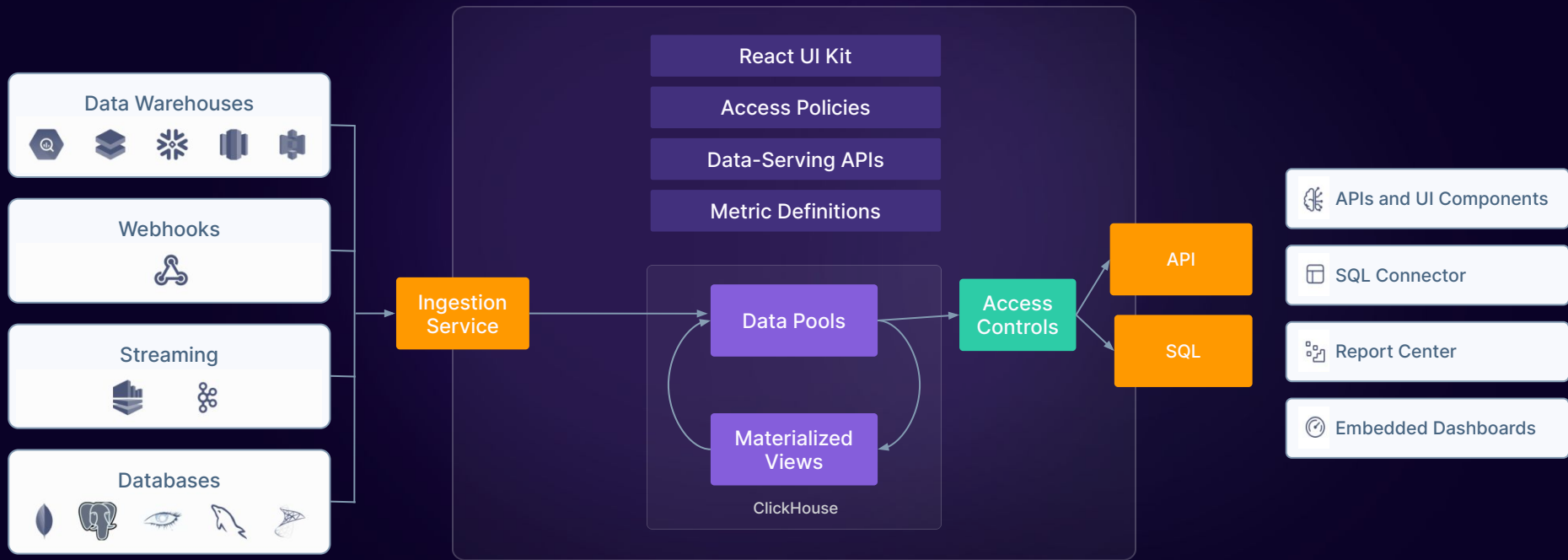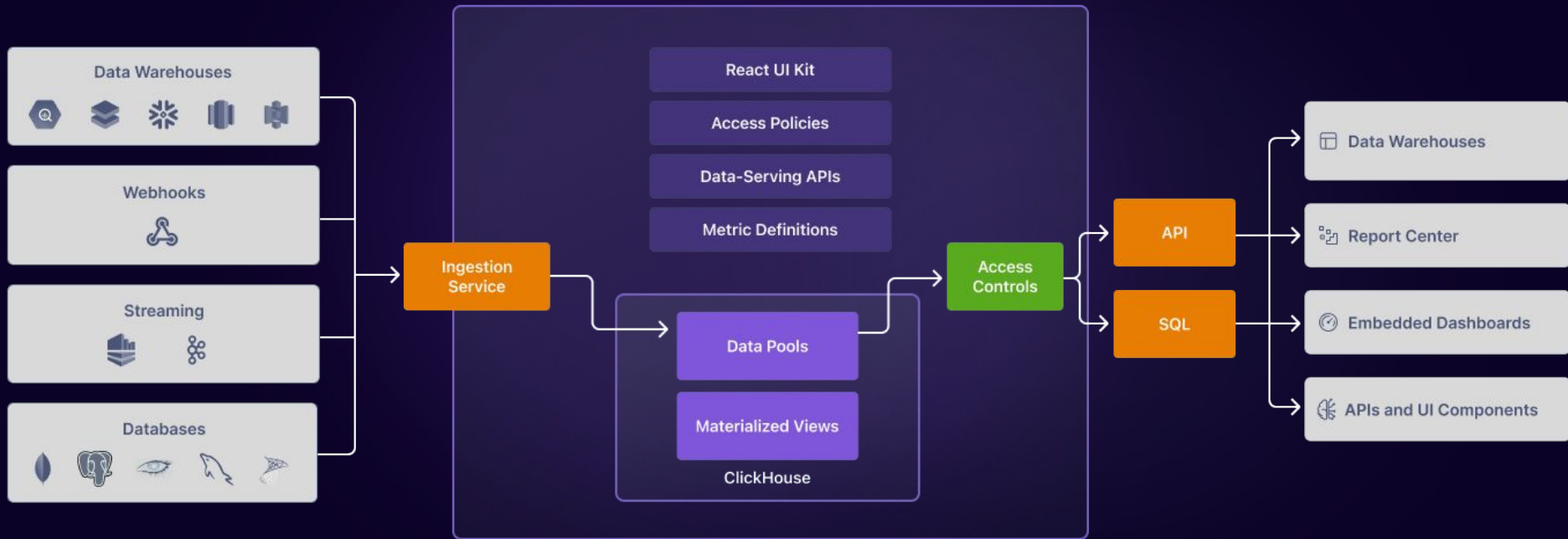
# See it live

TacoSoft, our imaginary SaaS company, ships its first real-time analytics product

- ➜ Ingest data from our data warehouse
- ➜ Ingest streaming data from Postgres-Kafka
- ➜ Transform and enrich it in real time
- ➜ Set multi-tenant access controls
- ➜ Expose it via the API
- ➜ Build it into a customer-facing dashboard

Demo recap

- **Ingest**
  - Ingested data from Snowflake
  - Ingested data from Kafka

- **Transform**
  - Flattened JSON from Kafka Messages
  - Enriched data joining datasets

- **Serve**
  - Served it via Analytics APIs
  - Types of APIs: Time Series

- **Visualize**
  - Custom dashboard in React
  - Using the Propel React components

Ingest data

Transform in real-time

Serve via API

Deliver the UI
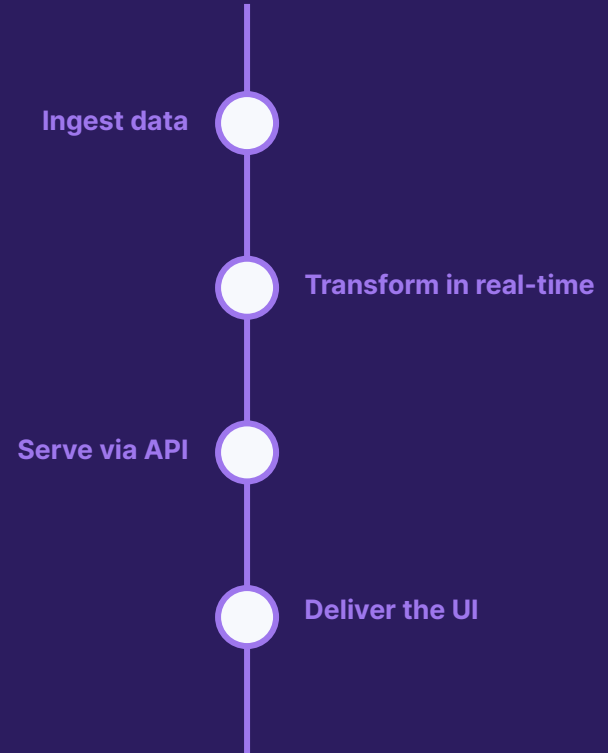
If you remember 3 things from this talk:

1. Best analytics products are those that customers are **willing to pay for**.

2. To successfully productize analytics, iterate quickly the **Cost-Effectiveness**, **Performance**, and **Flexibility** trade-off.

3. The highest-value, low effort products are **not always obvious**.

Propel is now generally available!

✅ Serverless ClickHouse
✅ Fully self-service.
✅ Generous free tier (10 GB)

www.propeldata.com

2024 | Turn customer data requests into
revenue-generating analytics products

Propel

www.propeldata.com