

Enforcing Data Contracts with Kafka Schema Registry

SERIALIZING DATA IN MESSAGING SYSTEMS



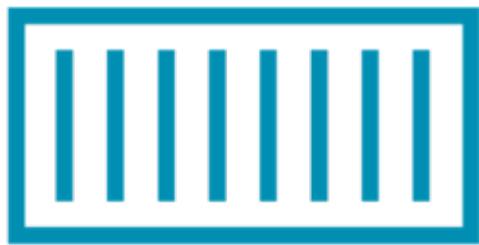
Bogdan Sucaciu

SOFTWARE ENGINEER

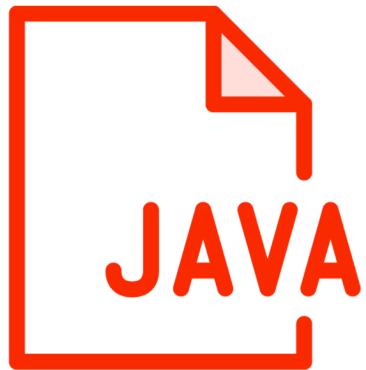
@BSucaciu bsucaciu.com



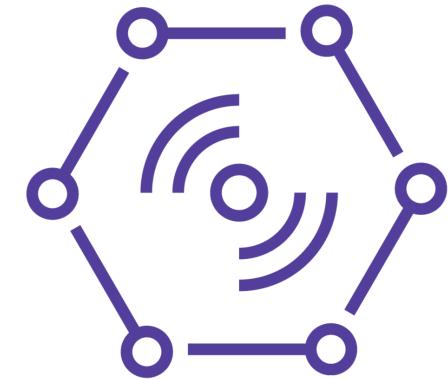
Prerequisites



Docker
Starting and stopping
containers; Docker
Compose



Java 8+
Understand Syntax



Apache Kafka
Build Producers and
Consumers; Topics,
Brokers, Kafka
Connect



Pluralsight Courses

Getting Started with Apache Kafka's Architecture

Getting Started with Apache Kafka

Ryan Plant COURSE AUTHOR

by Ryan Plant

This course will introduce you to Apache Kafka and provide a thorough tour of its architecture so you can start building your next enterprise system with it.

How Apache Kafka Starts to Differentiate

Producers Consumers

LinkedIn: 1,400 brokers => 2 petabytes per week

APP Broker APP

Consuming from Kafka

Handling Streaming Data with a Kafka Cluster

Step 1: Establish TCP Connection

broker-1 services-0 broker-2 services-0 broker-3 services-0

by Bogdan Sucaciuc

This course will teach you how to handle different scenarios that are commonly encountered when using a streaming platform inside your organization.

Modeling Data Streams

Admin Client Data Governance Create Topics Schemas Naming

Kafka Connect Fundamentals

Kafka Connect Distributed

OS / Container Worker 1

Source Sink config-storage-topic

by Bogdan Sucaciuc

Kafka Connect Cluster

Admin Client

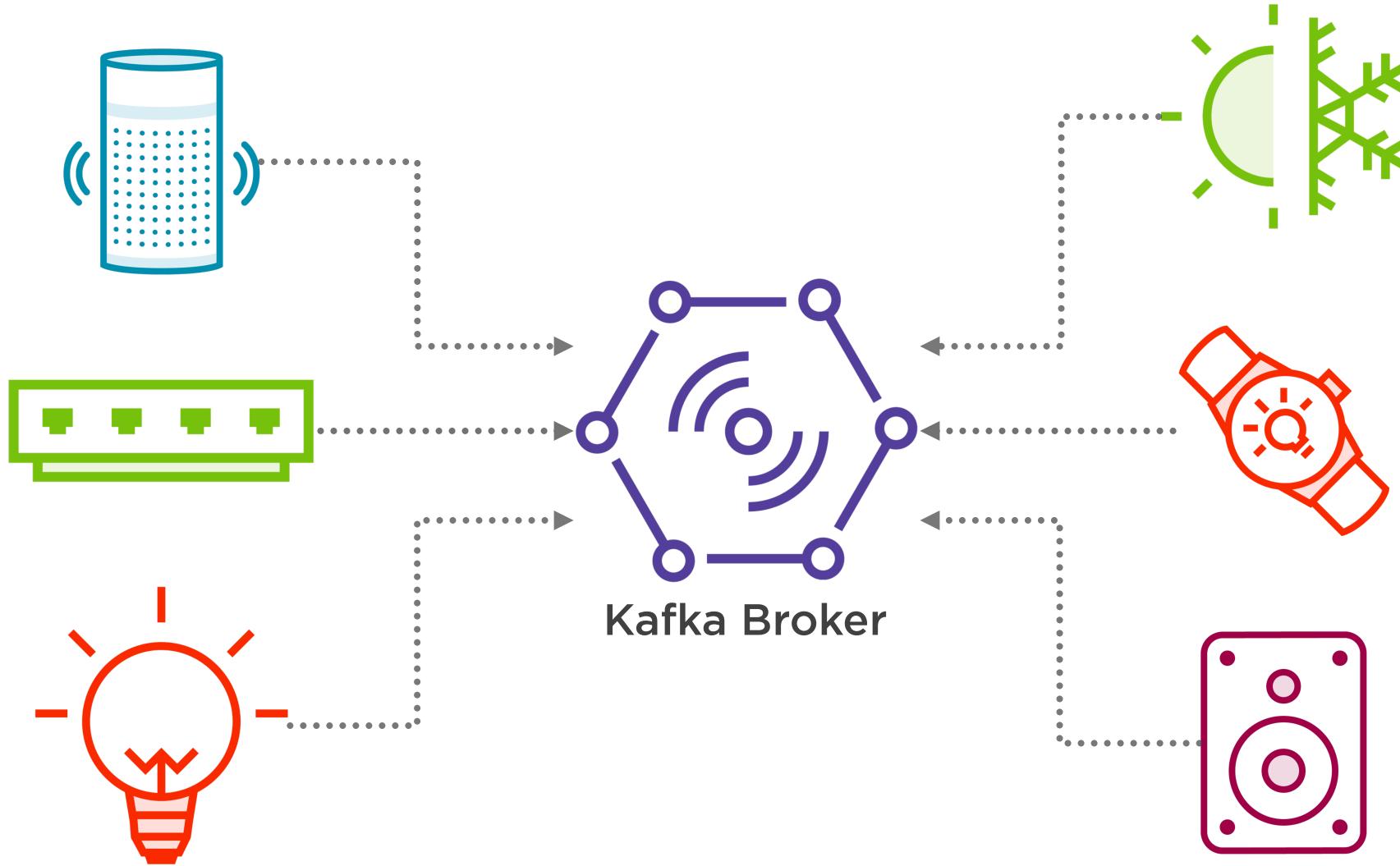
Kafka Producer Kafka Broker Kafka Consumer

This course is a technical introduction to Kafka Connect and how to use it to quickly transfer data in and out of the Apache Kafka ecosystem, to and from common data systems like databases, key-value stores, search indexes, and file systems.

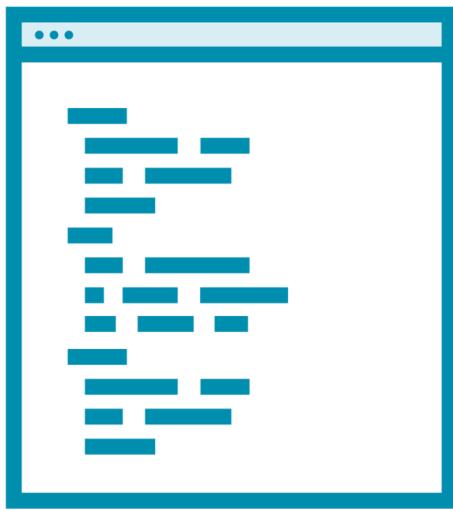
A smiling man in a blue shirt is visible on the right side of the thumbnail.



IoT Devices



Data Serialization



Software Application

Bare Metal



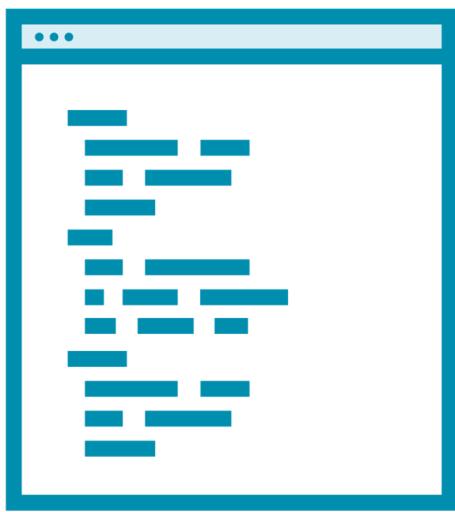
Machine

Virtual Machine

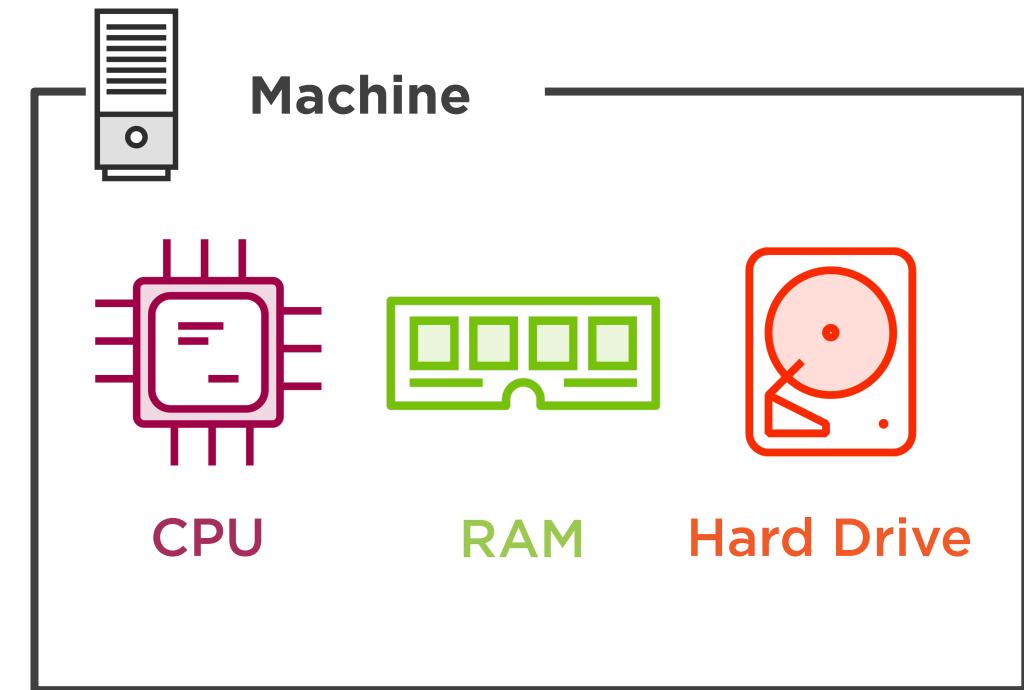
Container



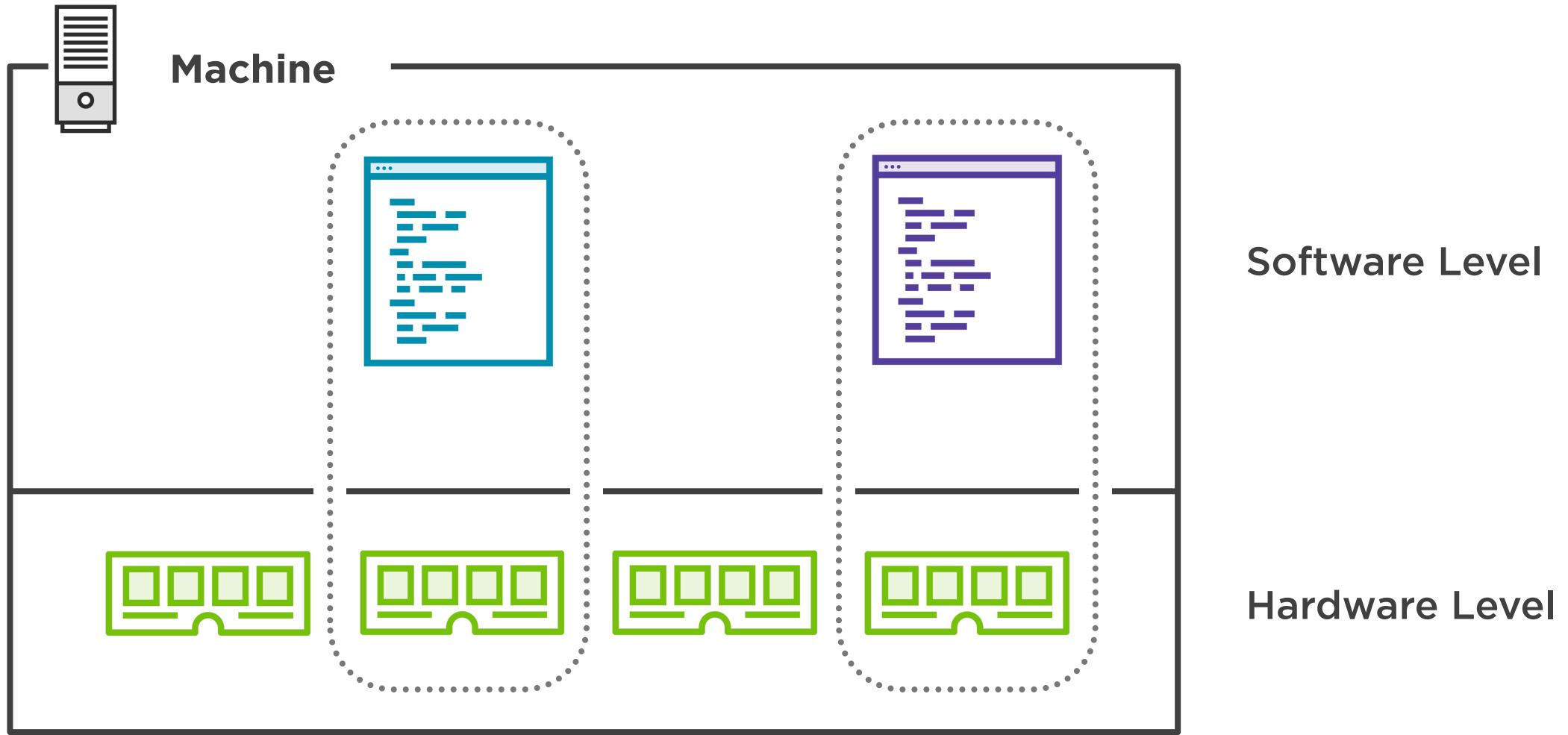
Data Serialization



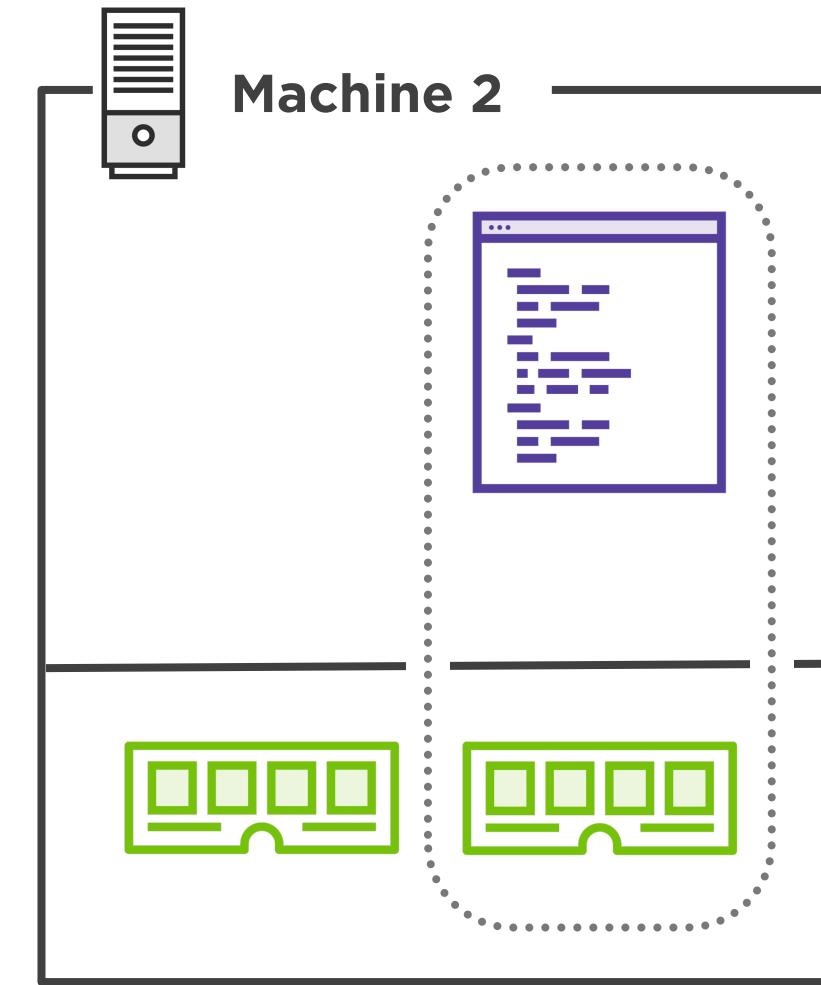
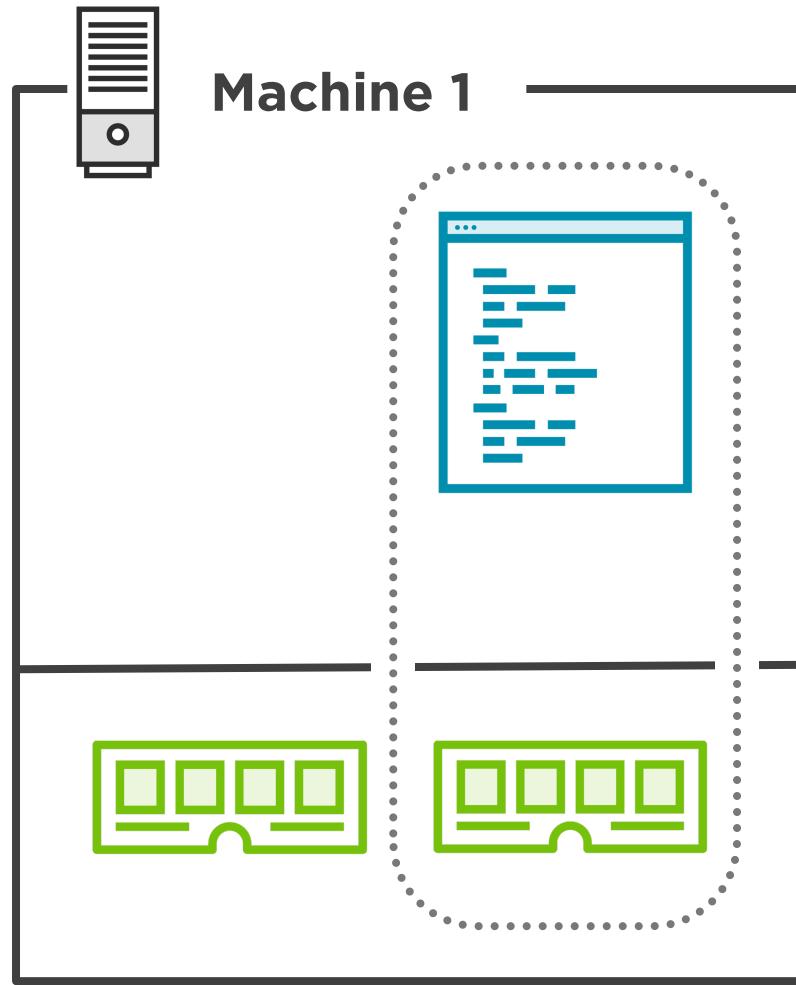
Software Application



Data Serialization



Data Serialization



Data Serialization



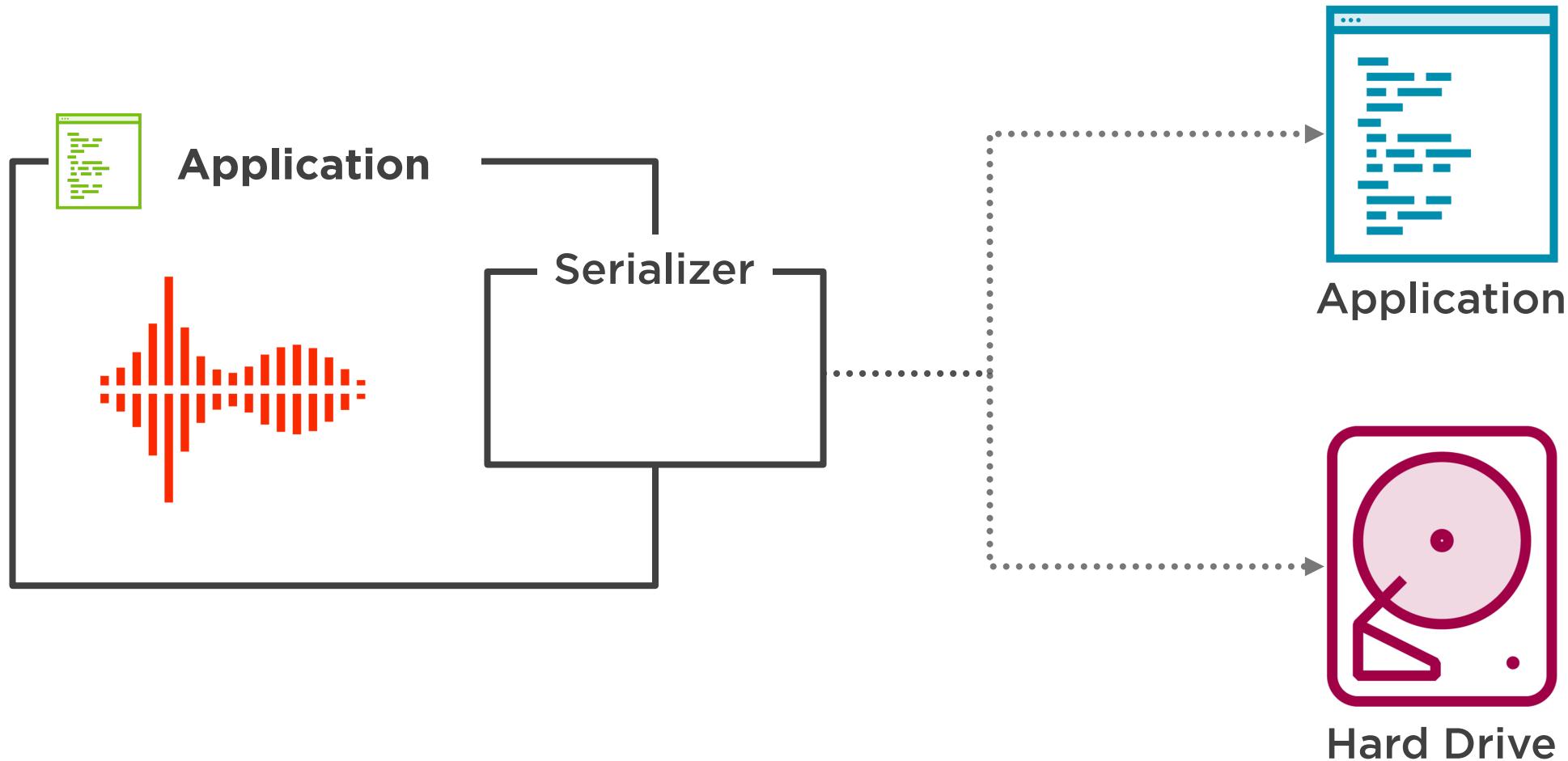
Serialization

The process of translating **data structures** or **object states** into a format that can be **stored** or **transmitted** and reconstructed later, possibly in a **different** computer environment.

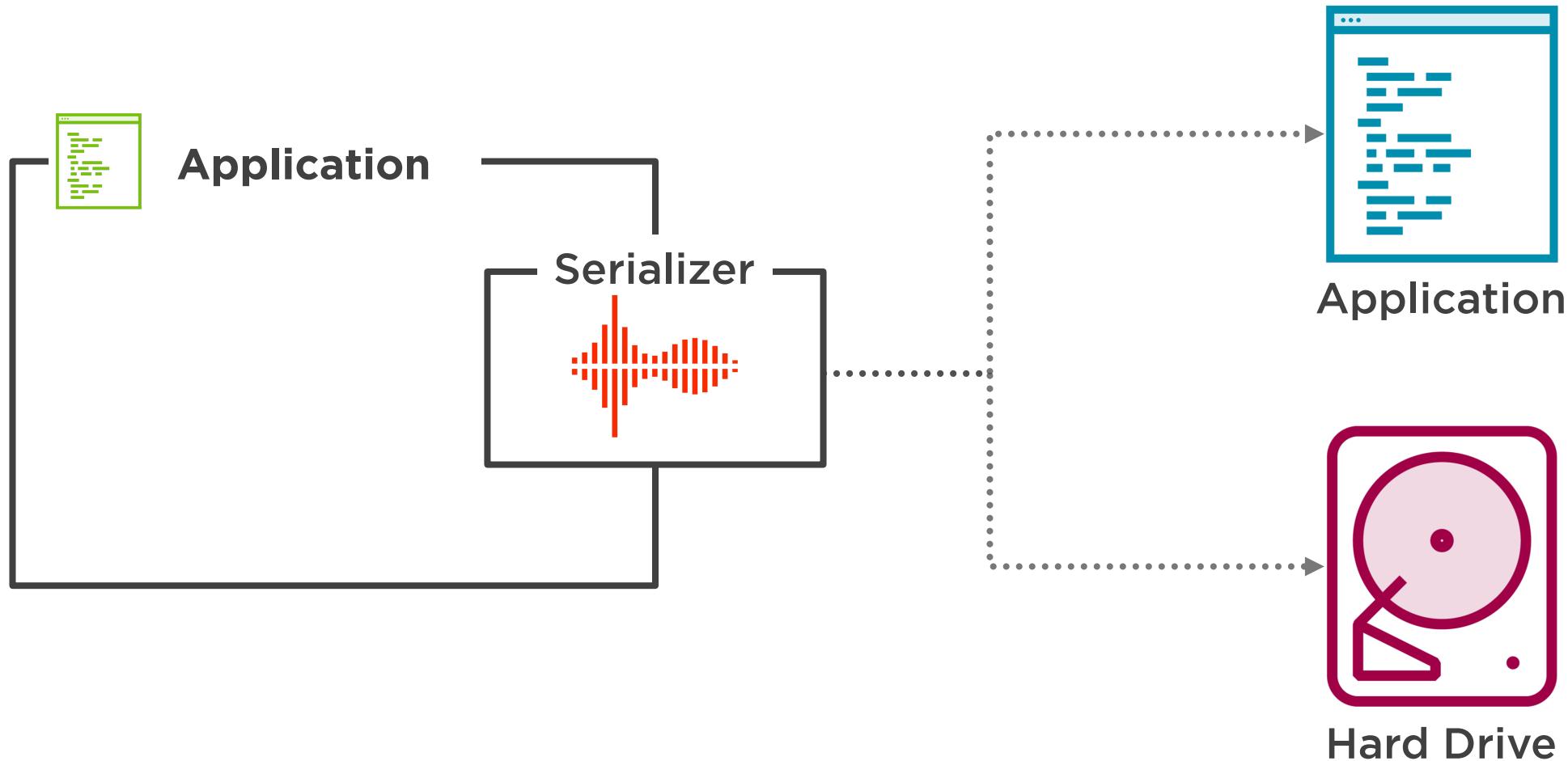
-Wikipedia



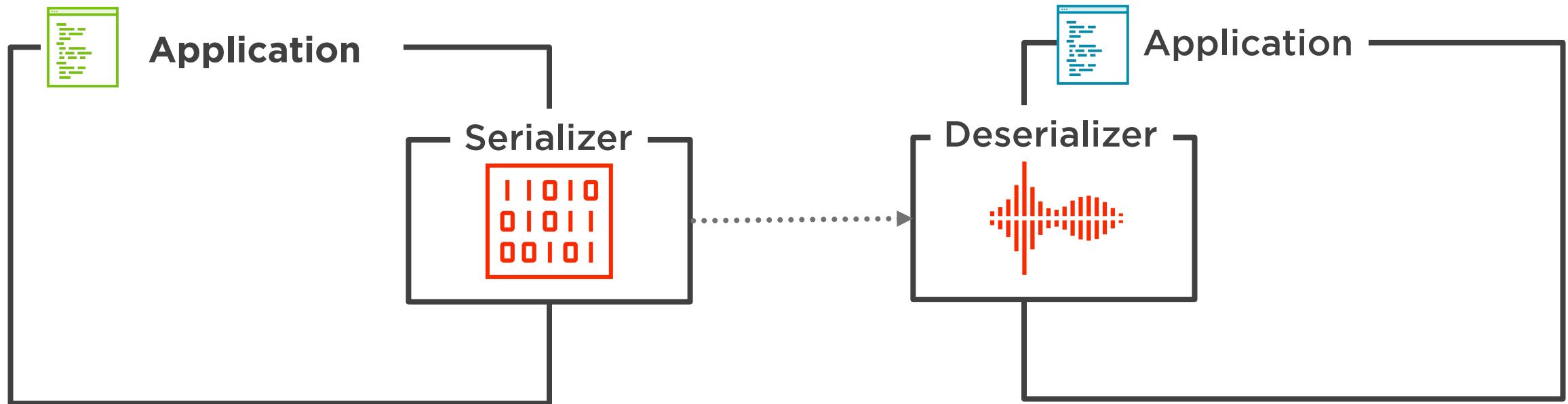
Data Serialization



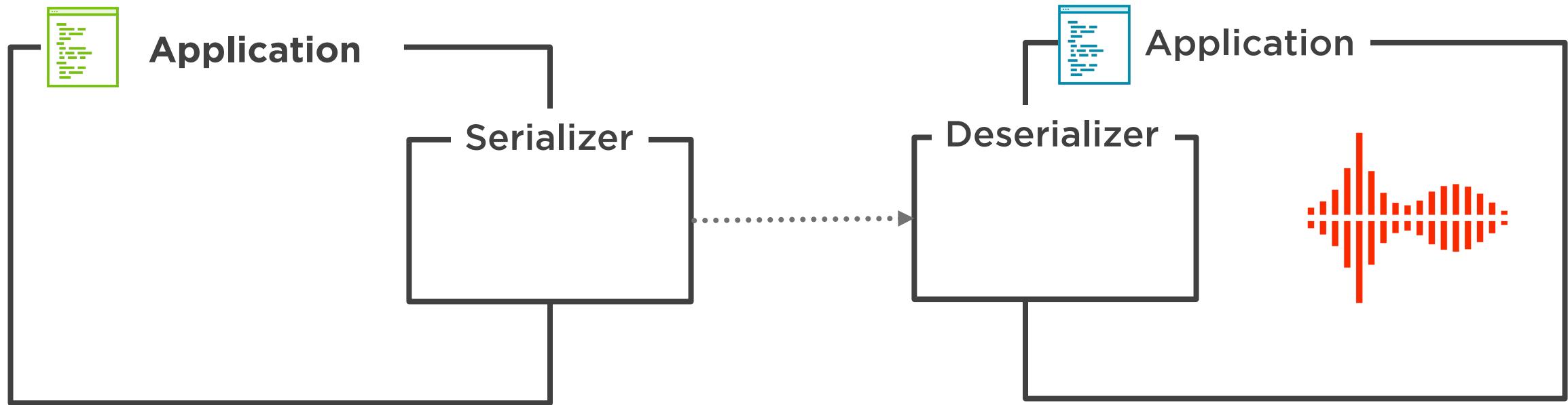
Data Serialization



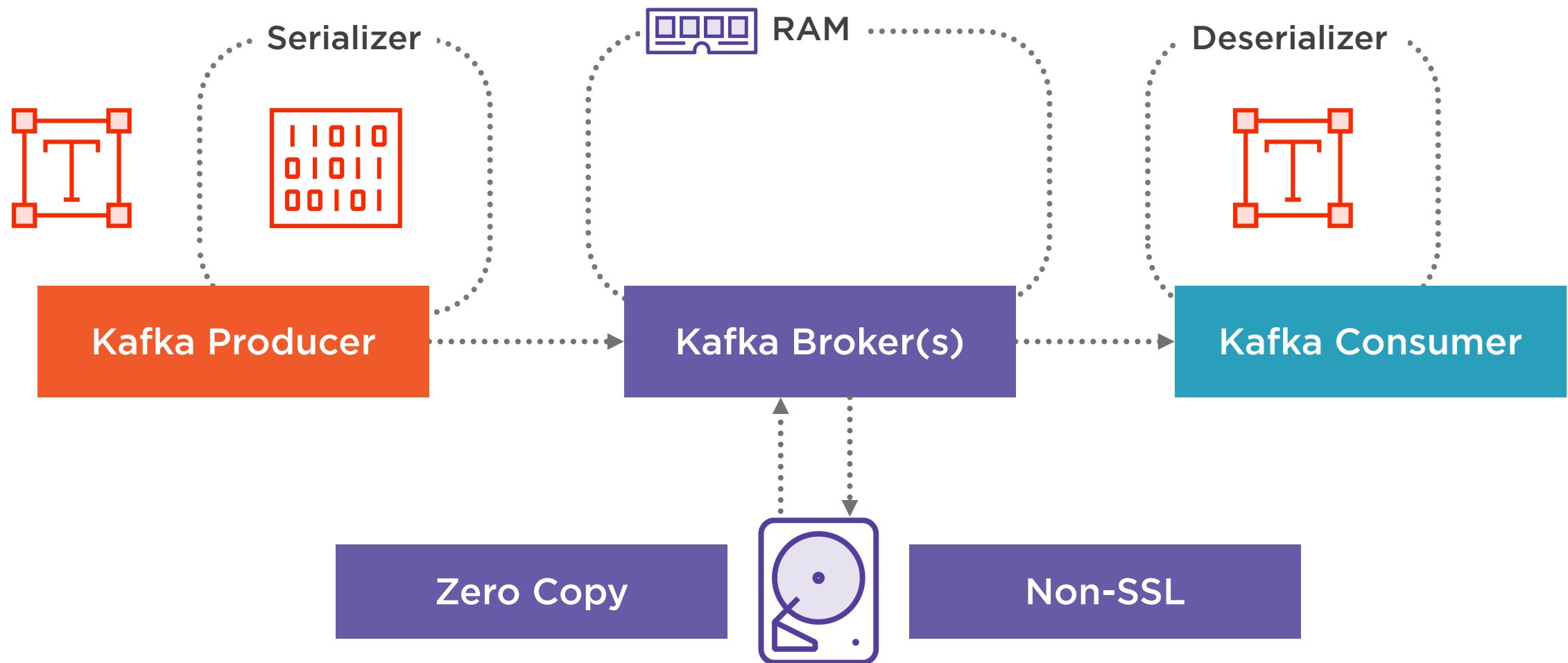
Data Serialization



Data Serialization



Data Serialization in Apache Kafka



How is data serialized?



It depends...



It depends...

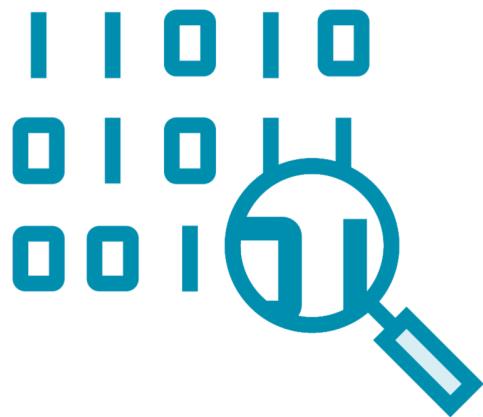
on the **Serialization Format!**



Serialization Formats



Serialization Formats



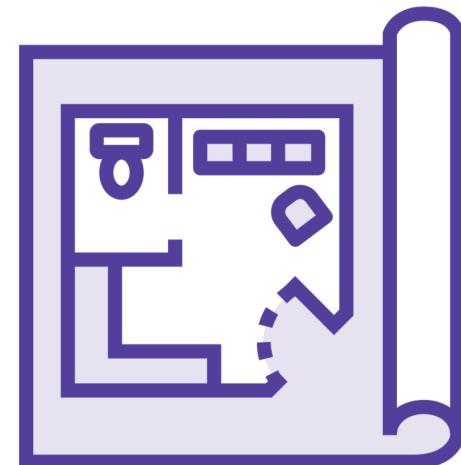
Binary?

Data is transmitted /
stored as bytes or
plaintext



Human-readable?

Can be read without
referring to an
encoding table?

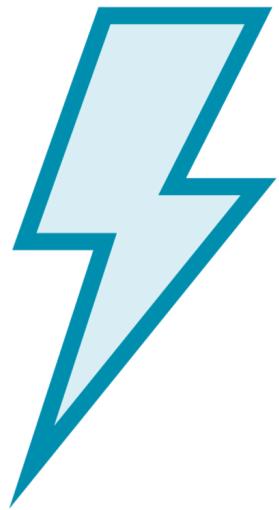


Schema?

Fixed data structure or
Flexible

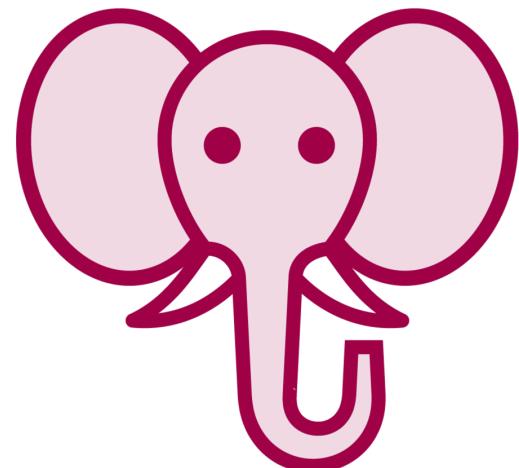


Serialization Formats



Speed

How fast can data be serialized?



Size

How much space will
serialized data take?



Serialization Formats

Name	Binary	Human-readable	Schema - IDL
JSON	✗	✓	✗ ✓
XML	✗	✓	✓
YAML	✗	✓	✗
AVRO	✓	✗	✓
Protobuf	✓	✗	✓
Thrift	✓	✗	✓



Serialization Formats with Apache Kafka

AVRO

- first class support using Schema Registry

JSON

- easy, no learning curve, extremely common

Protobuf

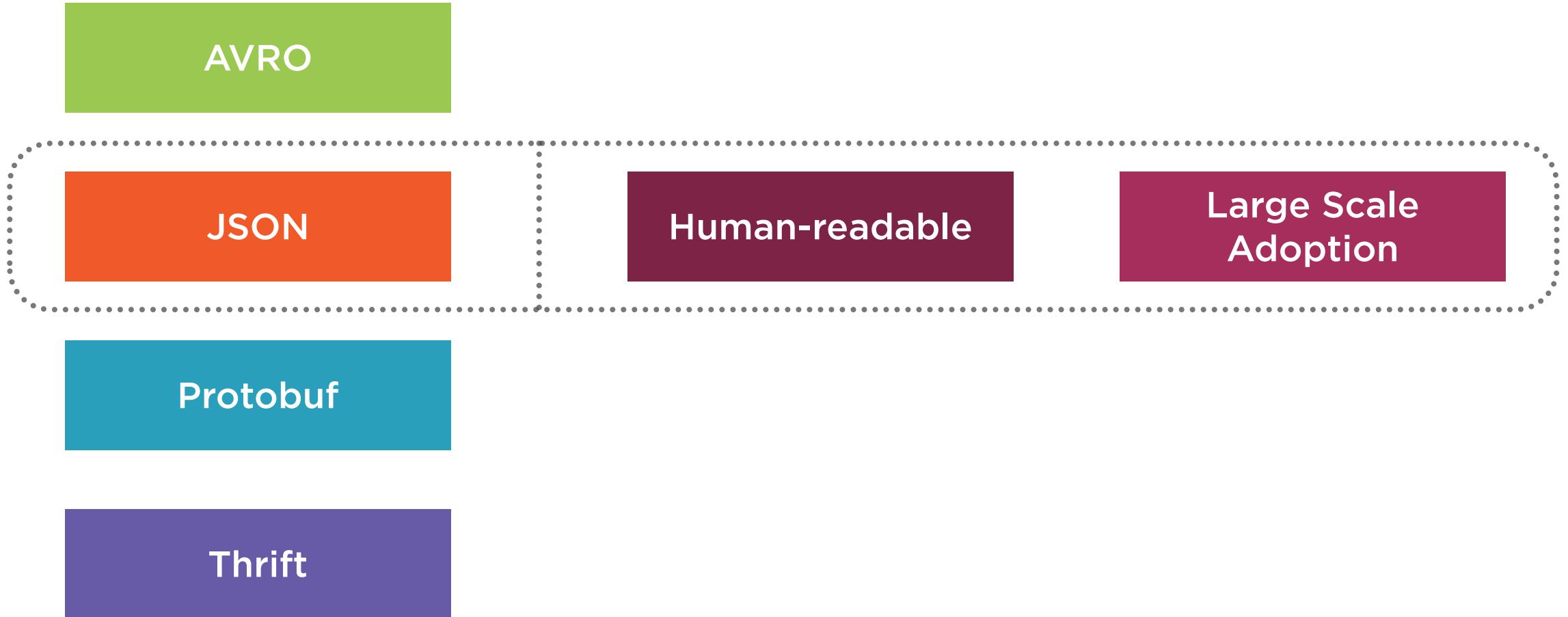
- highly efficient, gRPC

Thrift

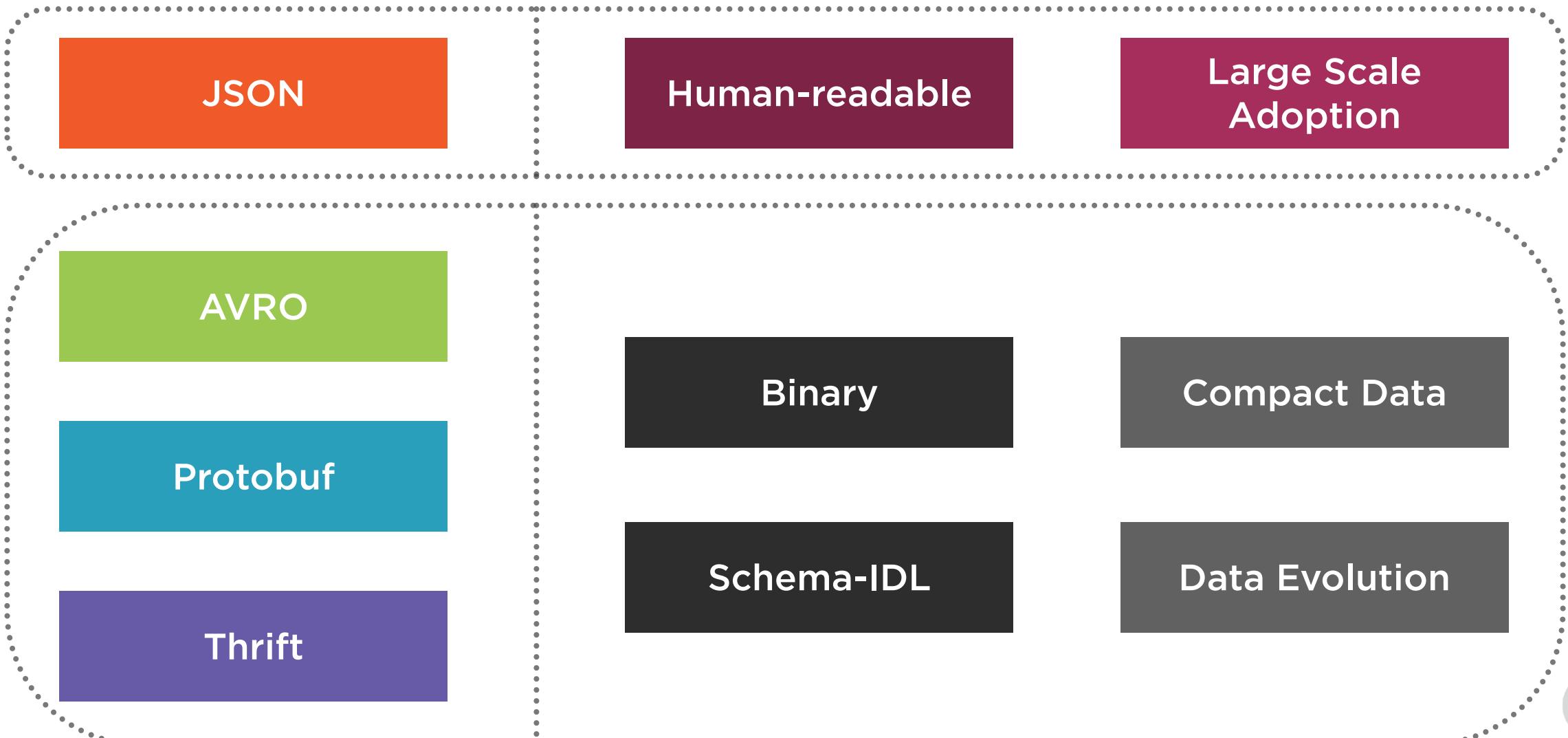
- Twitter Finagle, Apache Cassandra



Serialization Formats with Apache Kafka



Serialization Formats with Apache Kafka



Schema-IDL

schema.json

```
{  
  "type" : "string"  
}
```

schema.avsc

```
{  
  "type" : "string",  
  "name" : "command"  
}
```

schema.proto

```
syntax = "proto3";  
  
message AudioCommand {  
  string command = 1;  
}
```

schema.thrift

```
struct AudioCommand {  
  1: string command  
}
```

Schema-IDL

Fixed Data Structure

Fine-grained Control



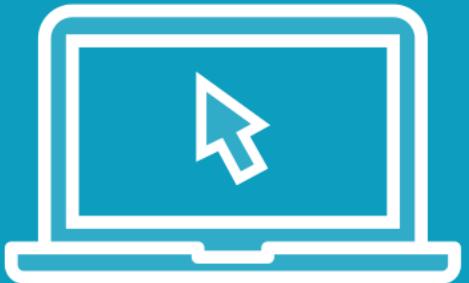
Data Contract

A formal **agreement** between a service and a client that abstractly **describes the data** to be exchanged.

- Microsoft



Demo



Install Apache Kafka

Kafka Console Producer

Kafka Console Consumer

Serializing and Deserializing Data

Prerequisites

- Docker



Summary



**Data serialization in Apache Kafka
Serialization Formats**

