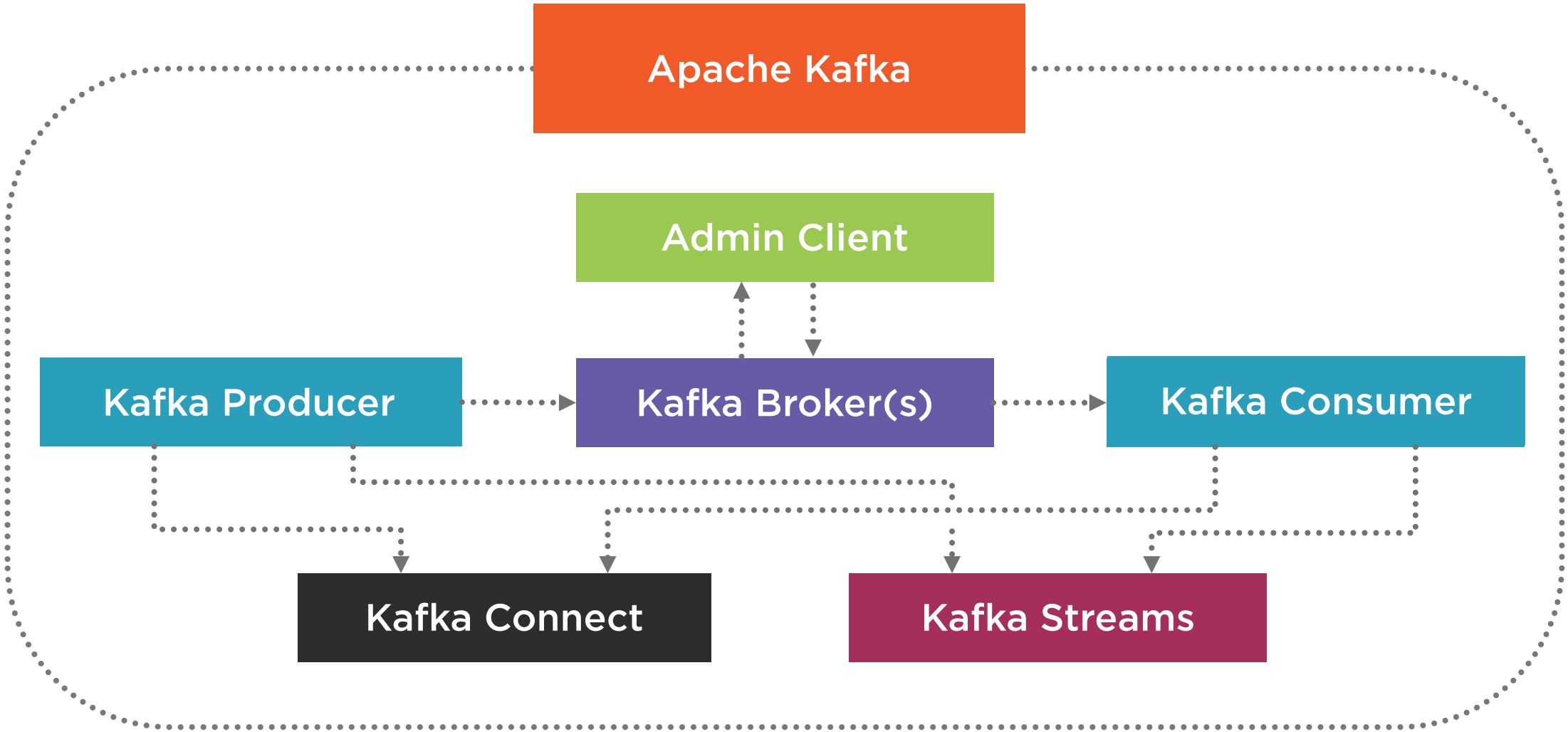
Managing Schemas

Bogdan Sucaciu
SOFTWARE ENGINEER

@BSucaciu bsucaciu.com


Apache Kafka
Admin Client
Kafka Producer
Kafka Broker(s)
Kafka Consumer
Kafka Connect
Kafka Streams

Schema Registry

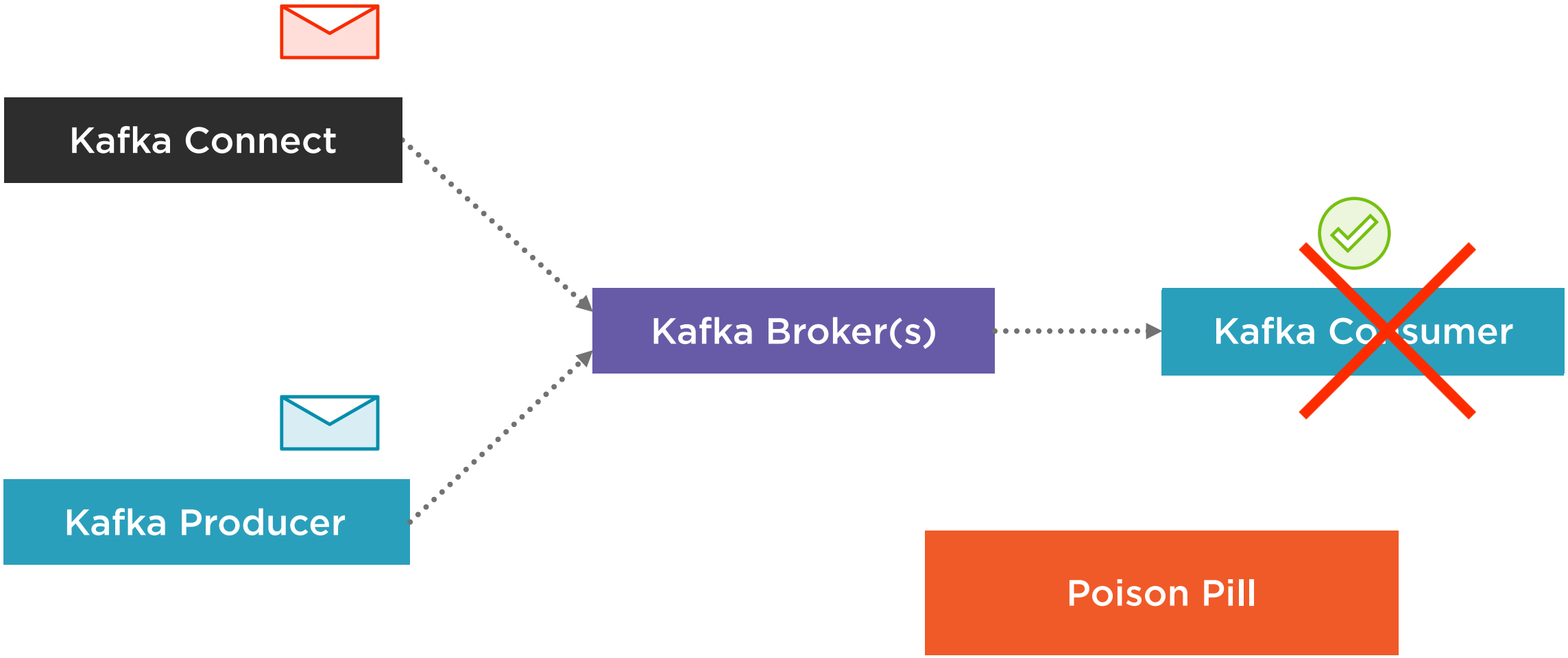
Enforcing Data Contracts

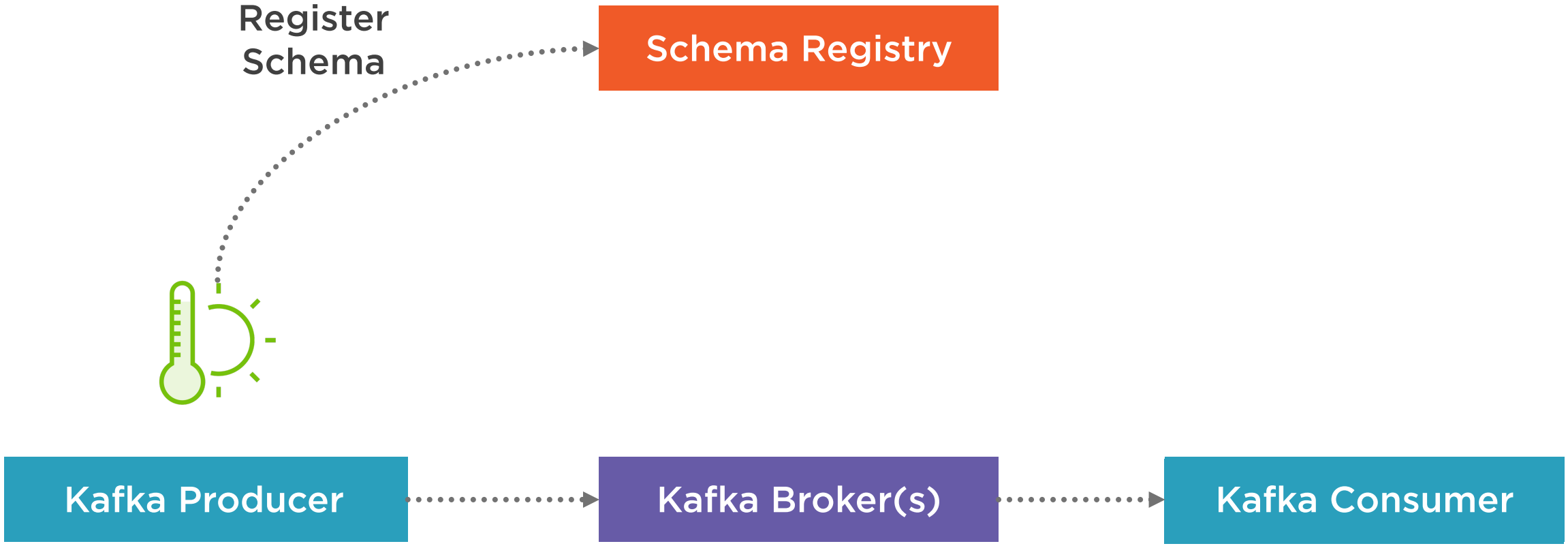
Handling Schema Evolution

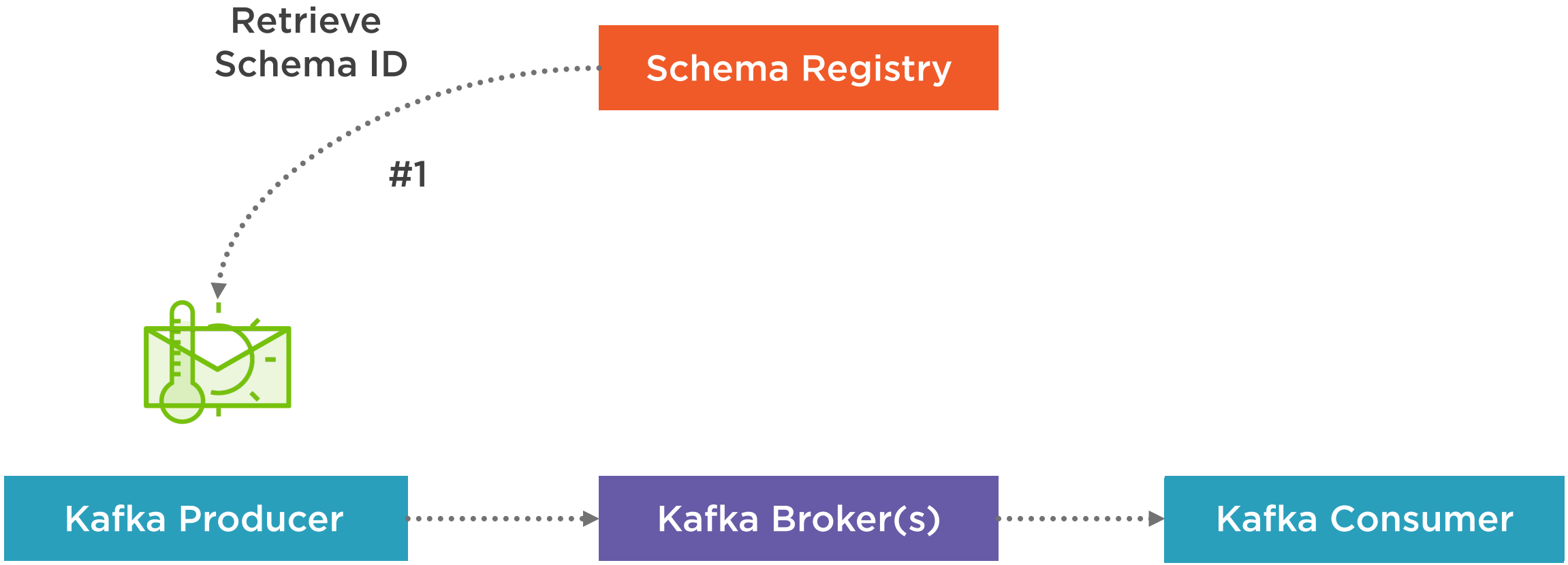
Confluent Community License

You can **access** the source code and **modify** or **redistribute** it; there is only one thing you cannot do, and that is use it to make a competing SaaS offering.

- Confluent

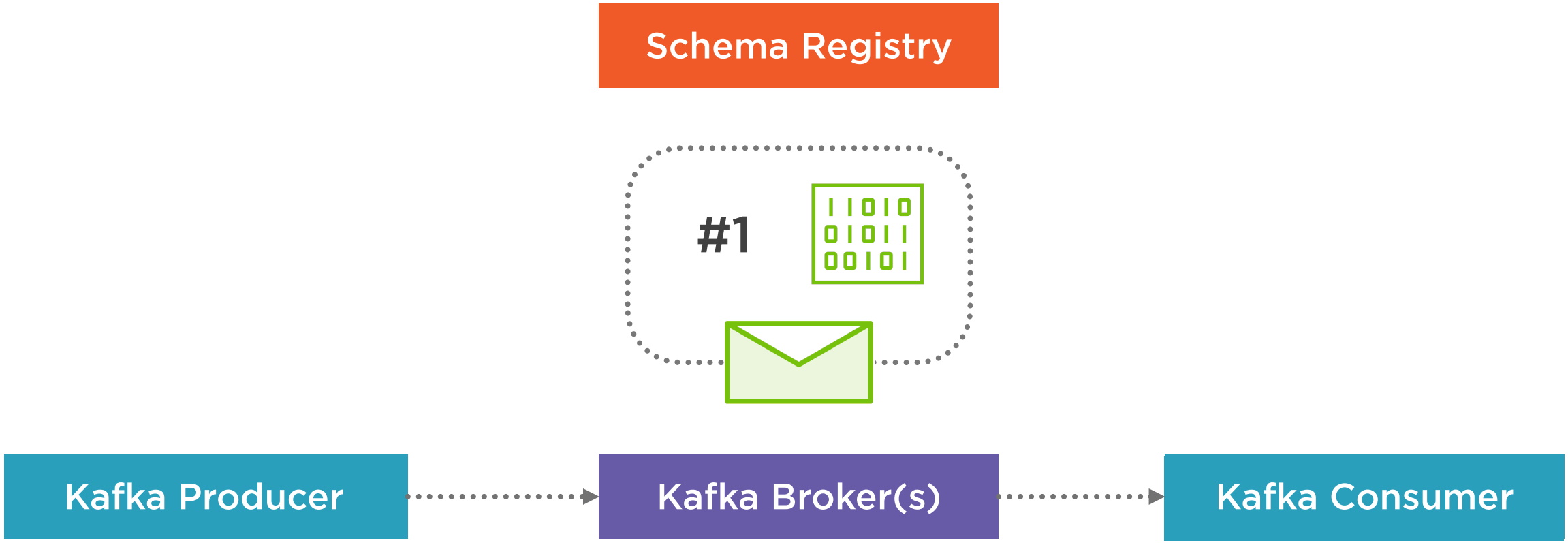

Kafka Connect
Kafka Broker(s)
Kafka Consumer
Kafka Producer
Poison Pill

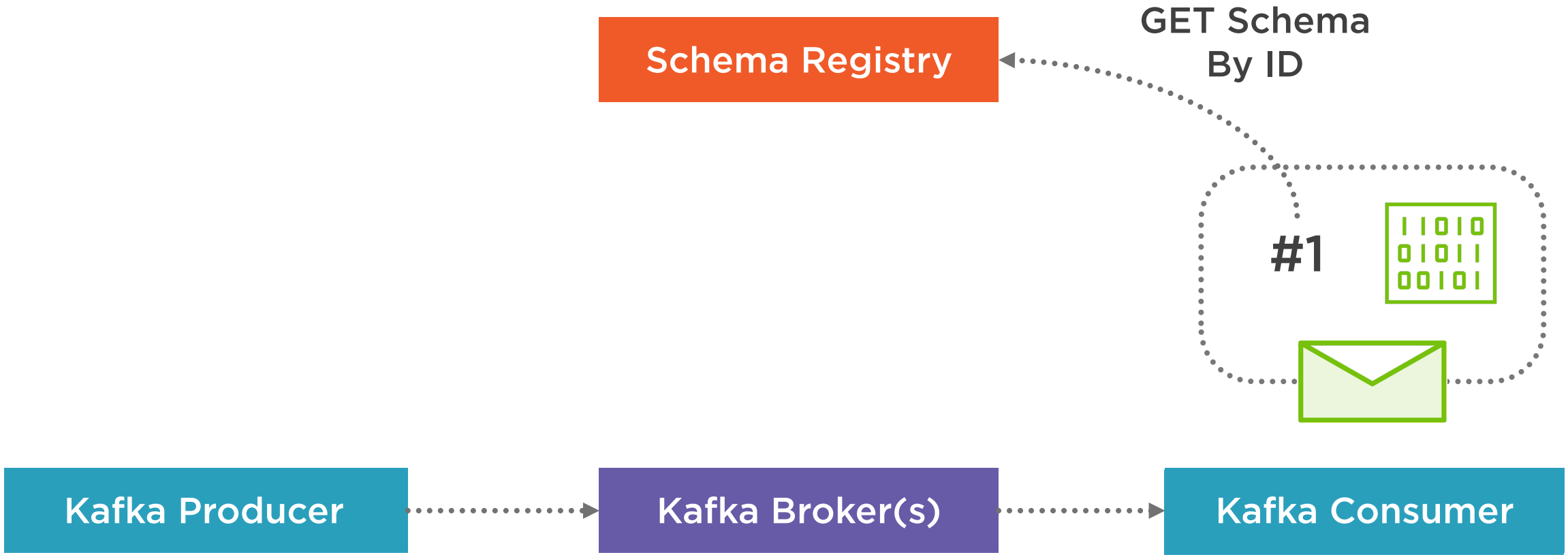

Register
Schema
Schema Registry
Kafka Producer
Kafka Broker(s)
Kafka Consumer

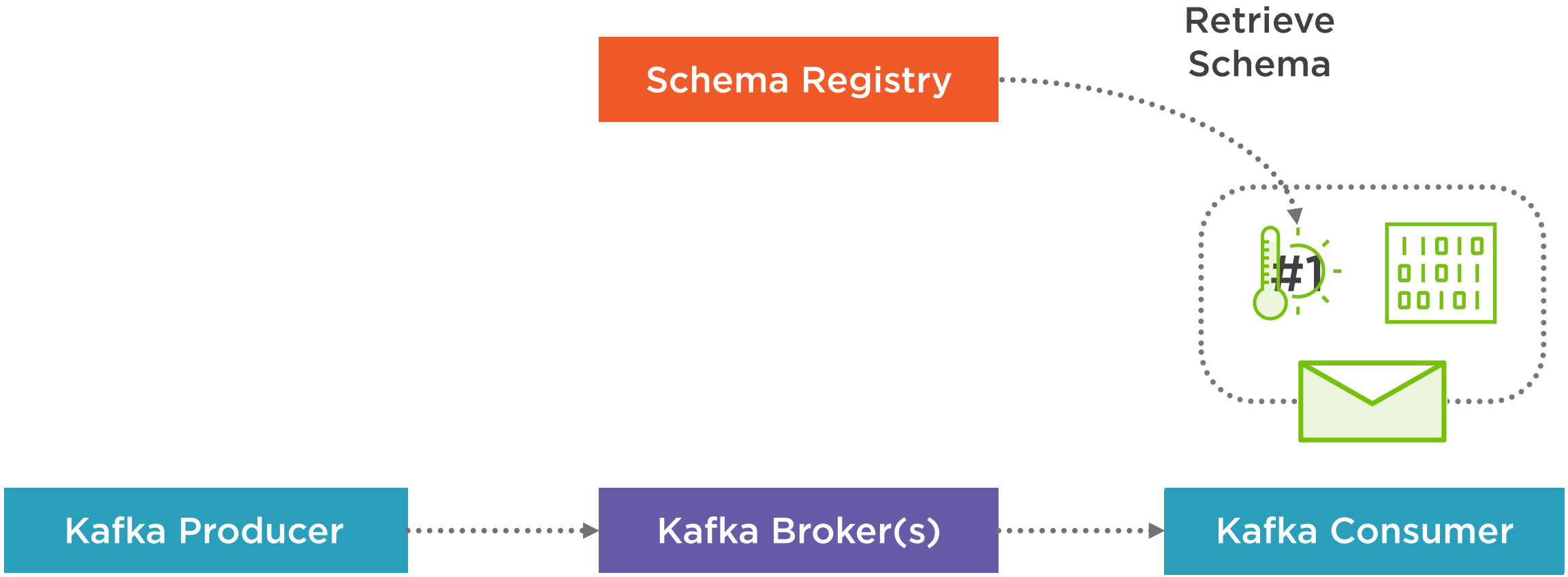

Retrieve
Schema ID
Schema Registry
#1
Kafka Producer
Kafka Broker(s)
Kafka Consumer

Schema Registry

Kafka Producer → Kafka Broker(s) → Kafka Consumer


Schema Registry
#1
Kafka Producer
Kafka Broker(s)
Kafka Consumer


Schema Registry
GET Schema
By ID
#1
Kafka Producer
Kafka Broker(s)
Kafka Consumer


Retrieve
Schema
Schema Registry
#1
Kafka Producer
Kafka Broker(s)
Kafka Consumer

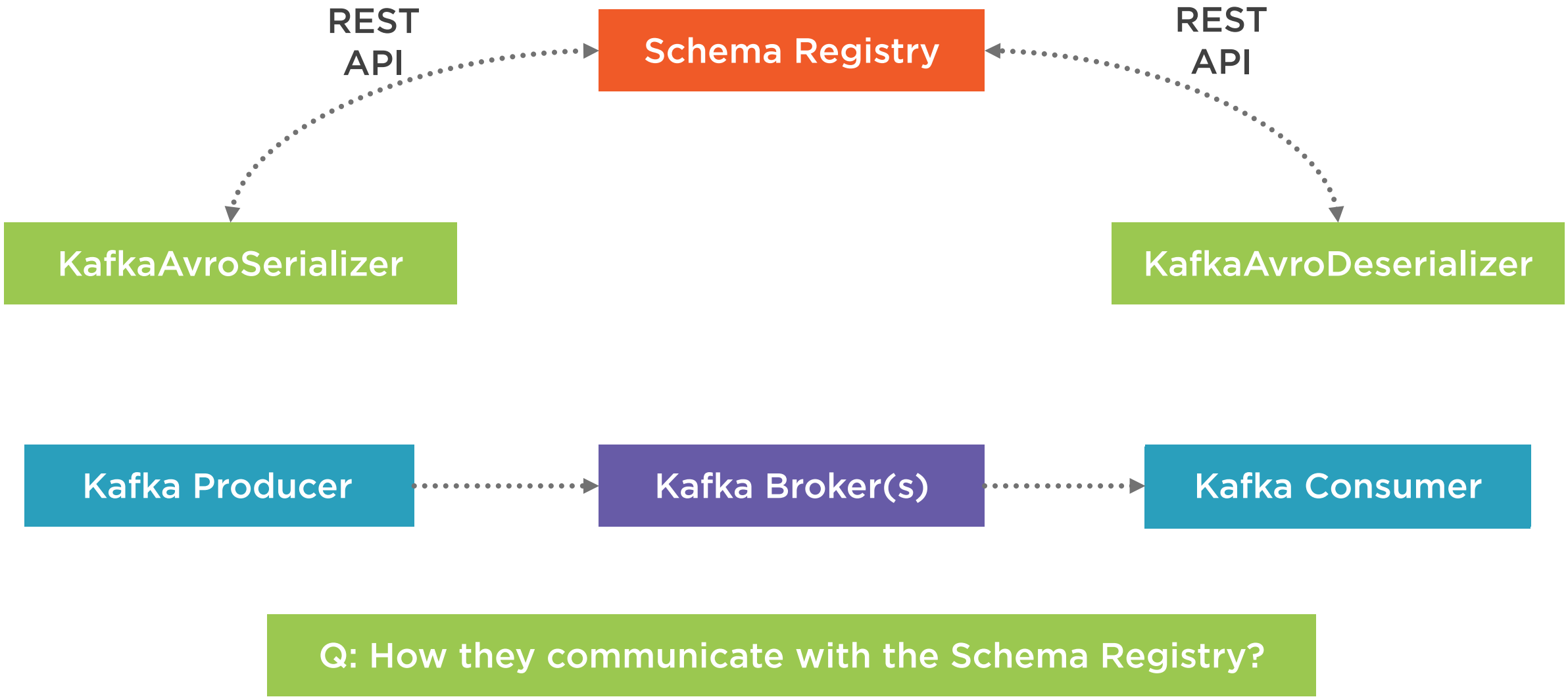
Schema Registry

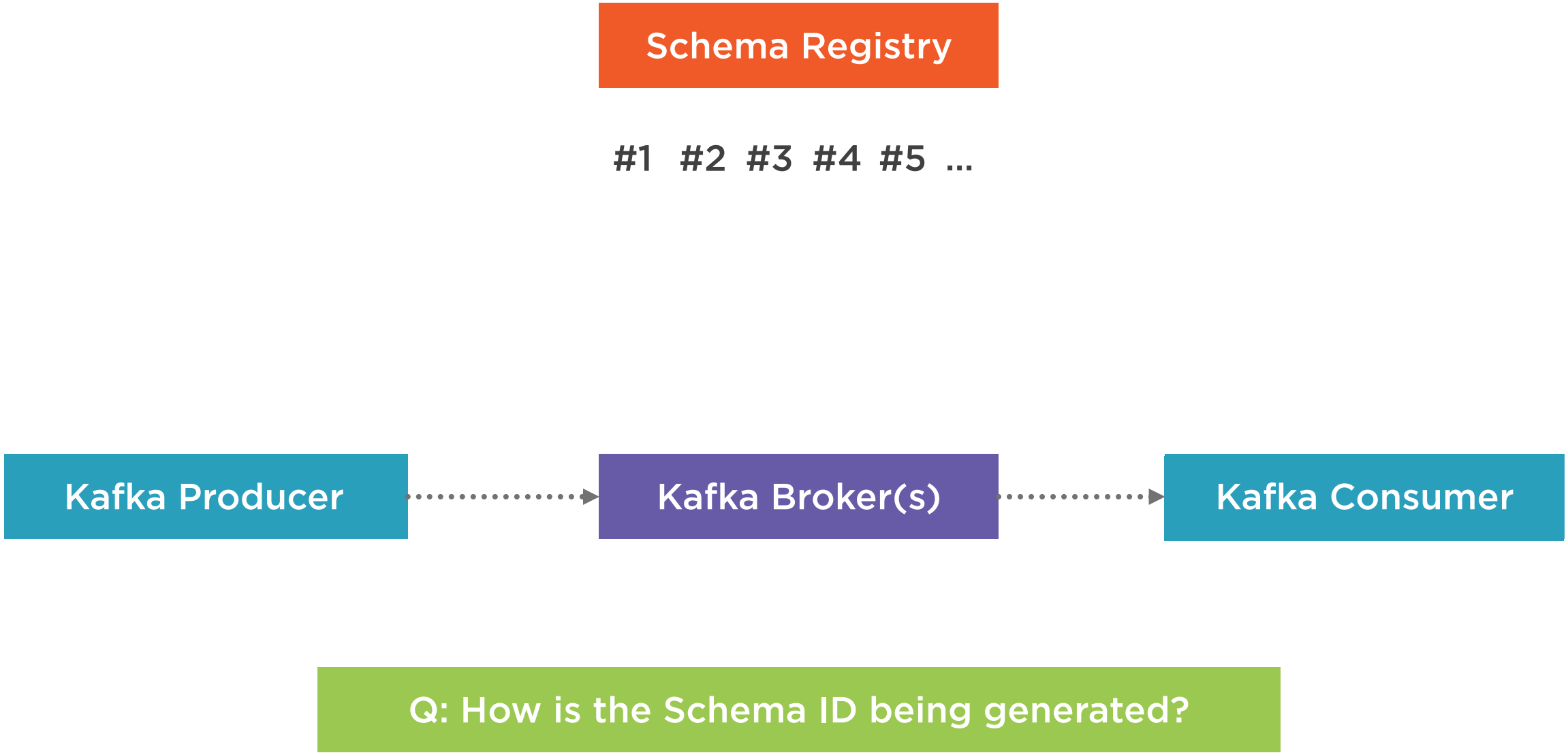
KafkaAvroSerializer

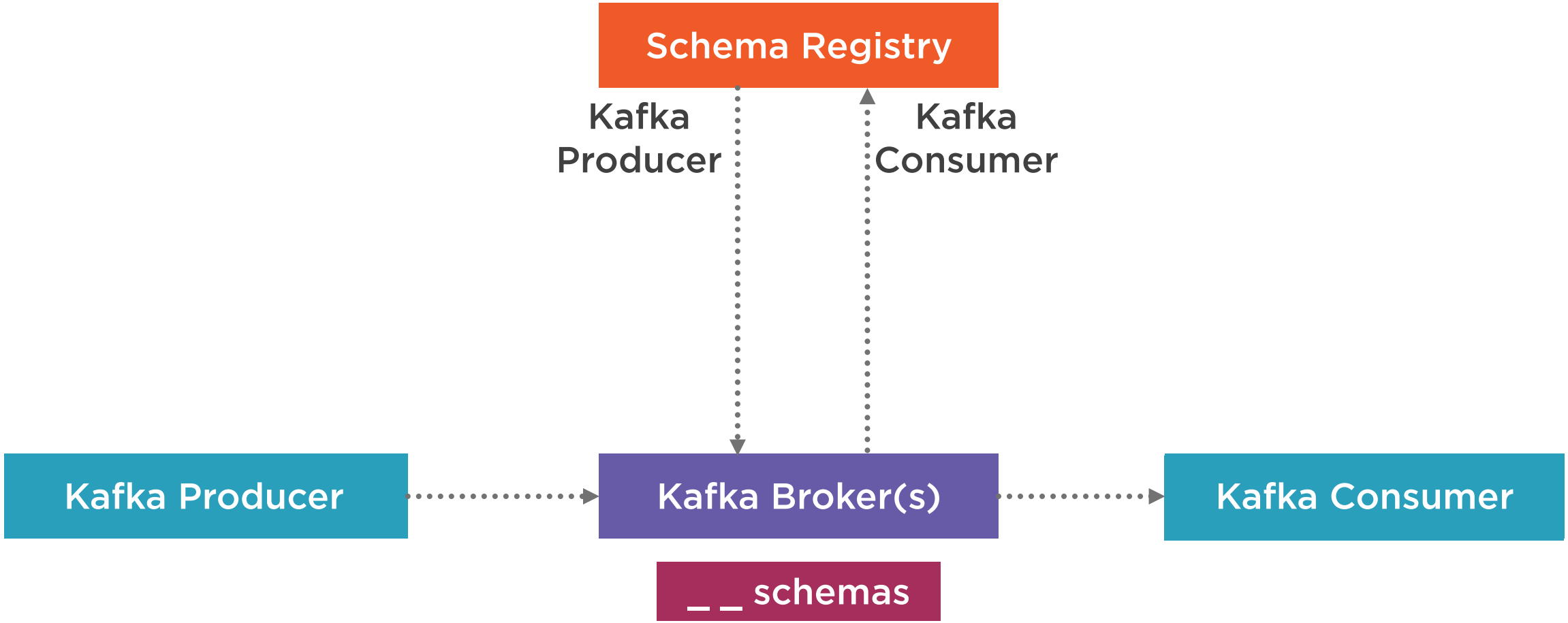
KafkaAvroDeserializer

Kafka Producer ····▶ Kafka Broker(s) ····▶ Kafka Consumer

Q: How does this process take place?


REST
API
Schema Registry
REST
API
KafkaAvroSerializer
KafkaAvroDeserializer
Kafka Producer
Kafka Broker(s)
Kafka Consumer
Q: How they communicate with the Schema Registry?


Schema Registry
#1 #2 #3 #4 #5 ...
Kafka Producer
Kafka Broker(s)
Kafka Consumer
Q: How is the Schema ID being generated?



Q: Where does Schema Registry store the schemas?

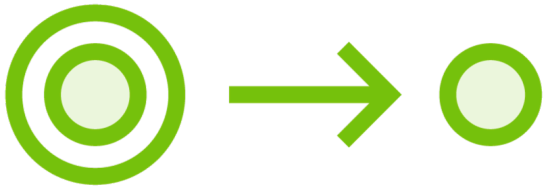
Demo

Modify Weather Producer / Consumer to use Schema Registry

Subject Name Strategy

Data Life Cycle

Created



Evolves

Deprecated / Removed

Schema Registry

Topic

Schema

Subject

Schema Registry

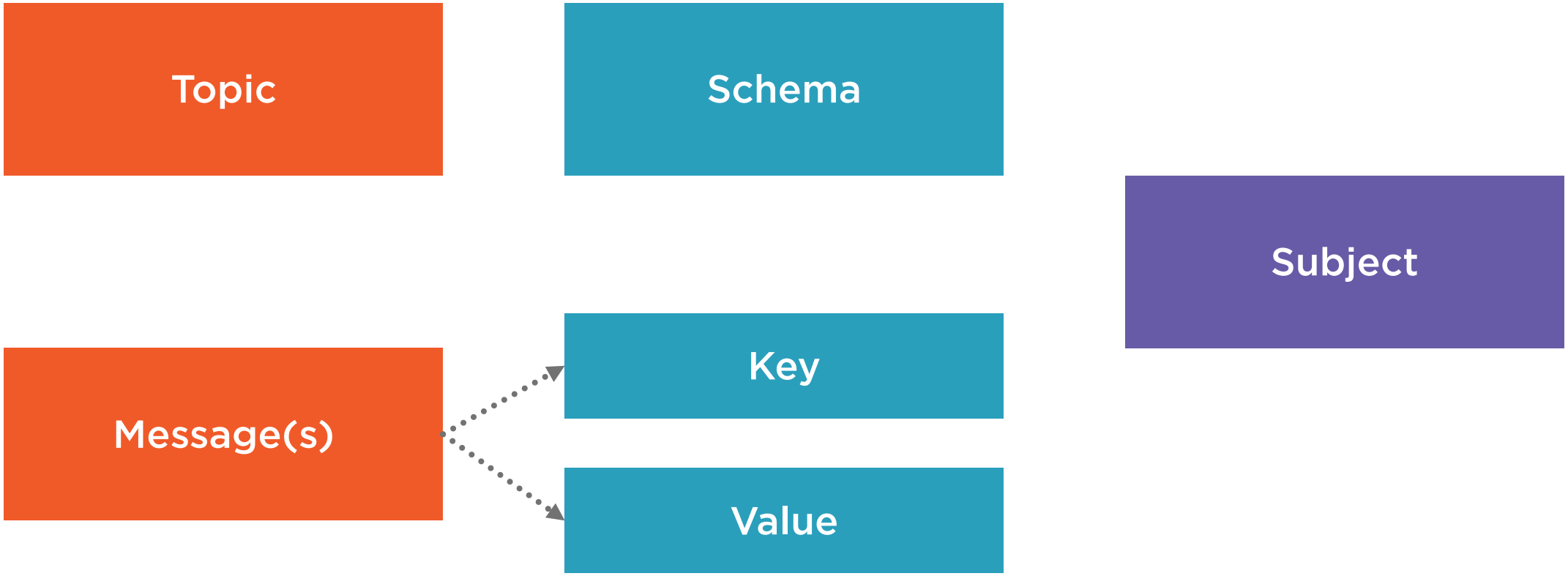
Topic

Schema

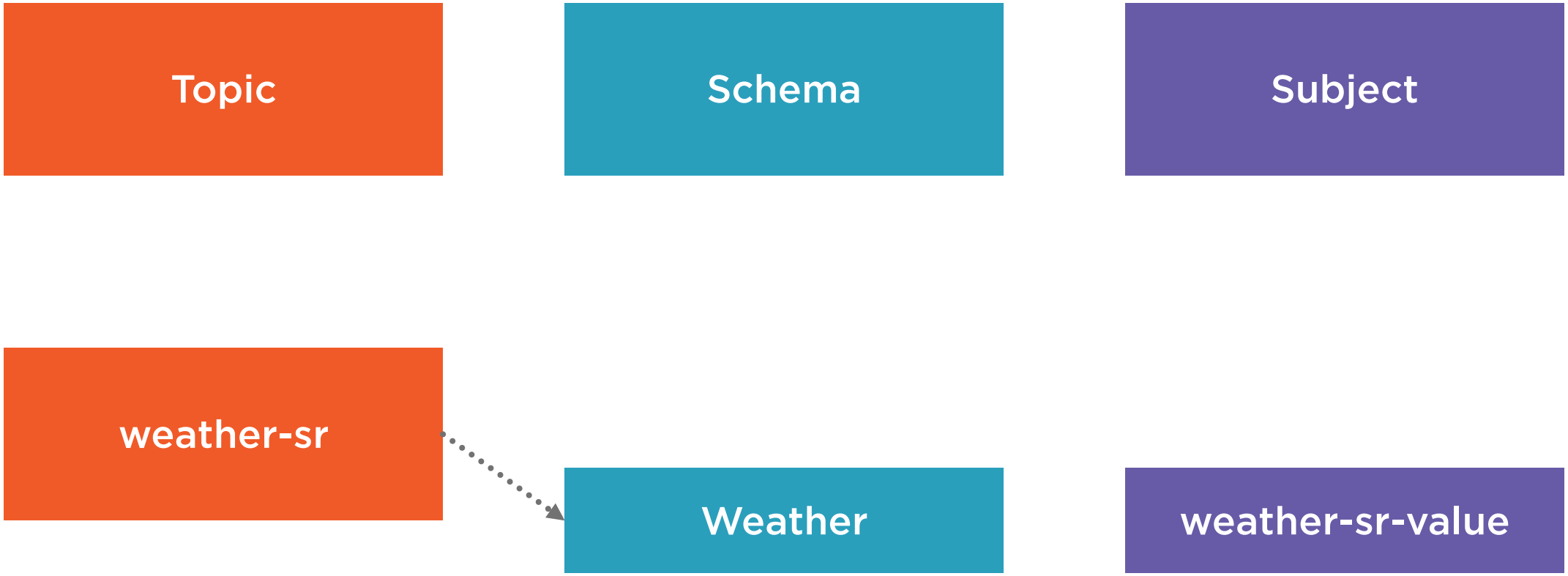
Subject

Message(s)

Schema Registry



Schema Registry



Subject Name Strategy

Topic Name Strategy

Record Name Strategy

Topic Record Name Strategy

Subject Name Strategy

Topic Name Strategy	Record Name Strategy	Topic Record Name Strategy
${topic}-key	${record}	${topic}-${record}
${topic}-value	${record}	${topic}-${record}

${topic} – Topic name

${record} – Schema name

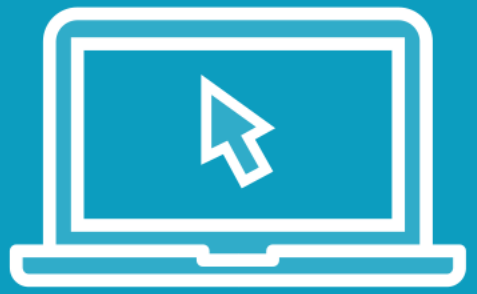
Subject Name Strategy

	Topic Name Strategy	Record Name Strategy	Topic Record Name Strategy
Default	✓	✗	✗
Unique Subject / Topic	✓	✗	✓
Multiple Record Types	✗	✓	✓
Schema Evolution	✓	✓	✓

Demo

Topic Name Strategy

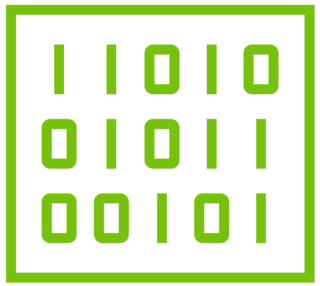
Demo



Record Name Strategy

Kafka AVRO Records

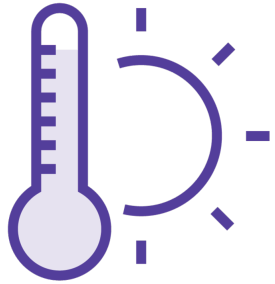
Byte Array Serializer



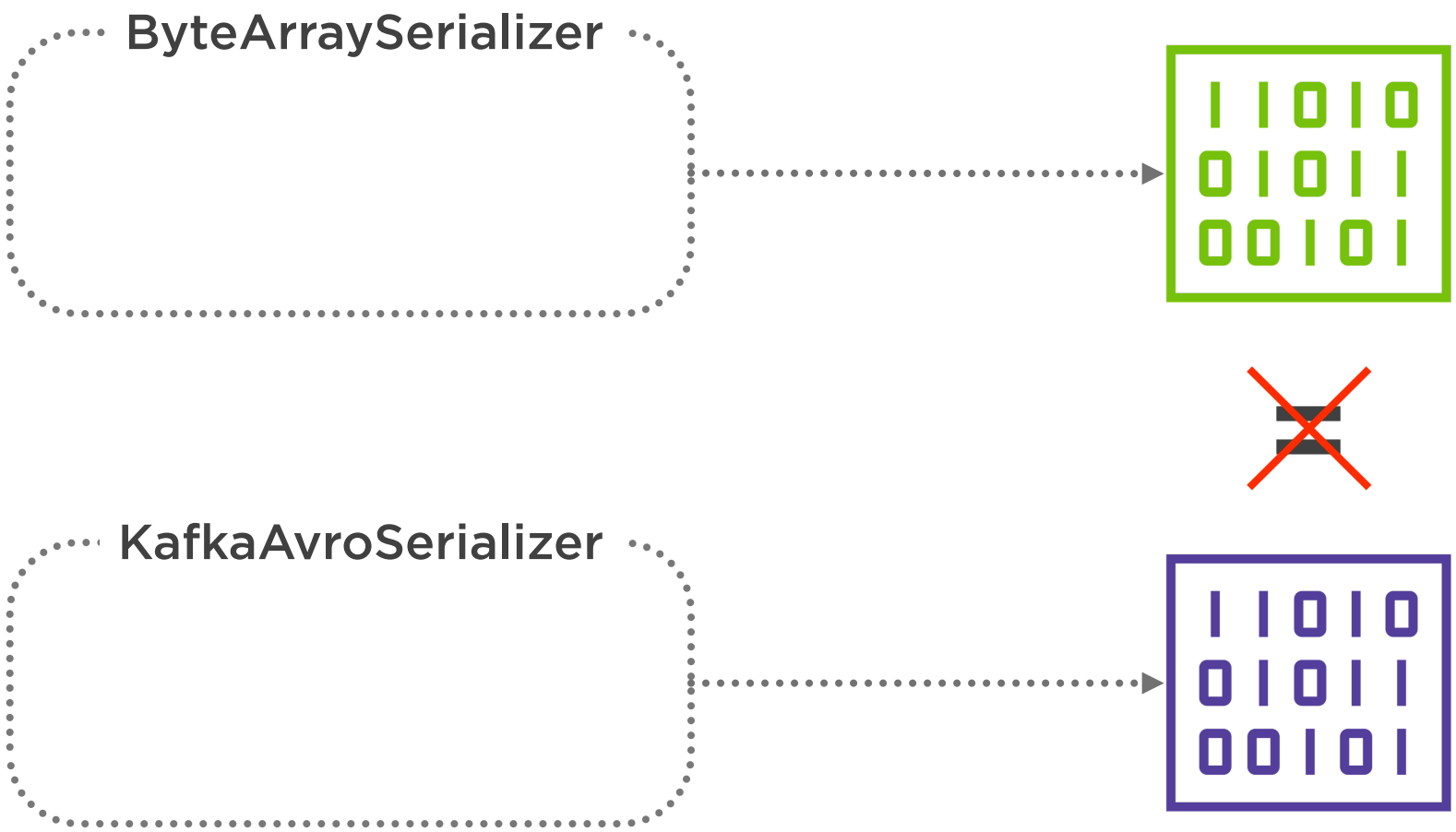
.toByteBuffer()
.array()

ByteArraySerializer

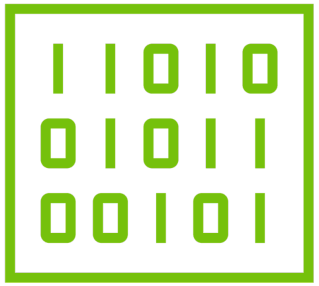
Kafka Avro Serializer



Serializers Comparison



Serializers Comparison

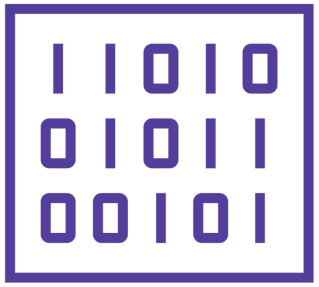


ByteArraySerializer

0010 1100 1111 0101 0011 1010 0001 1111 ...

Data

Confluent Wire Format



KafkaAvroSerializer

0

Magic Byte

0001

Schema ID

0010 1100 1111 0101 0011 1010 0001 1111 ...

Data

Kafka Avro Messages

0 0001 0010 1100 1111 0101 0011 1010 0001 1111 ...

Magic Byte

Schema ID

Data

Demo

Kafka Connect

Dump weather information to MongoDB

Summary

Enforcing Data Contracts

Subject Name Strategies

Kafka AVRO Records