Exploring the Kafka Ecosystem and Its Future
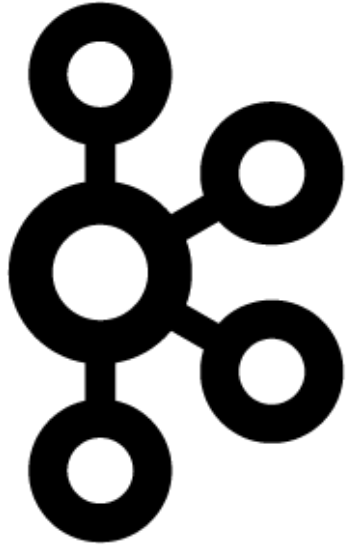
Ryan Plant
COURSE AUTHOR

@ryan_plant   blog.ryanplant.com

Primary Use Cases for Apache Kafka

Connecting disparate sources of data

Large-scale data movement pipelines
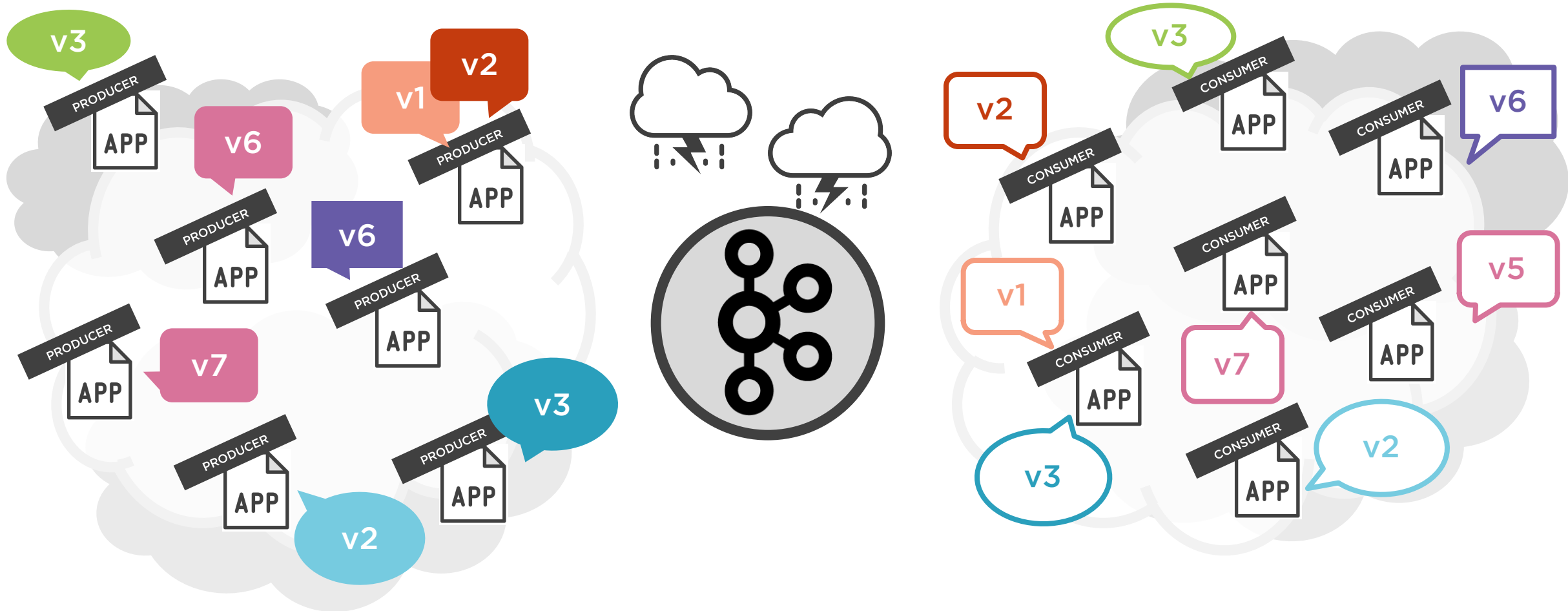
"Big Data" integration

Challenges Remain

Governance and Data Evolution

Consistency and Productivity

Big and Fast Data

Challenges in Governance and Evolution

KEY.SERIALIZER ········· KEY.DESERIALIZER

VALUE.SERIALIZER ········· VALUE.DESERIALIZER

Kafka Schema Registry

Apache Avro serialization format

First-class Avro serializers and deserializers
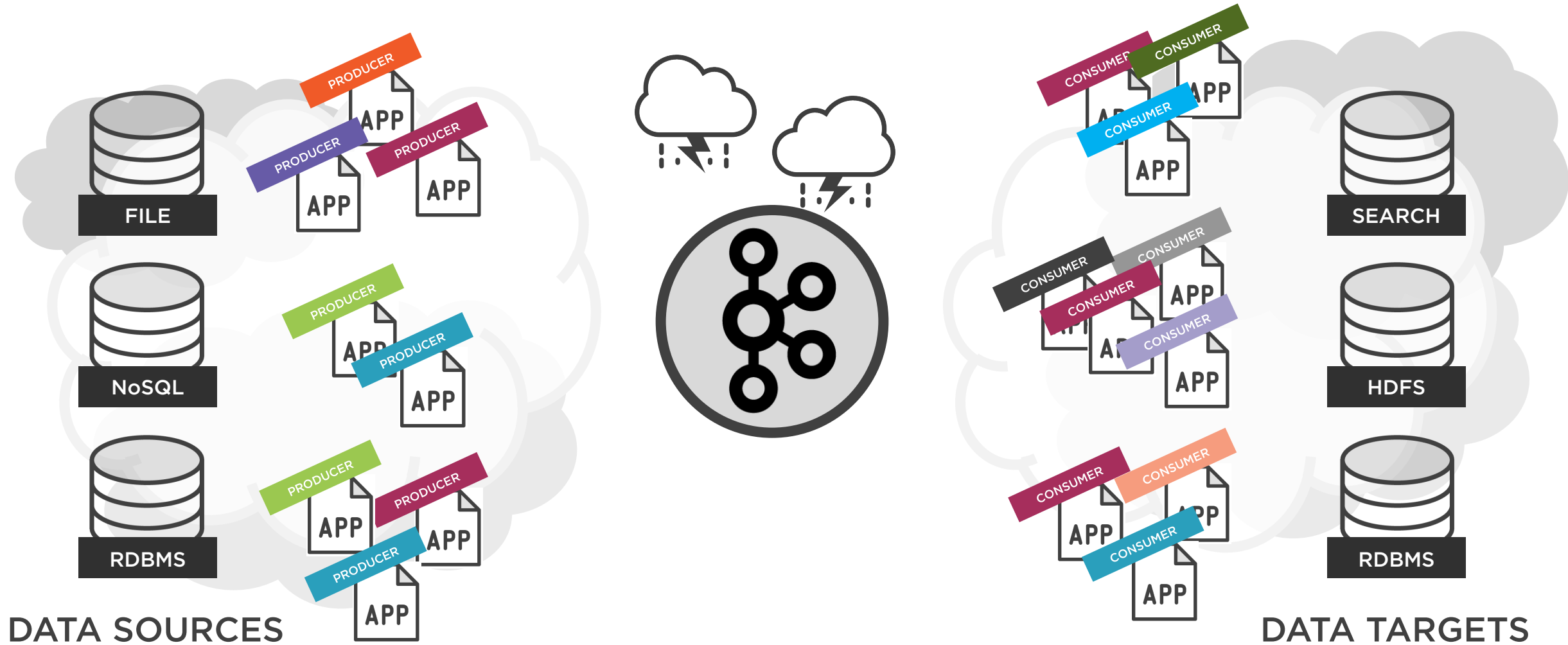
Schema registry and version management

RESTful service discovery

Compatibility broker

Open source with Apache license

Challenges with Consistency and Productivity

DATA SOURCES

DATA TARGETS

Apache Kafka Connect

Common framework for integration

– Standardization of common approaches

– Producers and Consumers

Platform Connectors
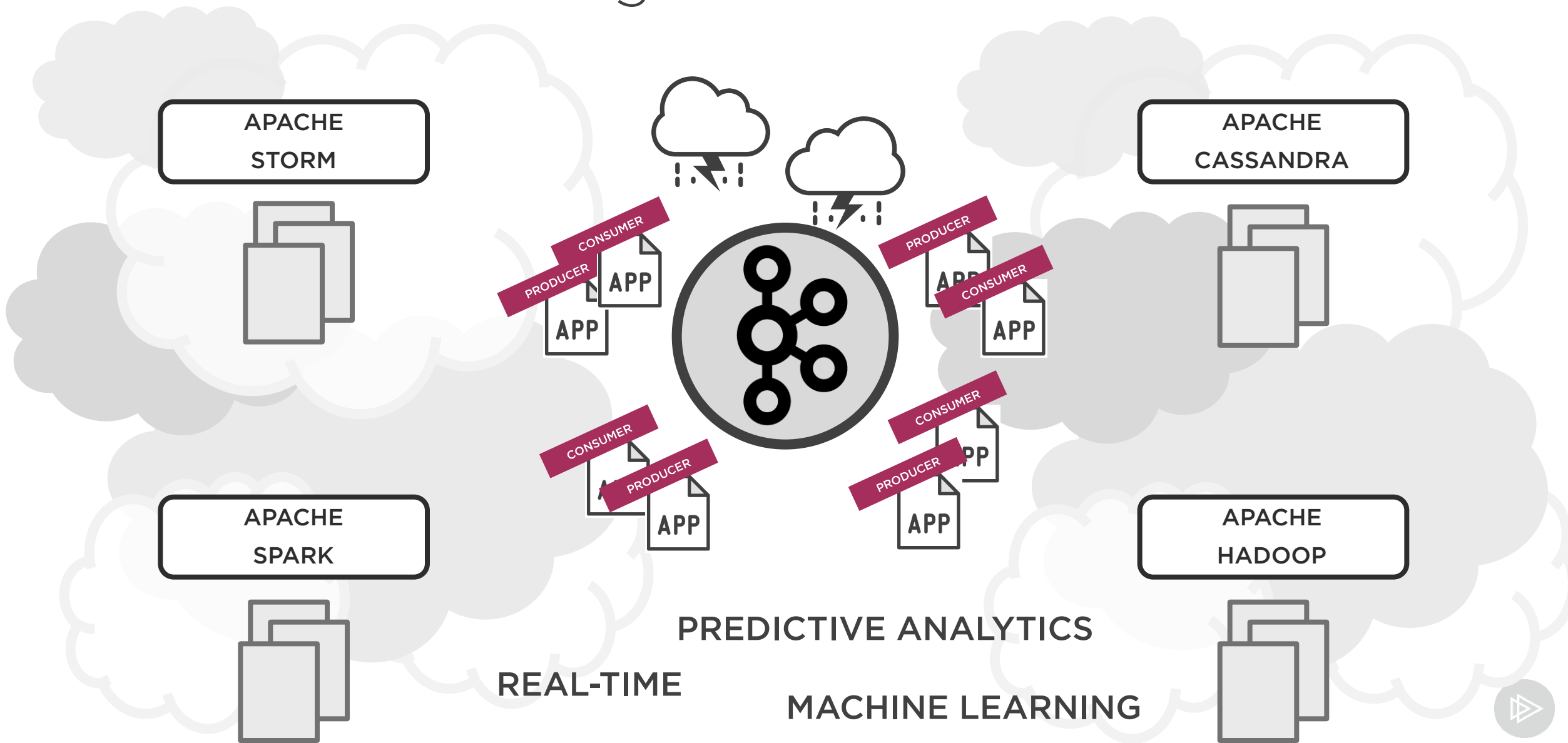
– Oracle, HP, etc.

– 50+ and growing

Connector Hub

Challenges with Fast Data

APACHE STORM

APACHE CASSANDRA

APACHE SPARK

APACHE HADOOP

CONSUMER

PRODUCER

APP

APP

PRODUCER

APP

CONSUMER

APP

CONSUMER

PRODUCER

APP

CONSUMER

PRODUCER

APP

PREDICTIVE ANALYTICS

REAL-TIME

MACHINE LEARNING

Kafka Streams

Leverages existing Kafka machinery

Single infrastructure solution
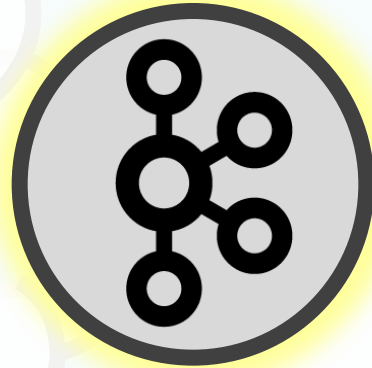- At least for streaming-based processing

Embeddable within existing applications

Growing and Healthy Ecosystem

LinkedIn

Netflix

Confluent

Twitter

Uber

Summary

Apache Kafka's unquestionable success

Continued challenges in an ever-changing industry

Kafka is evolving:

- Schema Registry
- Kafka Connect and Hub
- Kafka Streams

Lots of technology, little time

Apache Kafka is a solid bet