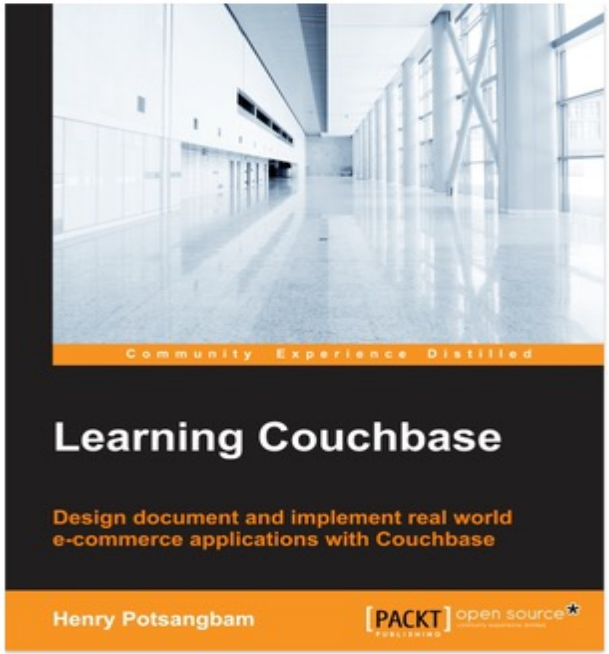
Kafka – Admins & Operations

Henry R.P

- **Generative AI - 360 Certification**
- **Confluence Kafka Accreditation**
- **Certified Cassandra Admin**
- **Mapr Certified – Hadoop Administrator**
- **IBM Certified Application Developer**
- **IBM Certified Solution Designer**
- **SAP Certified ABAP & Portal Consultant .**
- **CIPM – Certificate in Project Management.**
- **TOGAF – Enterprise Architect**

NOSQL, Streaming Platform & Bigdata

IT Architect & Corporate trainer
20 +Year of IT Experience



Author

Clientele

Tos

Outline

Tos

Overview of Kafka –

- Topics, Partitions, Brokers,
- Producer, Cluster, Offset, Producer, Consumer

Kafka Logging

Kafka Architecture & Ecosystem

Kafka Streams

KSQL DB

Kafka Security

Introduce Yourself.

Tos

Name
Year of Experience.
Skills Level
    Java / Linux
    Messaging System / Kafka
Expectation, if any.

Note: Basic knowledge of Java & Linux are required.

Schedule

Tos

Time	
9.30 – 11.00 AM	Session I
11.00 AM to 11.15 AM	Tea Break
11.15 AM to 12.45 PM	Session II
12.45 PM to 1.45 PM	Lunch Break
1.45 PM to 3.15 PM	Session III
3.15 PM to 3.30 PM	Tea Break
3.30 PM to 5.30 PM	Session IV

15 July 2024

Kafka – An Overview

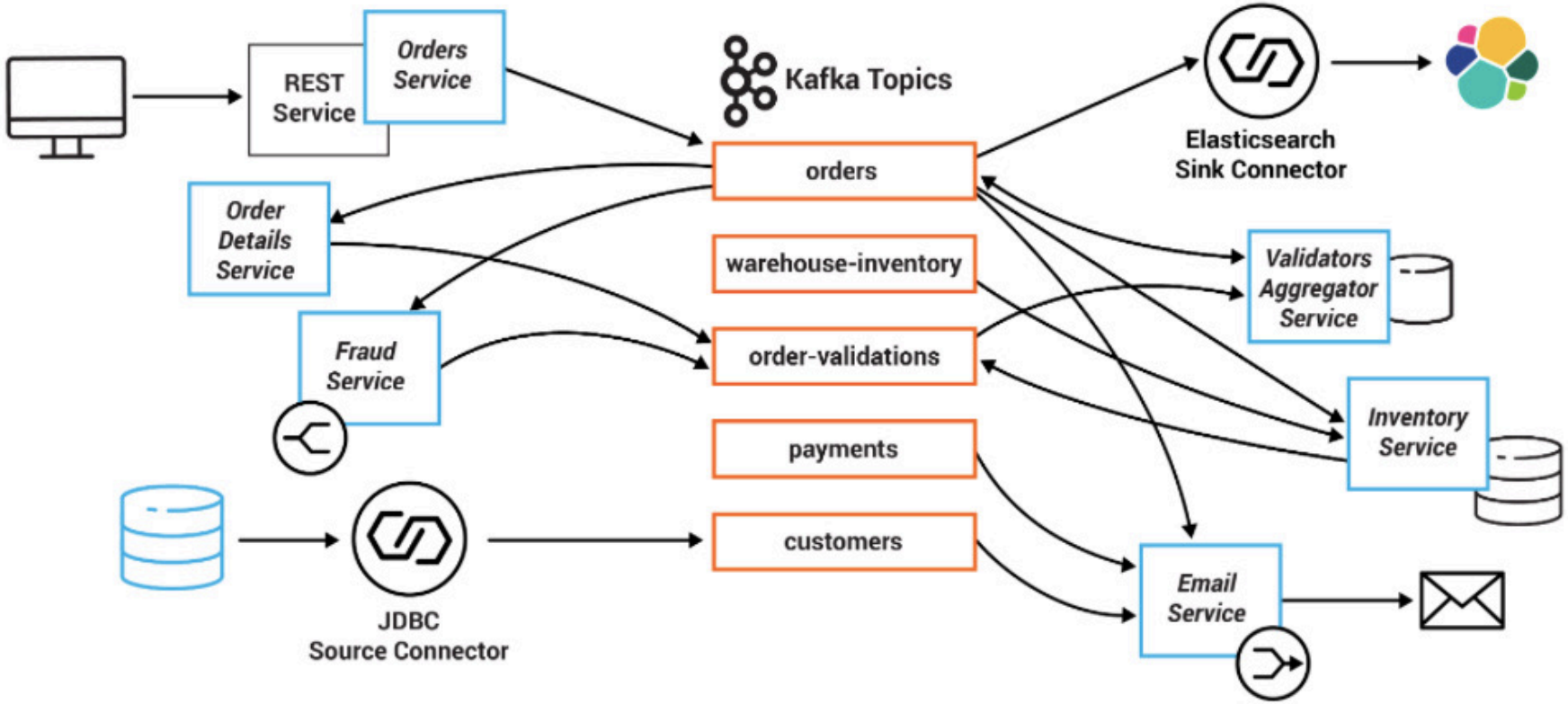
Kafka growth exploding

Tos

- 1/3 of all Fortune 500 companies
- Top ten travel companies, 7 of ten top banks, 8 of ten top insurance companies, 9 of ten top telecom companies
- LinkedIn, Microsoft and Netflix process 4 comma message a day with Kafka (1,000,000,000,000)
- Real-time streams of data, used to collect big data or to do real time analysis (or both)

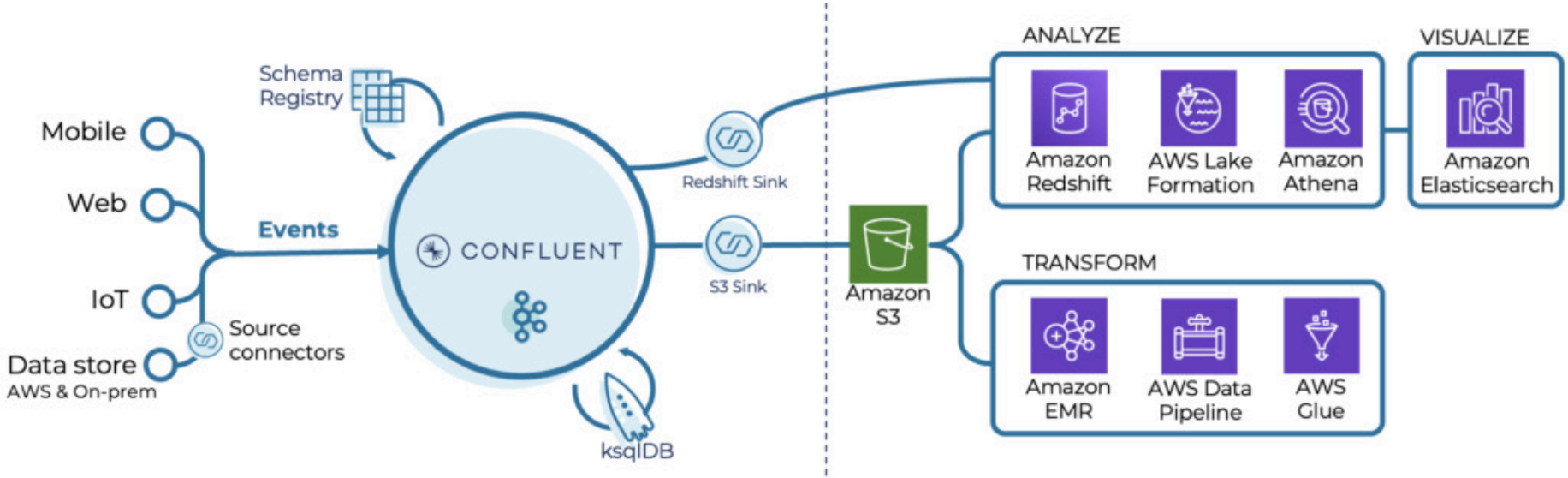
Why Kafka is Needed? - MicroService

Tos



Why Kafka is Needed? – Big Data

Tos



Why Kafka is Needed?

Tos

- Apache Kafka is a fast, scalable, durable, and fault- tolerant publish-subscribe messaging system
- Real time streaming data processed for real time analytics
- Service calls, track every call, IOT sensors
- Kafka is often used instead of JMS, RabbitMQ and AMQP
- higher throughput, reliability and replication

Why is Kafka needed? 2

Tos

- Kafka can works in combination with
  - Flume/Flafka, Spark Streaming, Storm, HBase and Spark for real-time analysis and processing of streaming data
  - Feed your data lakes with data streams
- Kafka brokers support massive message streams for follow- up analysis in Hadoop or Spark
- Kafka Streaming (subproject) can be used for real-time analytics

Kafka Usage

Tos

- Build real-time streaming applications that react to streams
  - Real-time data analytics
  - Transform, react, aggregate, join real-time data flows
  - Feed events to CEP for complex event processing
  - Feed data lakes
- Build real-time streaming data pipe-lines
  - Enable in-memory micro services (actors, Akka, Vert.x, Qbit, RxJava)

Why is Kafka Popular?

Tos

- ***Great performance***
- Operational Simplicity, easy to setup and use, easy to reason
- Stable, Reliable Durability,
- Flexible Publish-subscribe/queue (scales with N-number of consumer groups),
- Robust Replication,
- Works well with systems that have data streams to process, aggregate, transform & load into other stores

Why is Kafka so fast?

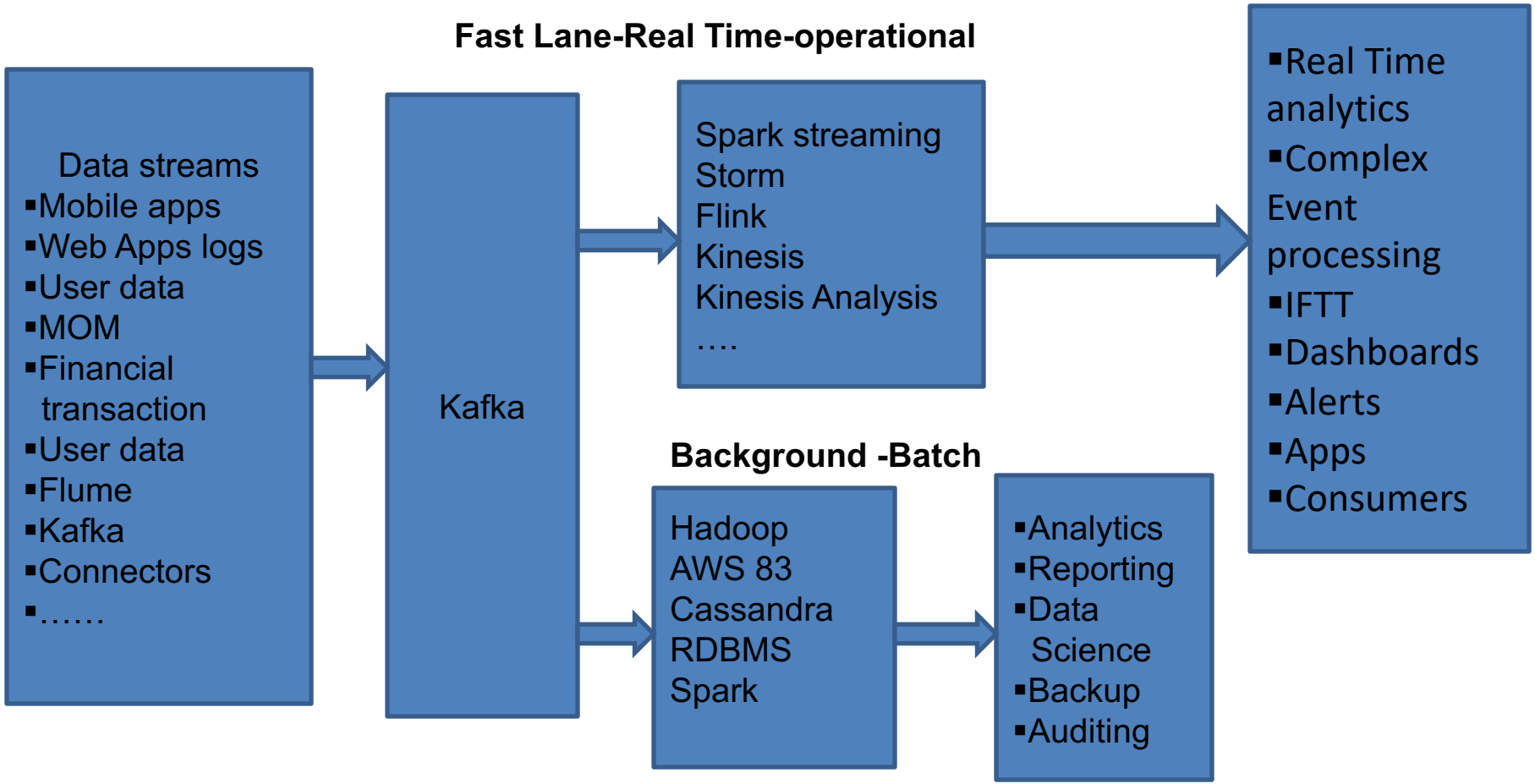
Tos

- Zero Copy - calls the OS kernel direct rather to move data fast
- **Batch Data in Chunks** - Batches data into chunks
  - end to end from Producer to file system to Consumer
  - Provides More efficient data compression. Reduces I/O latency
- **Sequential Disk Writes** - Avoids Random Disk Access
  - writes to immutable commit log. No slow disk seeking. No random I/O operations. Disk accessed in sequential manner
- **Horizontal Scale** - uses 100s to thousands of partitions for a single topic
  - spread out to thousands of servers
  - handle massive load

Kafka: A Stream Data Platform

Kafka Streaming Architecture

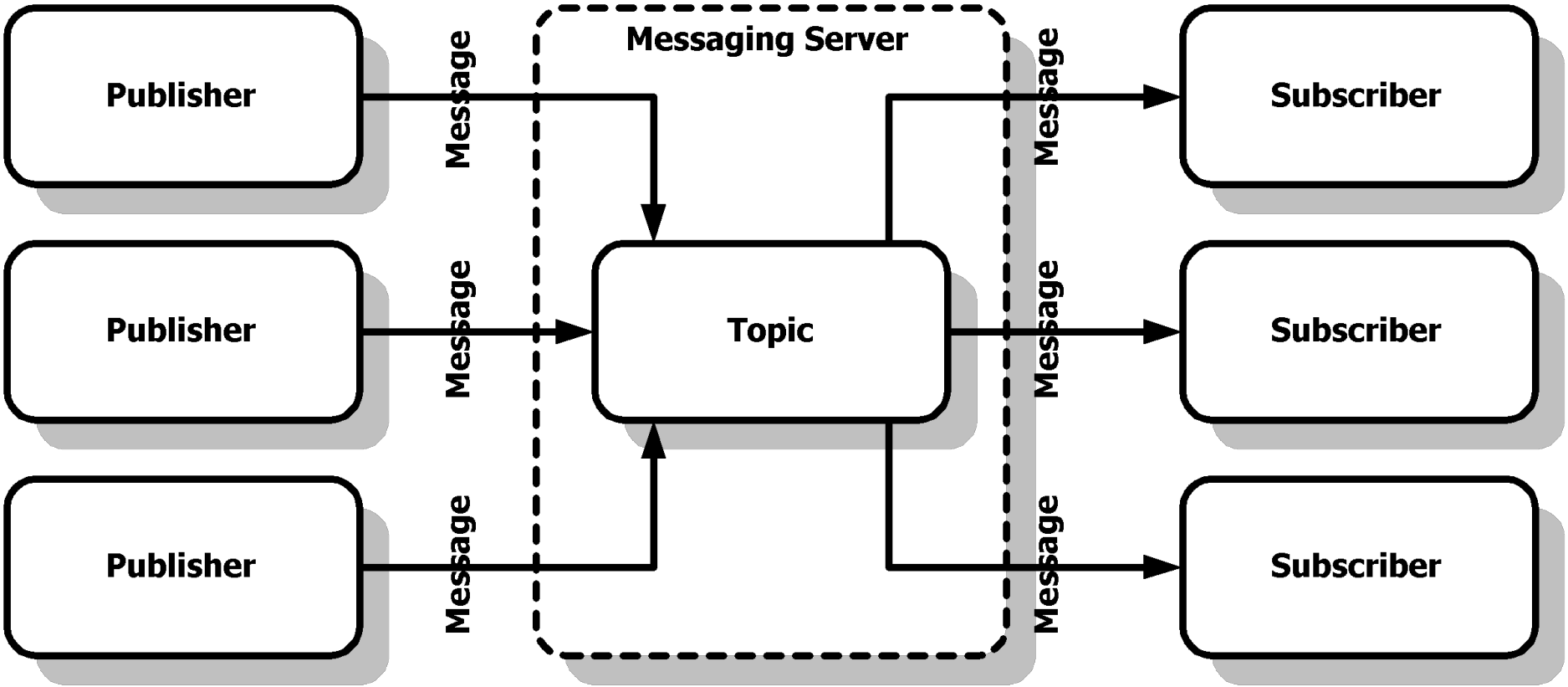
Tos



What is Kafka?

What is Kafka?

Tos



What is Kafka?

Tos

- Distributed Streaming Platform
  - Publish and Subscribe to streams of records
  - Fault tolerant storage
    - Replicates Topic Log Partitions to multiple servers
  - Process records as they occur
  - Fast, efficient IO, batching, compression, and more
- Used to decouple data streams

Kafka helps decouple data streams

Tos

- Kafka decouple data streams
- producers don’t know about consumers
- Flexible message consumption
  - Kafka broker delegates log partition offset (location) to Consumers (clients)

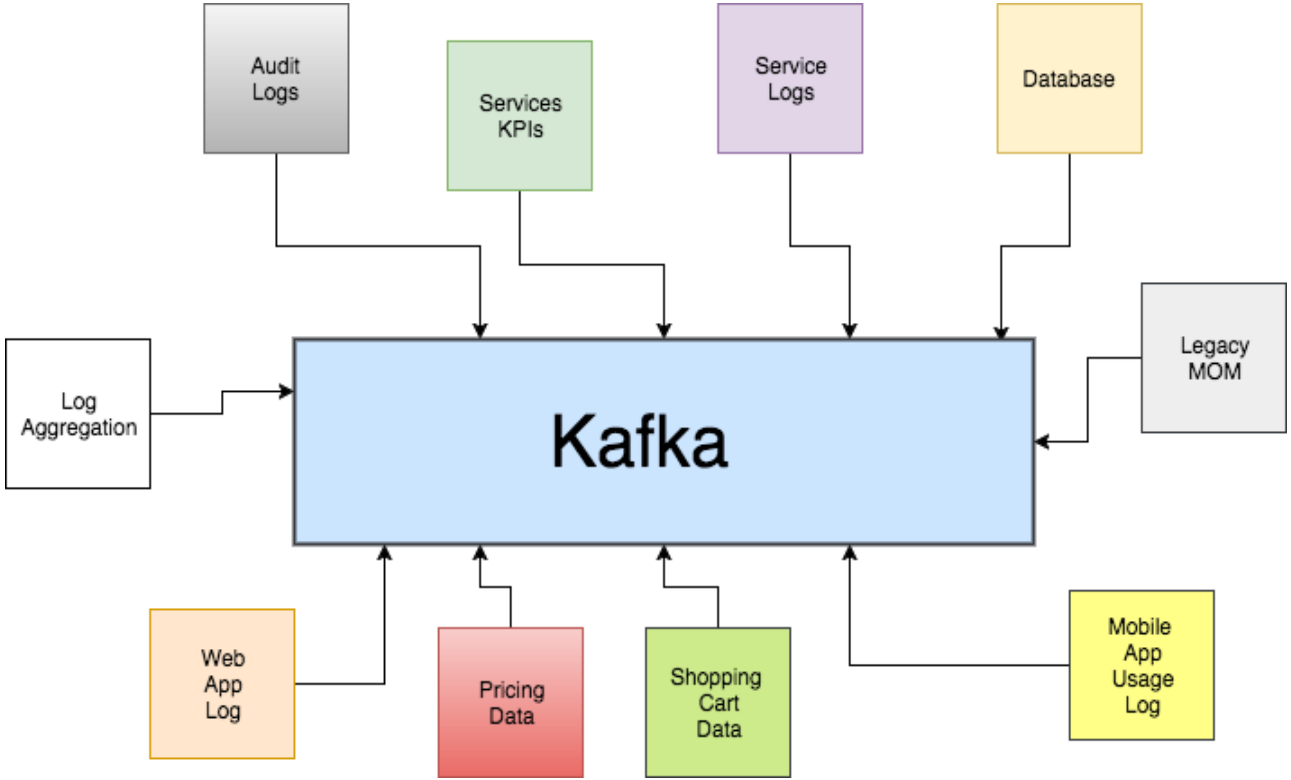
Kafka messaging allows

Tos

- Feeding of high-latency daily or hourly data analysis into Spark, Hadoop, etc.
- Feeding micro services real-time messages
- Sending events to CEP system
- Feeding data to do real-time analytic systems
- Up to date dashboards and summaries
- At same time

Kafka Decoupling Data Streams

Tos



Don’t couple the stream

Kafka Polyglot clients / Wire protocol

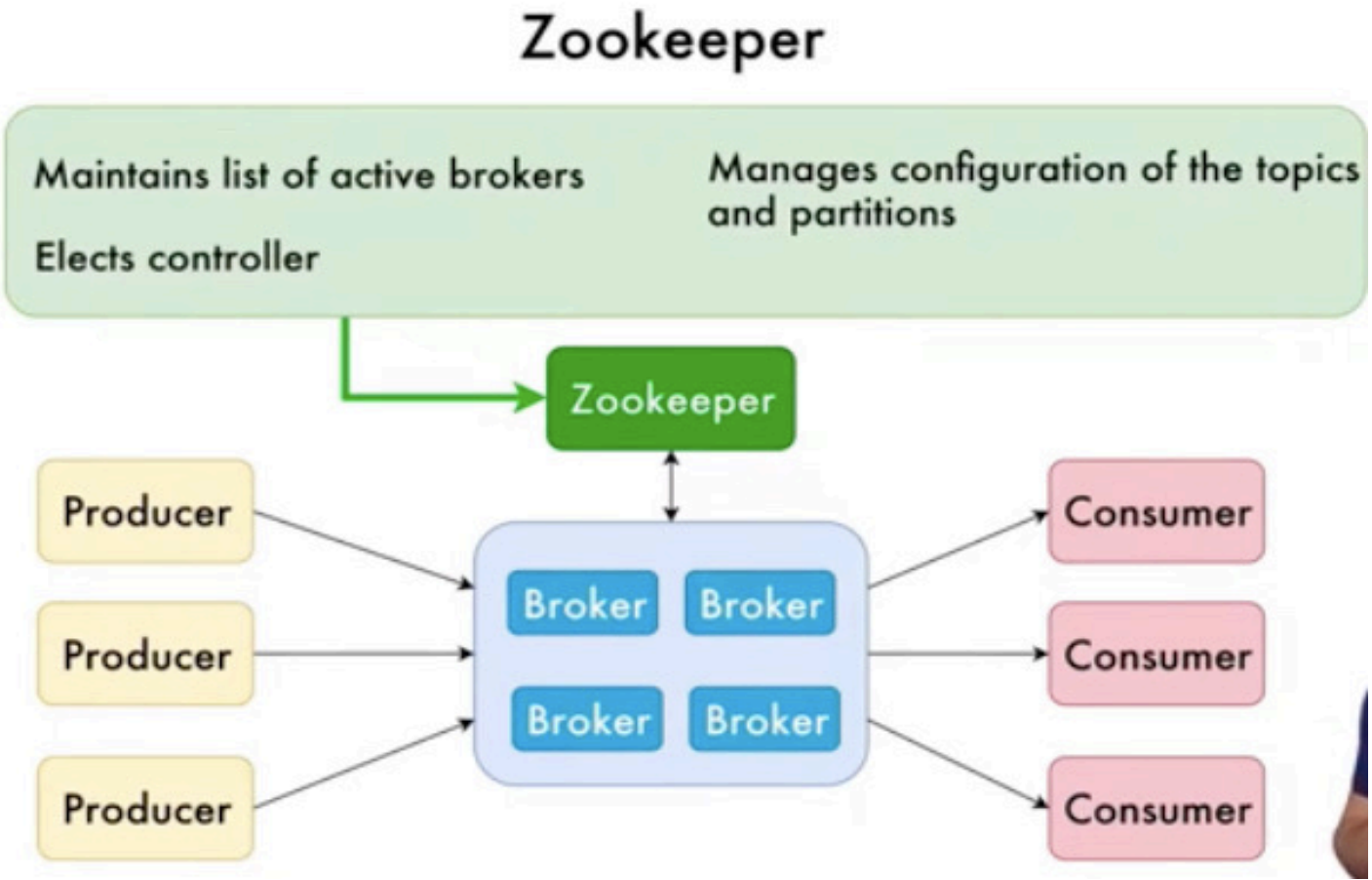
Tos

- Kafka communication from clients and servers wire protocol over TCP protocol
- Protocol versioned
- Maintains backwards compatibility
- Many languages supported
- Kafka REST proxy allows easy integration (not part of core)
- Also provides Avro/Schema registry support via Kafka ecosystem (not part of core)

Kafka Architecture

Kafka: Topics, Producers, and Consumers

Tos



Kafka Fundamentals

Tos

- ***Broker***: Kafka server that runs in a Kafka Cluster. Brokers form a cluster. Cluster consists on many Kafka Brokers on many servers.

- ***Zoo Keeper***: Does coordination of brokers/cluster topology. Consistent file system for configuration information and leadership election for Broker Topic Partition Leaders

Lab : **Installation of Kafka**