

How To Query A Stream?

Viktor Gamov, StarTree
@gamussa

Geecon, Kraków, 2024



Robert Zych ✓

@zychr



Had a great Q&A with John Roesler yesterday about IQ (ie Kafka streams interactive queries) v2 will it much easier to perform more complex queries. My question was basically when should use we use IQ vs something like [@ApachePinot](#) ?

7:32 AM · Jun 25, 2022



Jase

@jasonbelldata



It does.... Or a podcast. Or a blog.

8:47 PM · Jun 25, 2022

Viktor

GAMOV

Head of Developer Advocacy | StarTree

Twitter X: @gamussa



Monolith



ETL and CDC



A person wearing a red robe is seen from behind, holding a smartphone. They are standing in a dimly lit, ornate hallway with stone walls and arches. A bright light source is visible in the distance, creating a strong contrast and casting long shadows. The overall atmosphere is one of modern technology meeting a traditional, historical setting.

Mobile Era



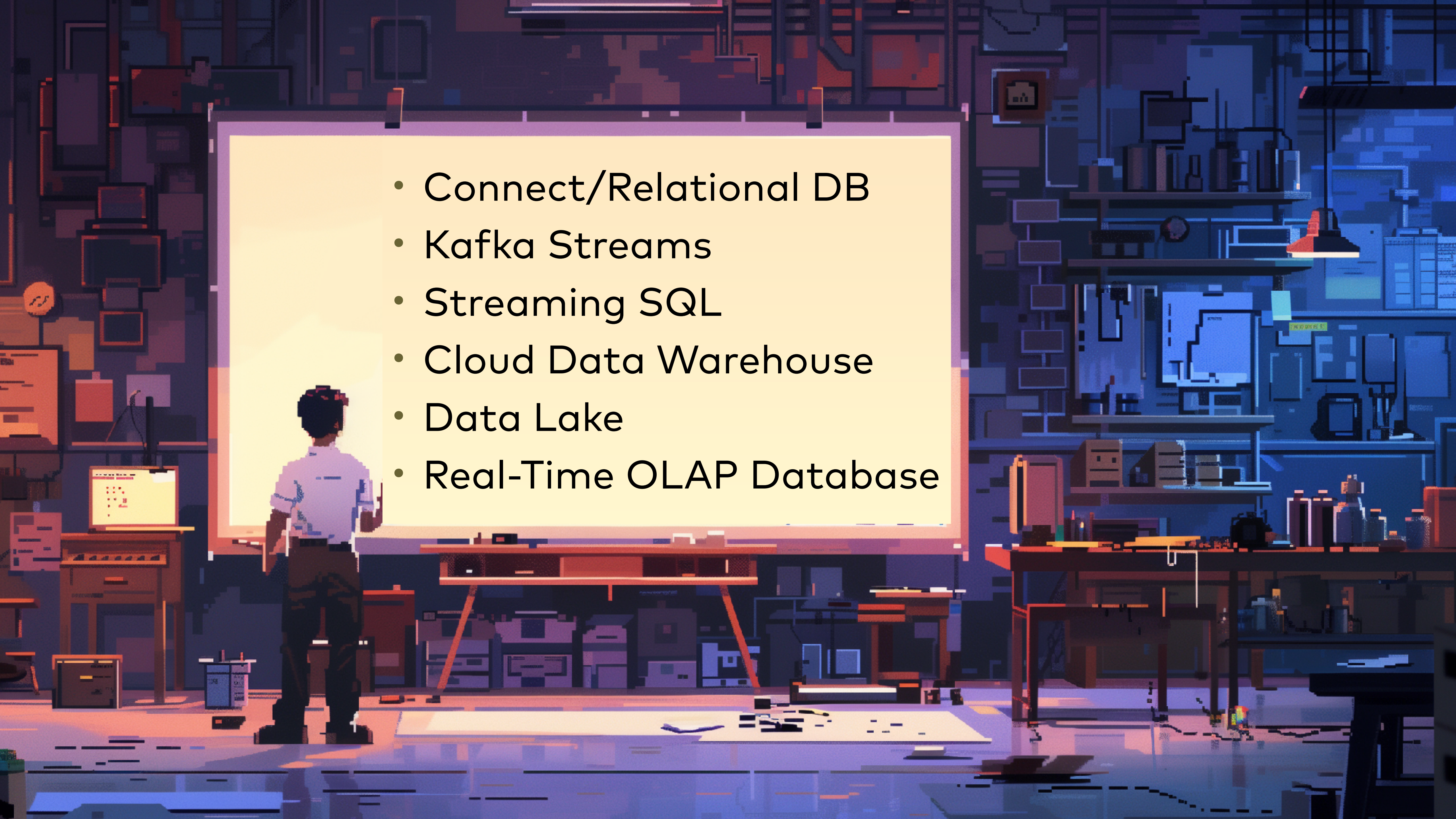
Data Pipelines and Microservices



LOG



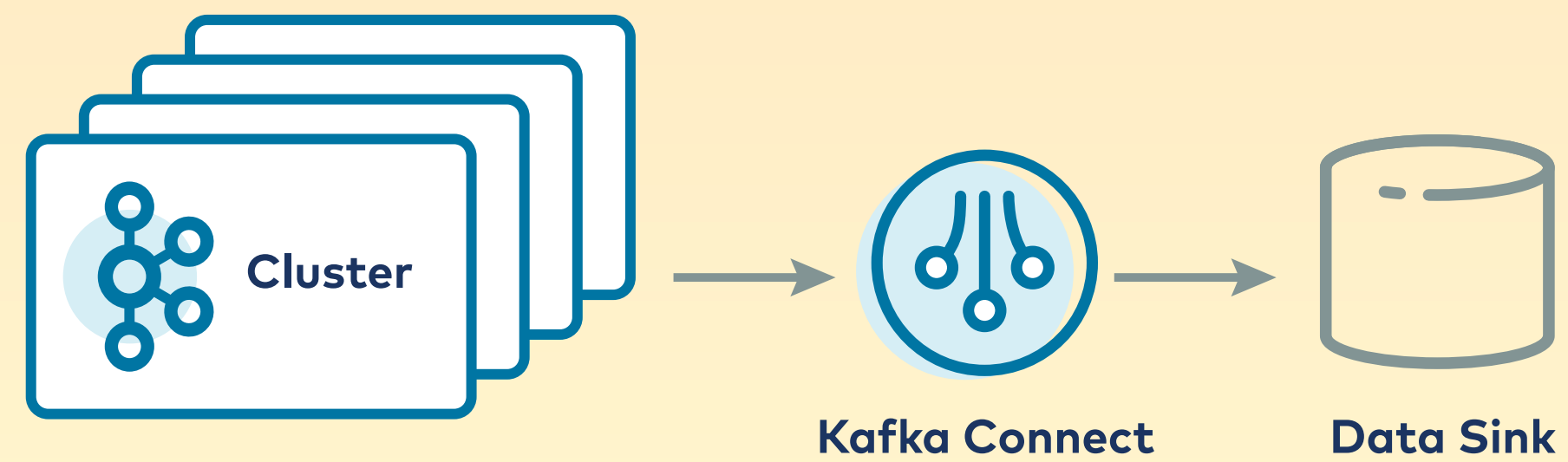
OLTP stream vs OLAP streams

- 
- A pixel-art style illustration of a person standing in a workshop or office, presenting a list of data technologies on a large screen. The person is seen from behind, wearing a white shirt and dark pants. The screen displays a list of six items, each preceded by a bullet point. The background is a detailed, colorful scene with various pieces of equipment, shelves, and a desk with a computer monitor. The lighting is warm and focused on the screen and the person.
- Connect/Relational DB
 - Kafka Streams
 - Streaming SQL
 - Cloud Data Warehouse
 - Data Lake
 - Real-Time OLAP Database

Kafka Connect

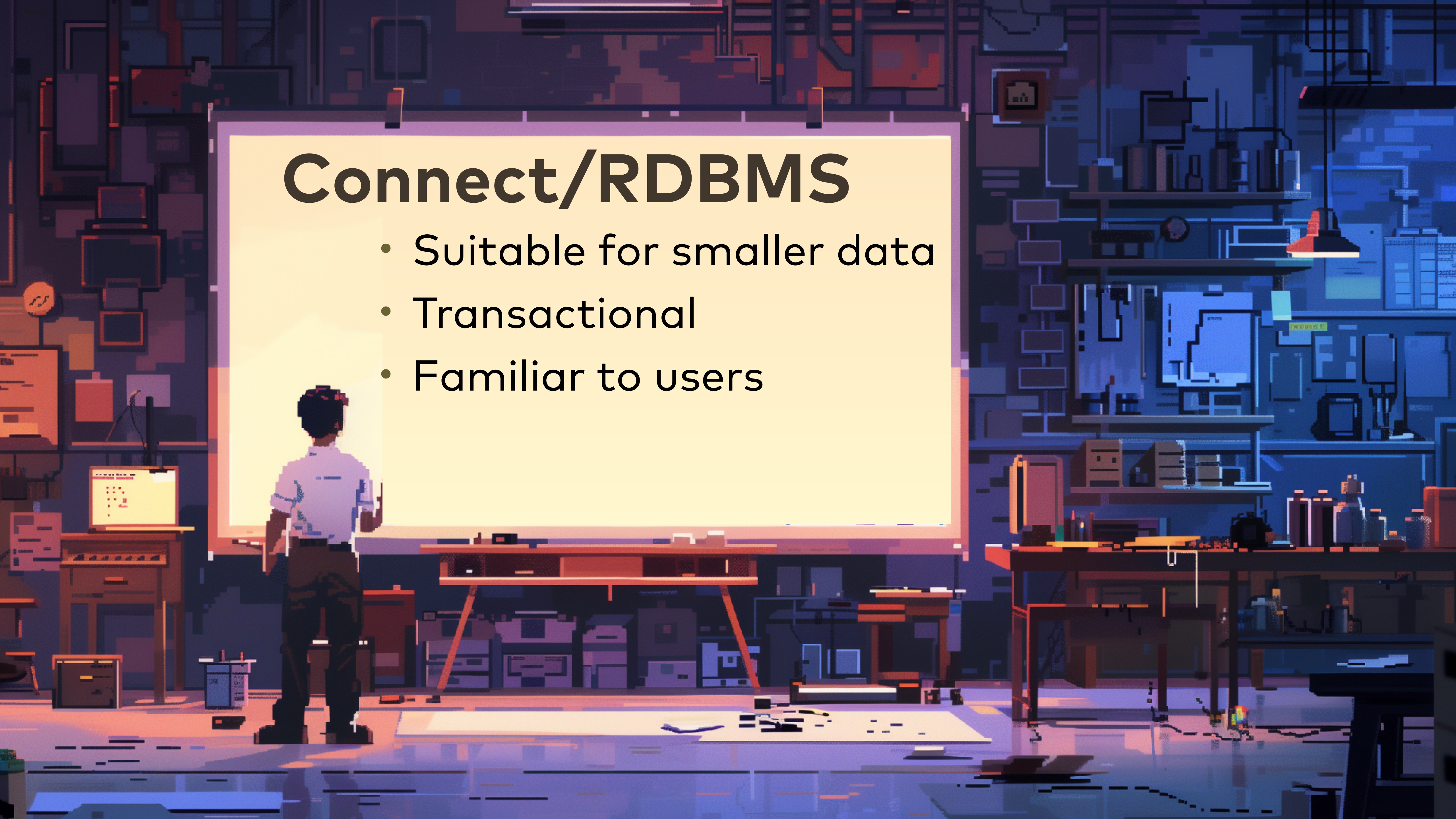


Connect/RDBMS



Connect/RDBMS

- Suitable for smaller data
- Transactional
- Familiar to users





Kafka Streams

Kafka Streams

(transactional)

- Ingests directly from a **topic**
- KTable
- Forms an **in-memory key/value** store suitable for querying by topic key
- Scalable across **members** of a **consumer group**
- Readable through **Interactive Queries**

Kafka Streams

(transactional)

```
final KStream<String, String> stream =  
    builder.stream(inputTopic,  
        Consumed.with(stringSerde, stringSerde));
```

```
final KTable<String, String> convertedTable =  
    stream.toTable(Materialized.as("stream-  
converted-to-table"));
```


Kafka Streams

(analytical)

- Full-featured **Java stream processing API**
- Arbitrary streaming computation
- Can emit new streams (not this talk)
- KTables **queryable by key**
- Every read pattern requires its own topology
- **Interactive Queries** again

Kafka Streams

(analytical)

```
KTable<String, Long> wordCounts = textLines
    .flatMapValues(textLine ->

Arrays.asList(textLine.toLowerCase().split("\\W+")))
    .groupBy((key, word) -> word)
    .count(Materialized.<String, Long, KeyValueStore<Bytes,
byte[]>>as("counts-store"));

wordCounts.toStream().to("WordsWithCountsTopic",
    Produced.with(Serdes.String(),
Serdes.Long()));
```


Streaming SQLs



Streaming SQL

- Materialize
- DeltaStream
- RisingWave
- ksqlDB



I love you, little
squirrel.



But this talk is
not about you.



Materialize

- Replacement data warehouse
- Integrates with Kafka, Postgres, dbt
- The **Materialized View** is the central abstraction
- Views are **persistent and queryable**
- Postgres wire-compatible
- Positioned as an analytics solution

Delta Stream

- Cloud-native **streaming SQL**
- **Serverless, BYOC**
- Kafka, Kinesis integration
- **Materialized views** and streaming pipelines
- streaming database and streaming analytics

Rising Wave

- Distributed SQL Streaming database
- Cloud and OSS versions
- Implementation of **Flink in Rust**
- **Kafka, Pulsar, Kinesis** integrations
- Flink+persistent views
- Postgres wire-compatible

ksqlDB

- «*Streaming Database*»
- Provides persistent TABLE abstraction
- Pull and Push queries
- Like **KafkaStreams**, but in **SQL**



Cloud Data Warehouses

Cloud Data Warehouses

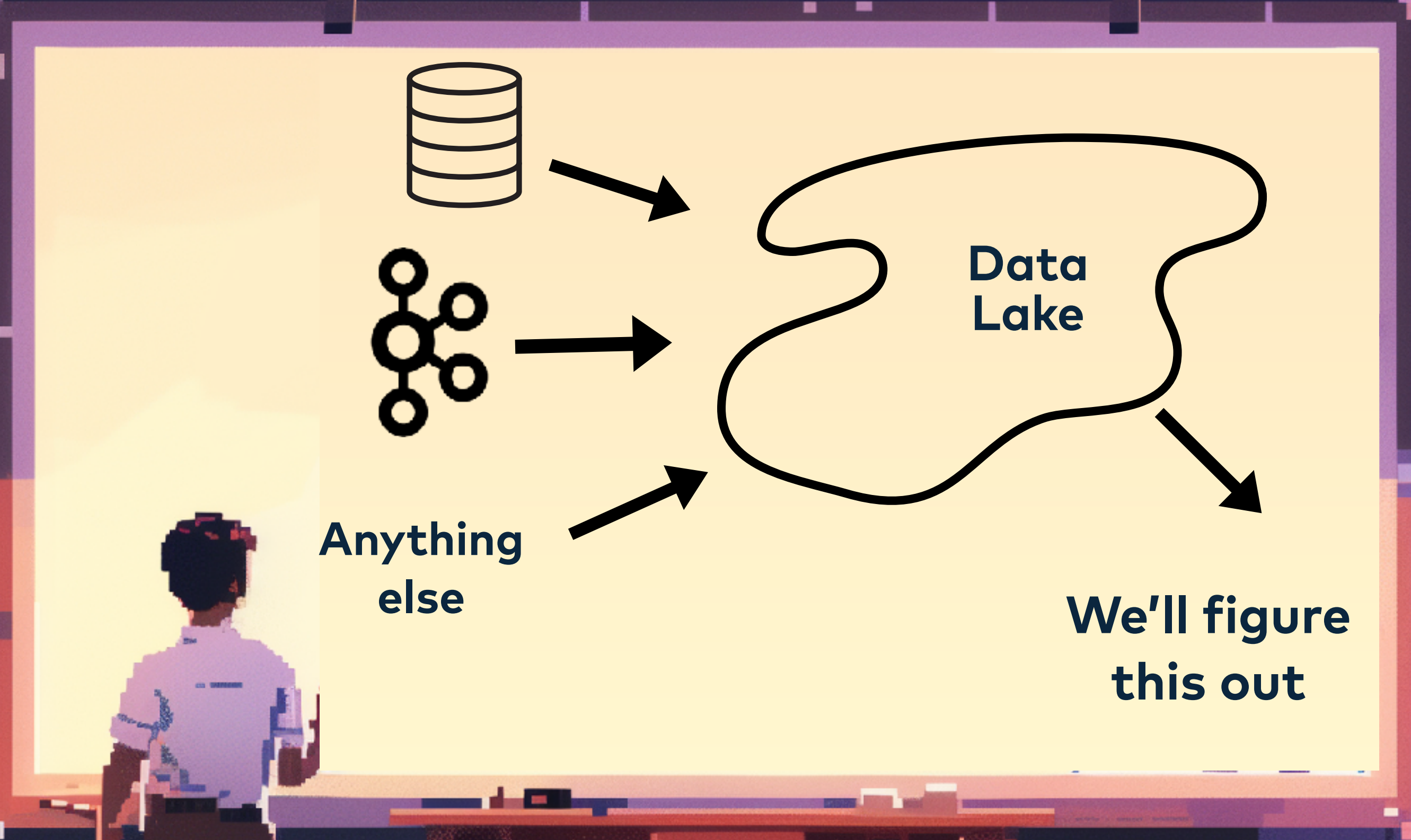


Cloud Data Warehouses

- The **cloud-based** heir of legacy DWH
- Ingest from **batch** and **streaming sources**
- Biased towards **structured data** and **batch access**

Data Lake





Data Lakes

- Started as the HDFS cluster
- Became S3
- That didn't help...
- ELT vs. ETL
- Iceberg/Hudi/DeltaLake

Data Lakes

- Storage and compute are radically decoupled
- Structure is relatively less important
- Reads are slow
- Streaming is historically difficult



Data Lakes

- Storage and compute are radically decoupled
- Structure is relatively less important
- Reads are slow
- Streaming is historically difficult



Real-Time Analytics Database

Real-Time OLAP

- Designed for **high concurrency**, **low latency** queries
- Ingests from **streaming** and **batch** sources
- Intimate **integration with Kafka**
- Conventional **tables and SQL**

Real-Time OLAP

- Analytics shaped like **real-time data**
- Analytics when **users are decision makers**





**No Solutions
only Trade Offs**



Sometimes you go
with what you know

A vibrant, stylized illustration of a man with dark hair and a mustache, wearing a yellow shirt and blue pants, sitting in a large, patterned armchair. He is smiling and reading a red book with a sun-like symbol on the cover. The background features a large window with a bright, stylized sun illustration and a bookshelf filled with books. A large green plant is visible on the left side. The overall scene is bathed in warm, golden light.

This is not bad!

A detailed steampunk illustration of a laboratory. In the center, a scientist with a large, spiky white hairpiece is seated at a desk, focused on a large, glowing blue circular device. The desk is cluttered with various scientific instruments, including a microscope, a large glass bottle, and several smaller containers. The background is filled with intricate mechanical components, including gears, pipes, and a large, ornate circular dial. Several human skulls are mounted on the wall, adding a macabre touch to the scene. The lighting is a mix of warm, golden-yellow tones from the desk and cooler, blue-green tones from the background and the scientist's hairpiece. The overall atmosphere is one of intense, somewhat dark, scientific inquiry.

Performance

Community



Differentiated Application Code

The background is a complex illustration of a fantastical world. It features several dragons of various colors (green, brown, blue) and sizes, some breathing fire. A sailing ship is visible in the upper middle. In the bottom right, there's a depiction of a city or settlement. The entire scene is overlaid with a grid pattern, and the edges of the illustration are jagged, resembling torn paper.

Area of
Exploration

Kafka

Need your feedback



Scan me

Need your feedback



It's Anonymous

Scan me

A stylized illustration of a vineyard. A road with yellow dashed lines runs through the center, flanked by rows of green grapevines. On the left, a large bunch of dark red grapes hangs from a vine. On the right, a red rectangular sign with white text reads "NO SPEED LIMIT". The sky is blue with white and yellow clouds. The overall style is a vibrant, pixelated or halftone aesthetic.

For more resources
on Apache Pinot:

dev.startree.ai

Viktor Gamov, StarTree
@gamussa