Building ETL Pipelines from Streaming Data with Kafka and KsqlDB

UNDERSTANDING KSQLDB AND KAFKA STREAMS

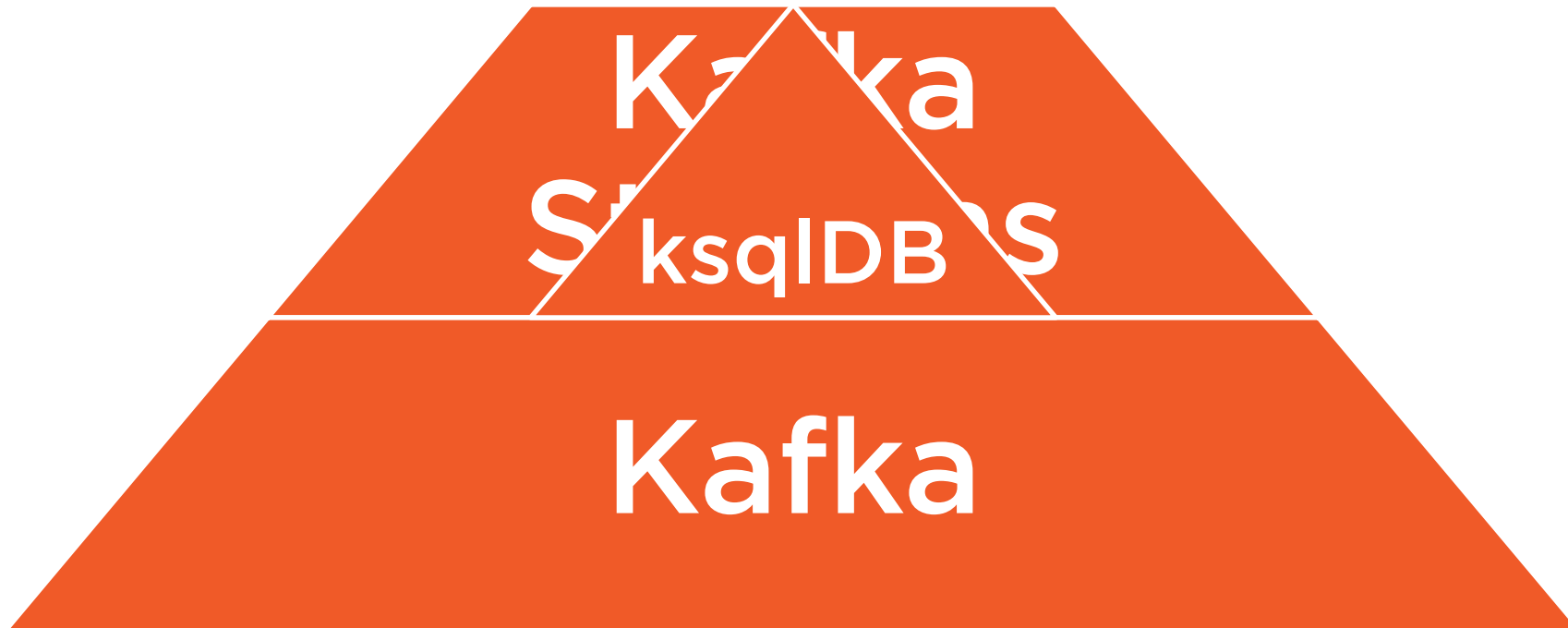**Eugene Meidinger**

BI CONSULTANT

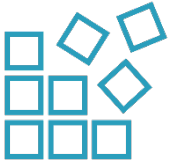@sqlgene   www.sqlgene.com

Overview

Topics

Deployment

Aggregations

Enrichment

Challenges

Kafka wasn't designed for stream processing.

Kafka is an event streaming platform

Apache Software Foundation

Kafka is an event streaming platform

Kafka is an **event streaming** platform

Kafka is an event streaming platform

Benefits of Kafka
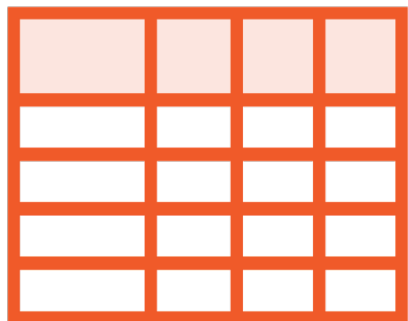
Decoupling

Scaling

Space

Time

What Is Kafka Streams?

Kafka Streams is a library that lets you speak in streams, not events.

What Does Kafka Streams Provide?
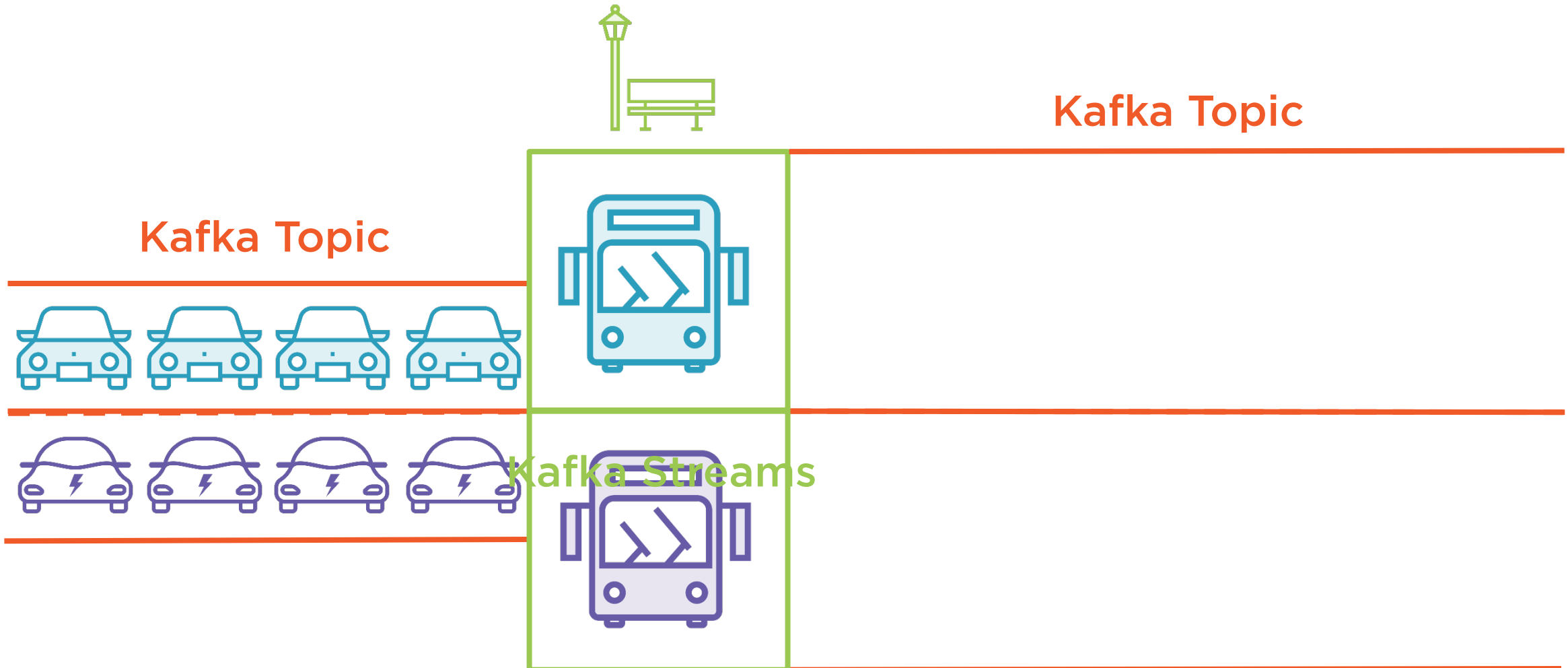
Nouns

Verbs

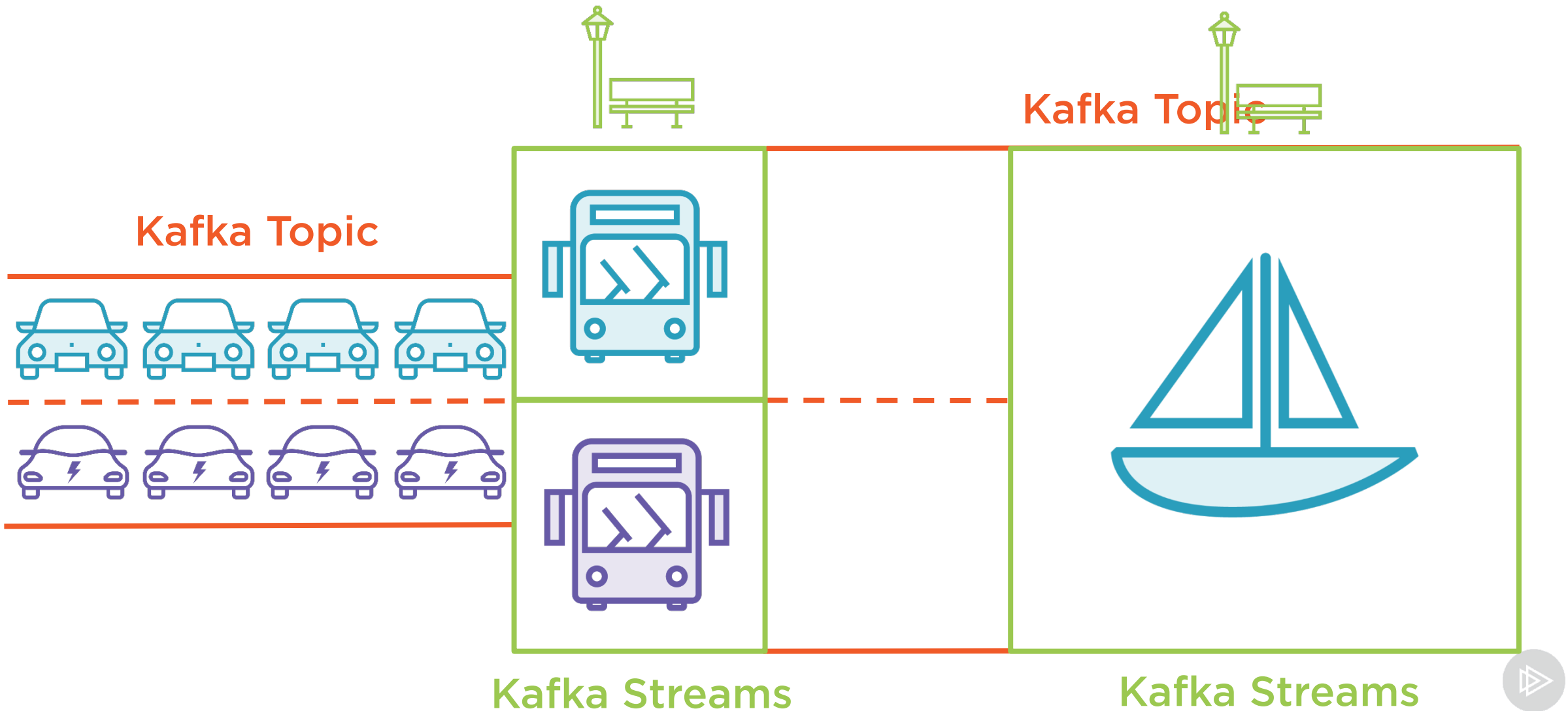**Scalability**

Reliability

Aggregating Data

Kafka Topic

Aggregating Data

Kafka Topic

Kafka Topic

Kafka Streams

Application Instances

Kafka Topic

Kafka Topic

Kafka Streams

Chaining Processors

Processor Topology

Streams and Tables

Streams and Tables

Stream

0	0	+5	+2	+1	-2	+1

?
4
?
3
?
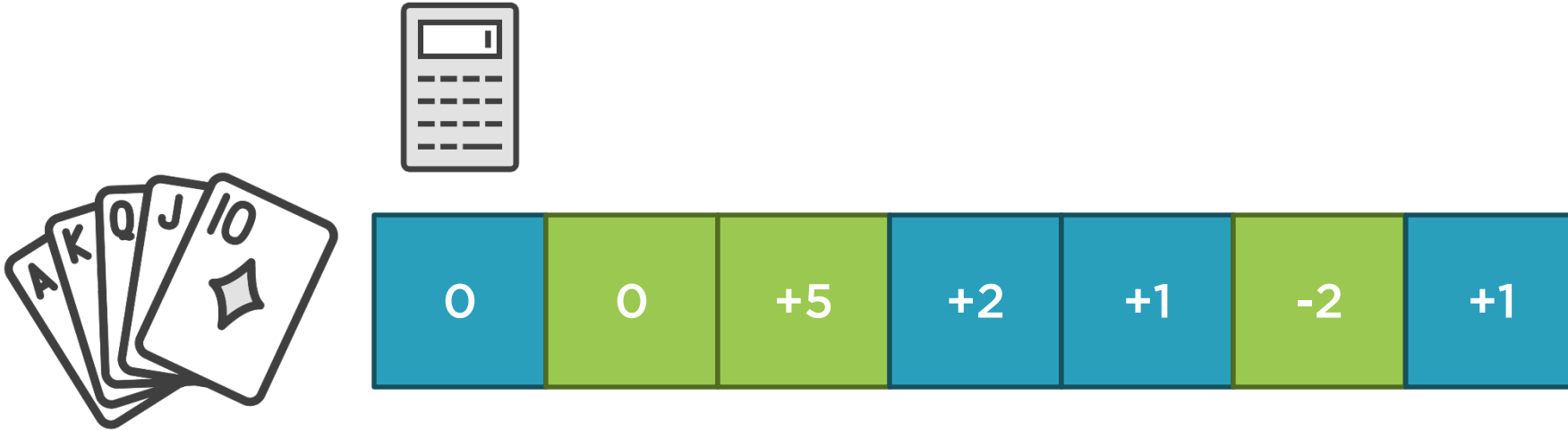
Table
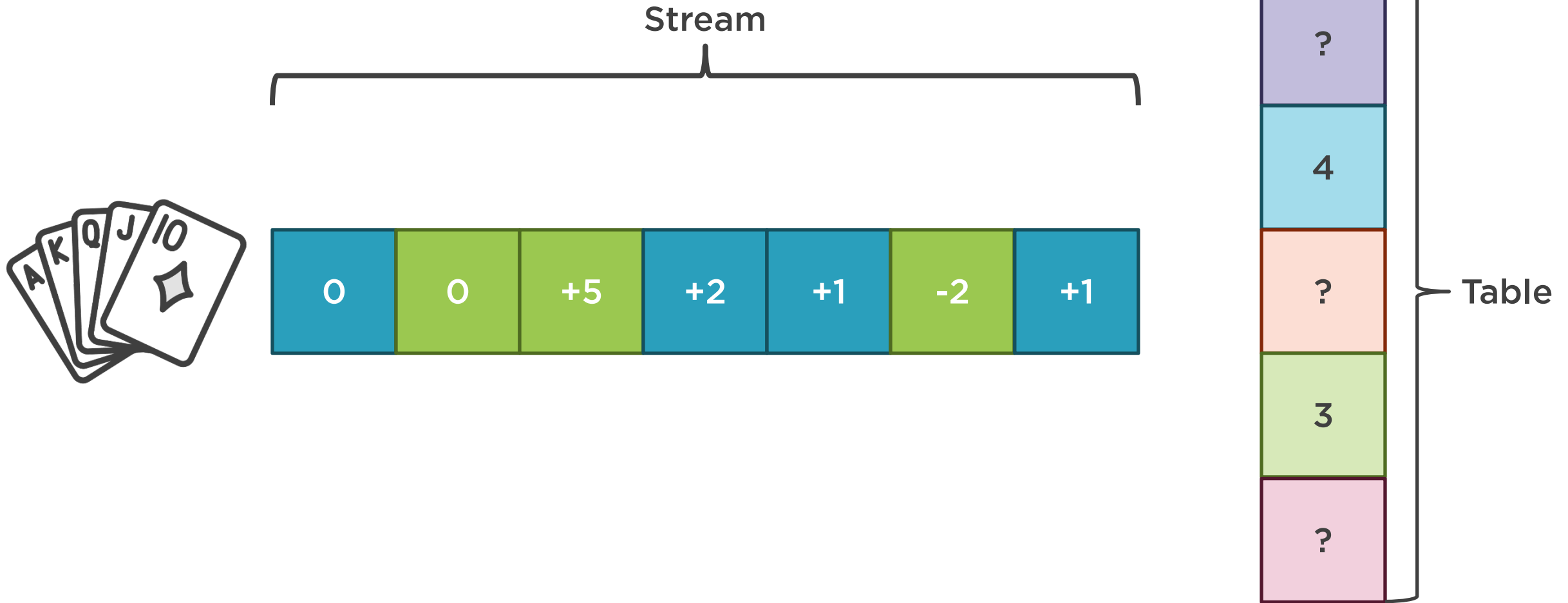
What is KsqlDB?

"ksqlDB is an event streaming database purpose-built to help developers create stream processing applications on top of Apache Kafka."

Confluent

"ksqlDB is an event streaming database [...] on top of Apache Kafka."
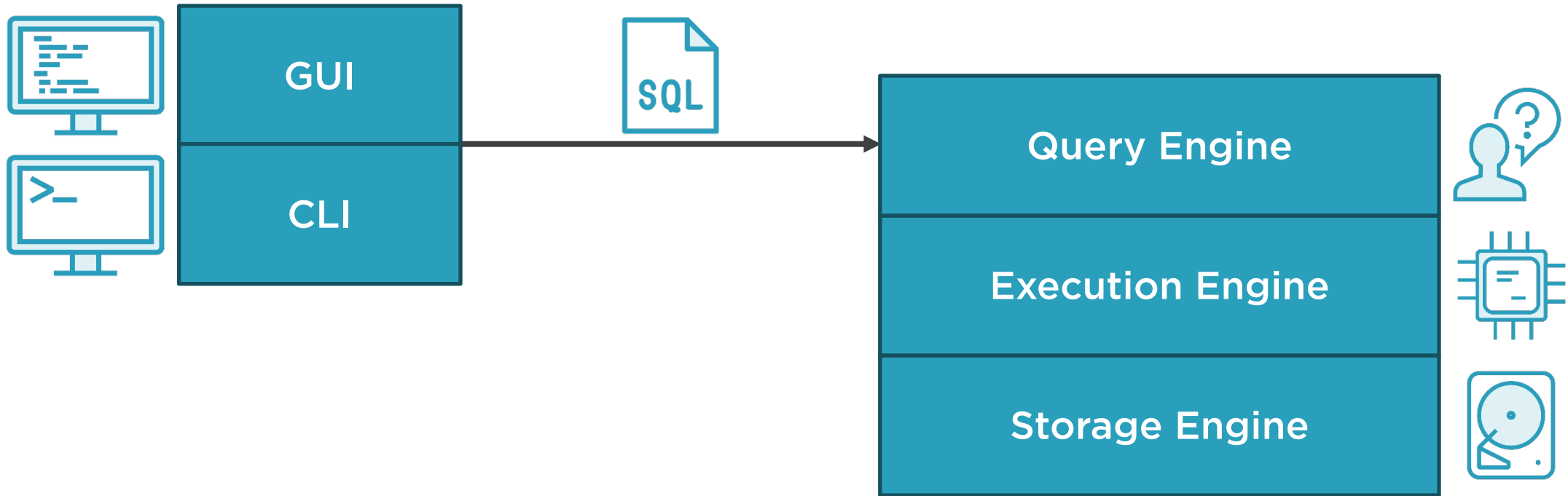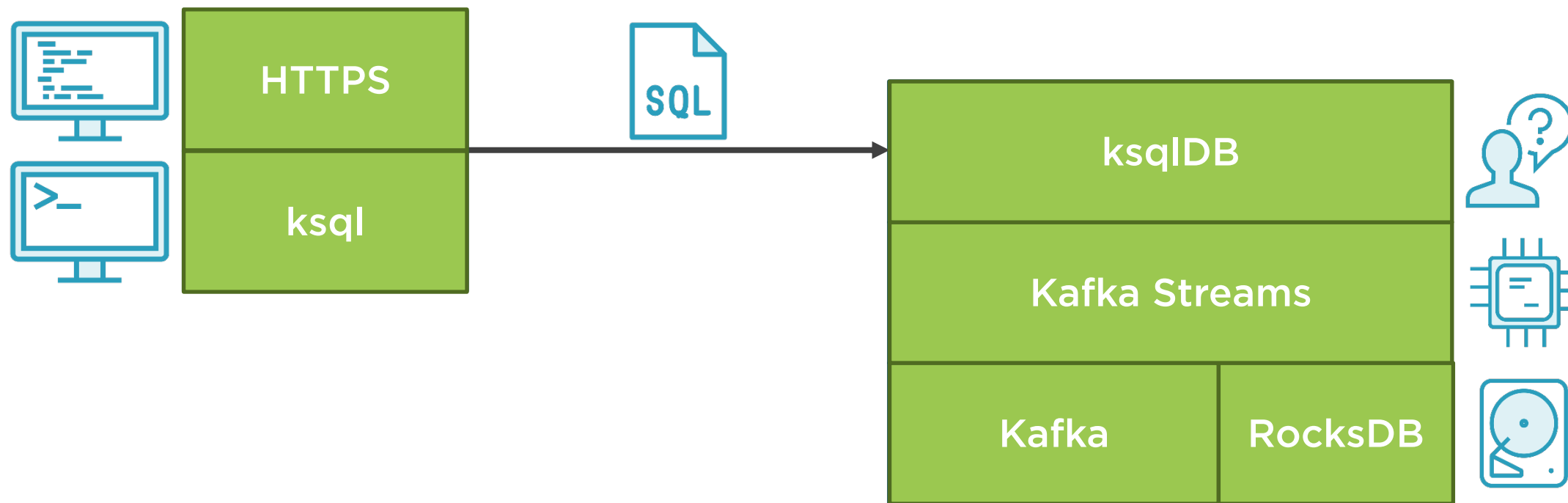
Confluent

What is MYSQL?
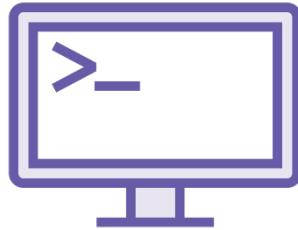
What is Postgres?

What Is a Database?

What Is a Database?

GUI

CLI

SQL

Query Engine

Execution Engine

Storage Engine

What Is a Database?

HTTPS

ksql

SQL

ksqlDB

Kafka Streams

Kafka

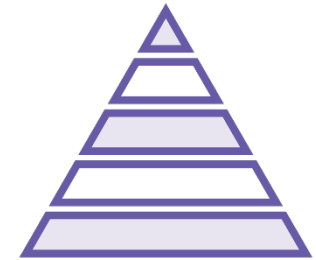RocksDB

What Is KsqlDB?

SQL Variant

Command Line Interface

Web Server

Top Layer

Overview

Kafka stores events
- Events are small historical records

Kakfa Streams simplifies ETL
- Uses streams and tables
- Manages scaling out

ksqlDB gives us an SQL interface
- No compiling, less work