Big Data Analytics

تحلیل کلان داده

dayche.com . گروه دایکه

تحلیل کلان داده

مقدمه و معرفی

❑ آپاچی کافکا (Apache Kafka)

APACHE KAFKA

More than 80% of all Fortune 100 companies trust, and use Kafka.

Apache Kafka is an open-source distributed event streaming platform used by thousands of companies for high-performance data pipelines, streaming analytics, data integration, and mission-critical applications.

Apache Kafka® is an event streaming platform. What does that mean?

تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

❑ What is Event Streaming?

Event streaming is the digital equivalent of the human body's central nervous system. It is the technological foundation for the 'always-on' world where businesses are increasingly software-defined and automated, and where the user of software is more software.

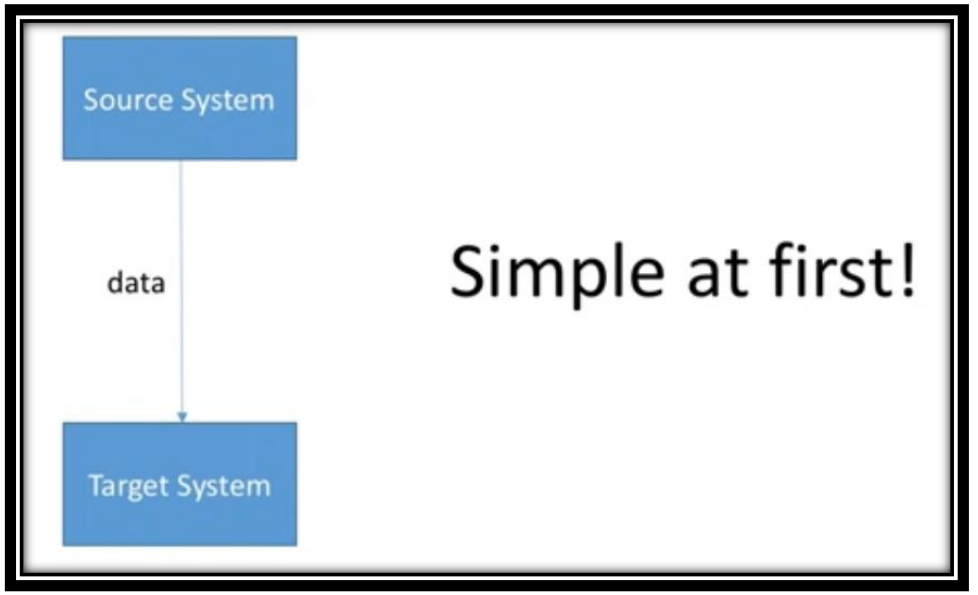
Technically speaking, **event streaming is the practice of capturing data in real-time from event sources like databases, sensors, mobile devices, cloud services, and software applications in the form of streams of events**; storing these event streams durably for later retrieval; manipulating, processing, and reacting to the event streams in real-time as well as retrospectively; and routing the event streams to different destination technologies as needed.

❑ What can I use event streaming for?

- To process payments and **financial transactions in real-time**, such as in stock exchanges, banks, and insurances.
- **To track and monitor** cars, trucks, fleets, and shipments in real-time, such as in logistics and the automotive industry.
- To continuously capture and analyze sensor data from **IoT devices** or other equipment, such as in factories and wind parks.
- To collect and immediately react to **customer interactions** and orders, such as in retail, the hotel and travel industry, and mobile applications.
- To monitor patients in hospital care and predict changes in condition to ensure timely treatment in emergencies.
- To connect, store, and make available data produced by different divisions of a company.
- To serve as the foundation for data platforms, event-driven architectures, and microservices.

❑ آپاچی کافکا (Apache Kafka)

- If you have 4 source systems, and 6 target systems, you need to write 24 integrations!
- Each integration comes with difficulties around
  - Protocol – how the data is transported *(TCP, HTTP, REST, FTP, JDBC…)*
  - Data format – how the data is parsed *(Binary, CSV, JSON, Avro…)*
  - Data schema & evolution – how the data is shaped and may change
- Each source system will have an increased load from the connections



LinkedIn™

تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

مقدمه و معرفی

❑ آپاچی کافکا (Apache Kafka)

Apache Kafka® is an event streaming platform. What does that mean?

Kafka combines three key capabilities so you can implement your use cases for event streaming end-to-end with a single battle-tested solution:

1. To **publish** (write) and **subscribe to** (read) streams of events, including continuous import/export of your data from other systems.
2. To **store** streams of events durably and reliably for as long as you want.
3. To **process** streams of events as they occur or retrospectively.

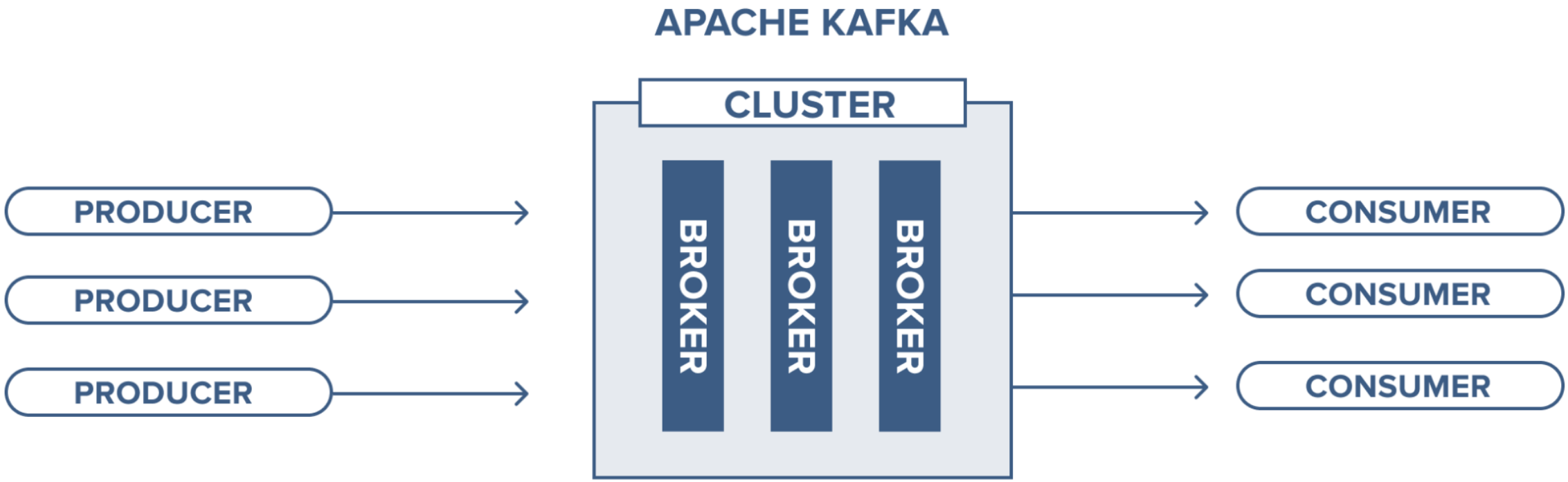
And all this functionality is provided in a distributed, highly scalable, elastic, fault-tolerant, and secure manner. Kafka can be deployed on bare-metal hardware, virtual machines, and containers, and on-premises as well as in the cloud. You can choose between self-managing your Kafka environments and using fully managed services offered by a variety of vendors.

تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

مقدمه و معرفی

❑ آپاچی کافکا (Apache Kafka)

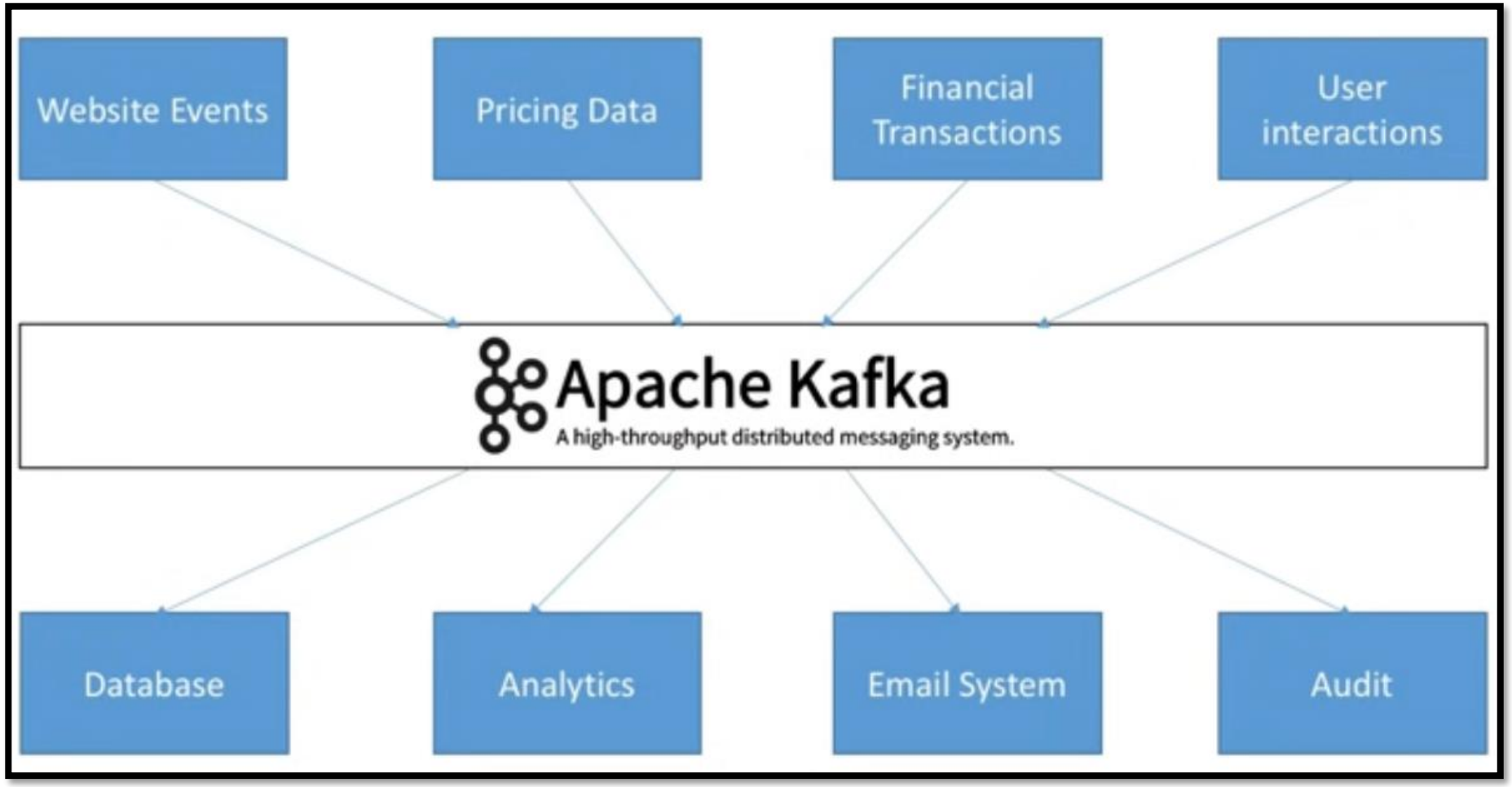


تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

مقدمه و معرفی

❑ آپاچی کافکا (Apache Kafka)

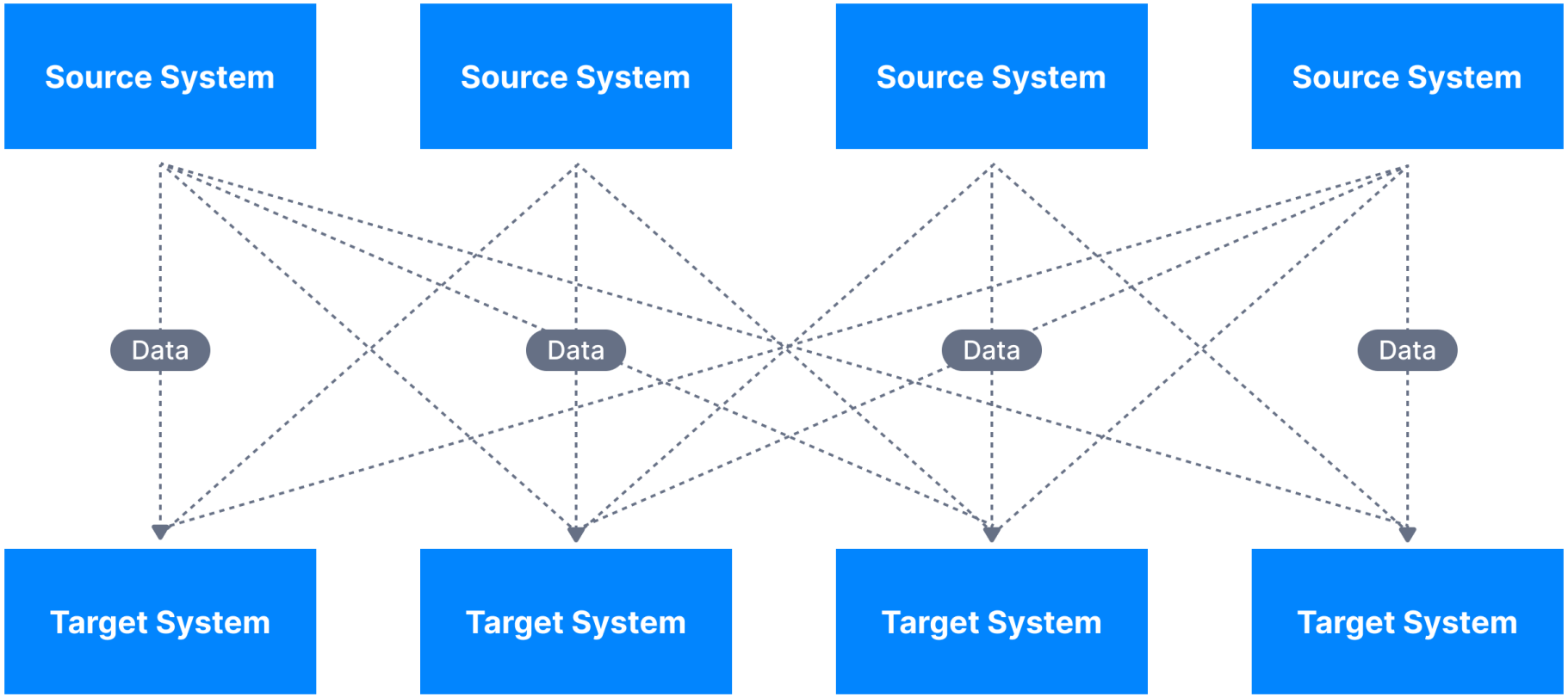


تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

مقدمه و معرفی

❑ آپاچی کافکا (Apache Kafka)



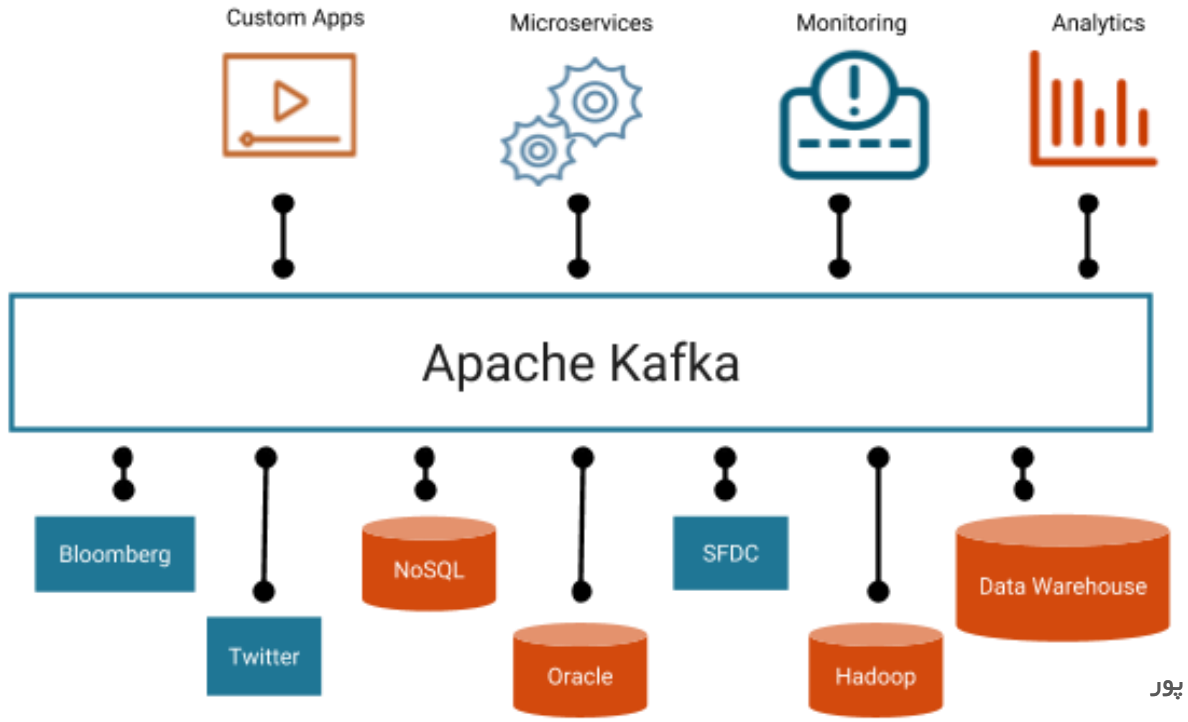
تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

مقدمه و معرفی

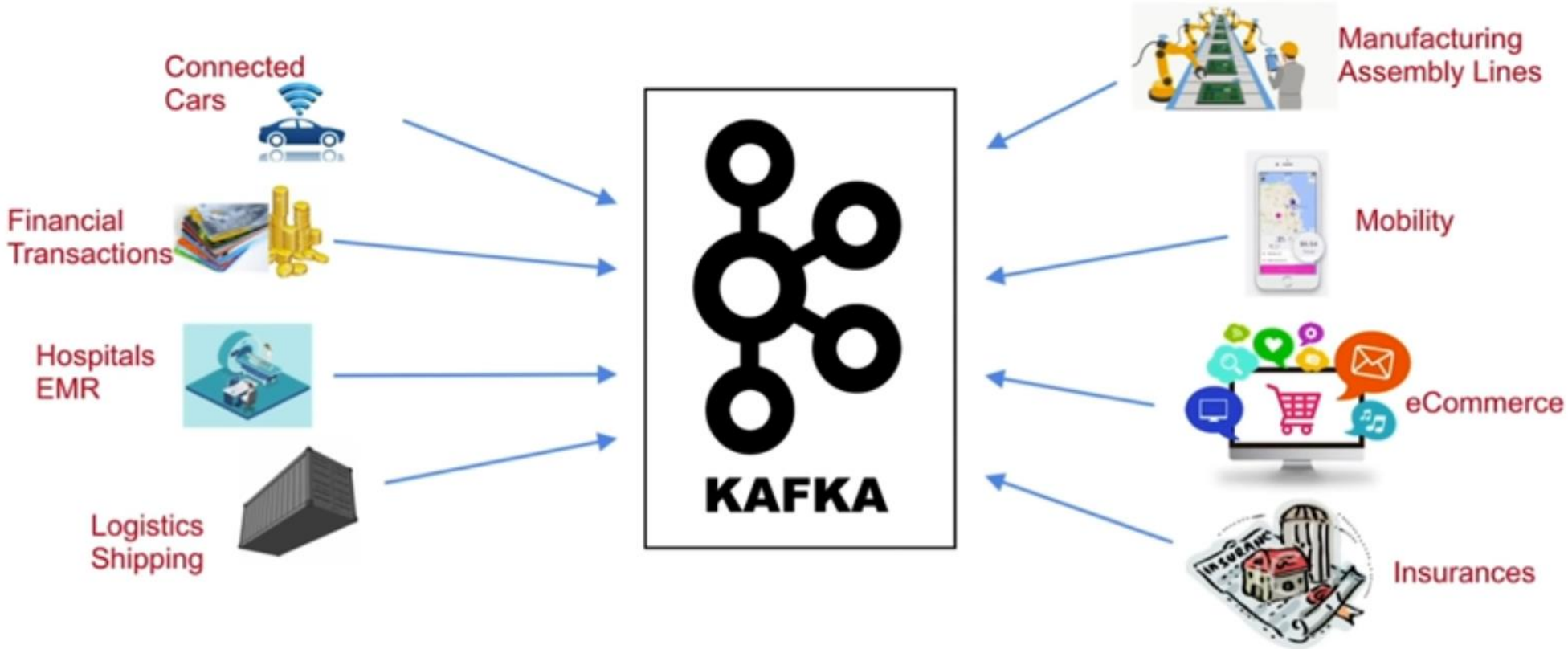
An **event** records the fact that "something happened" in the world or in your business. It is also called record or message in the documentation. When you read or write data to Kafka, you do this in the form of events. Conceptually, an event has a key, value, timestamp, and optional metadata headers. Here's an example event:

- Event key: "Alice"
- Event value: "Made a payment of $200 to Bob"
- Event timestamp: "Jun. 25, 2020 at 2:06 p.m."

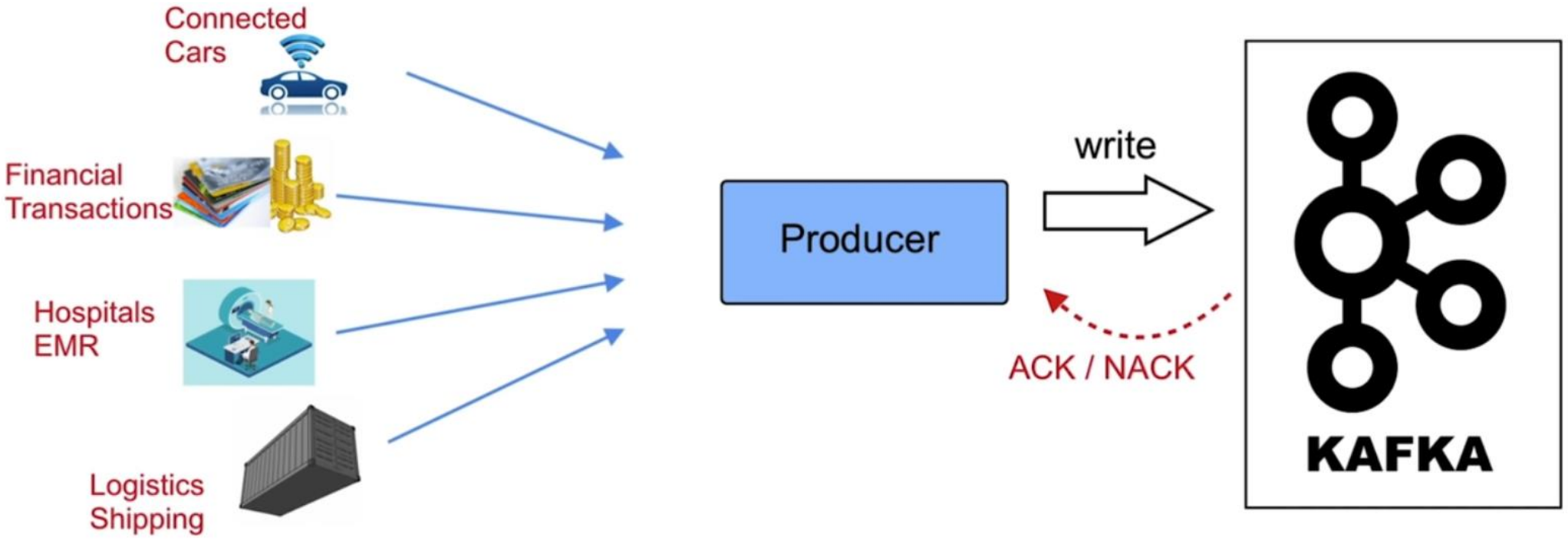


Kafka is a distributed system consisting of **servers** and **clients** that communicate via a high-performance TCP network protocol.
It can be deployed on bare-metal hardware, virtual machines, and containers in on-premise as well as cloud environments.

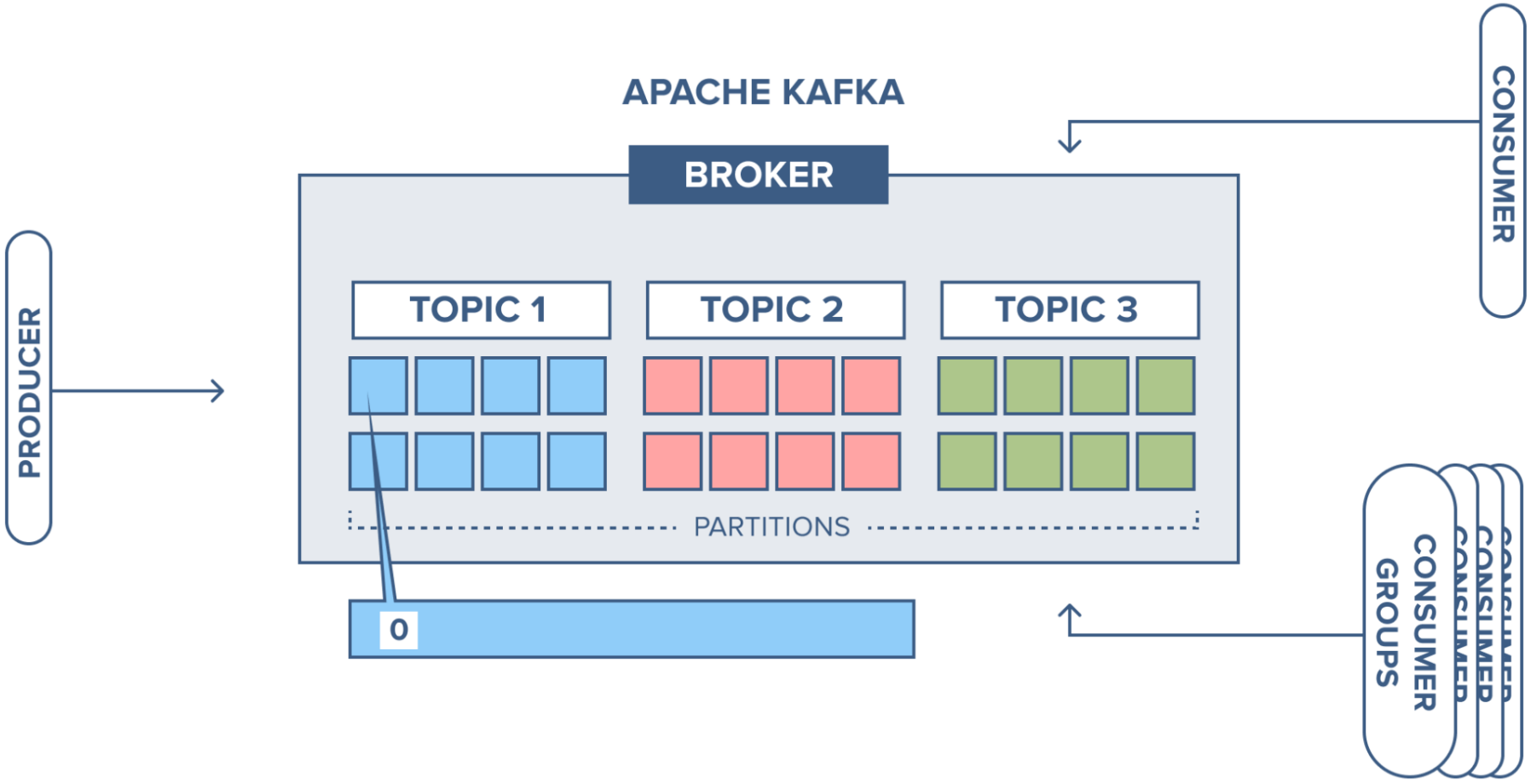
❑ آپاچی کافکا (Apache Kafka)



❑ نقش Producer در آپاچی کافکا (Apache Kafka)

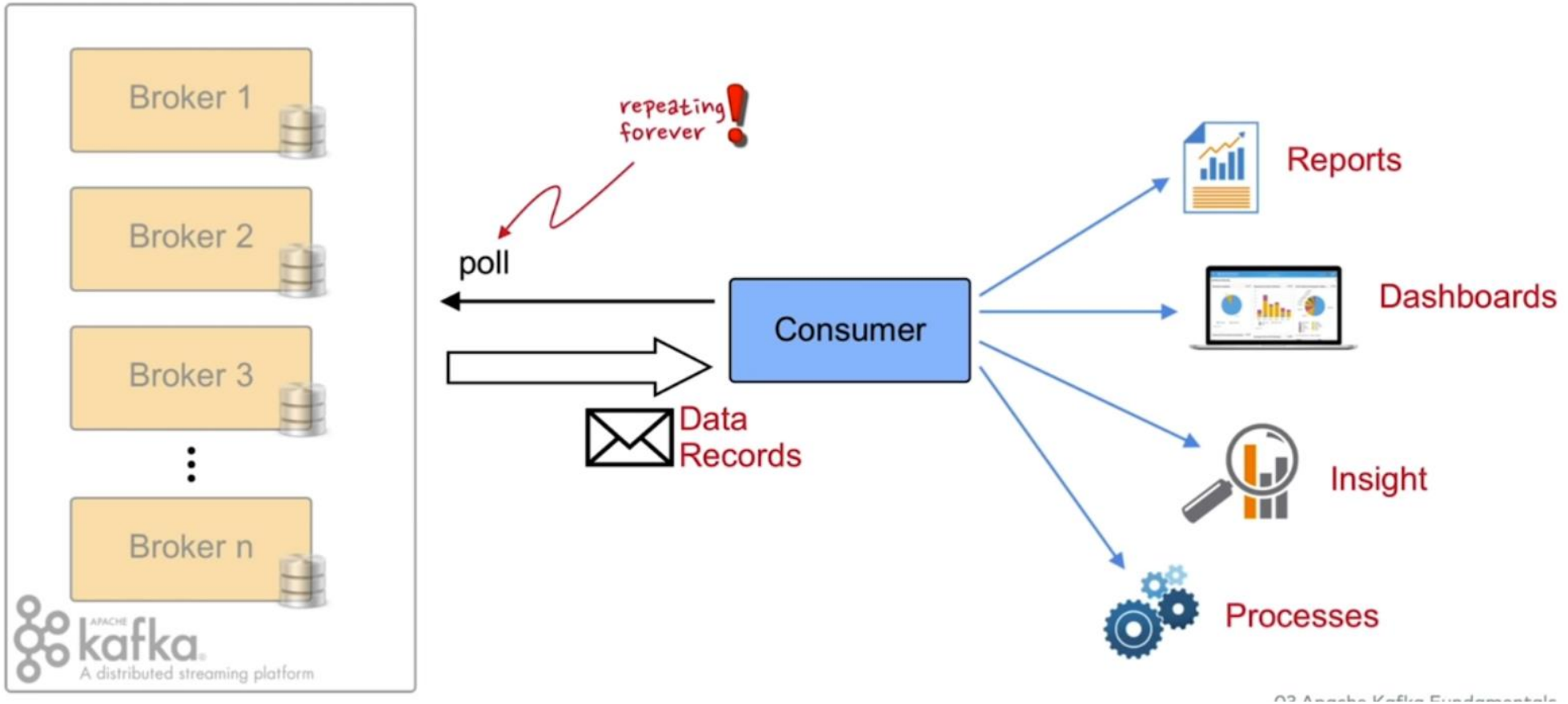


❑ نقش Broker در آپاچی کافکا (Apache Kafka)



تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

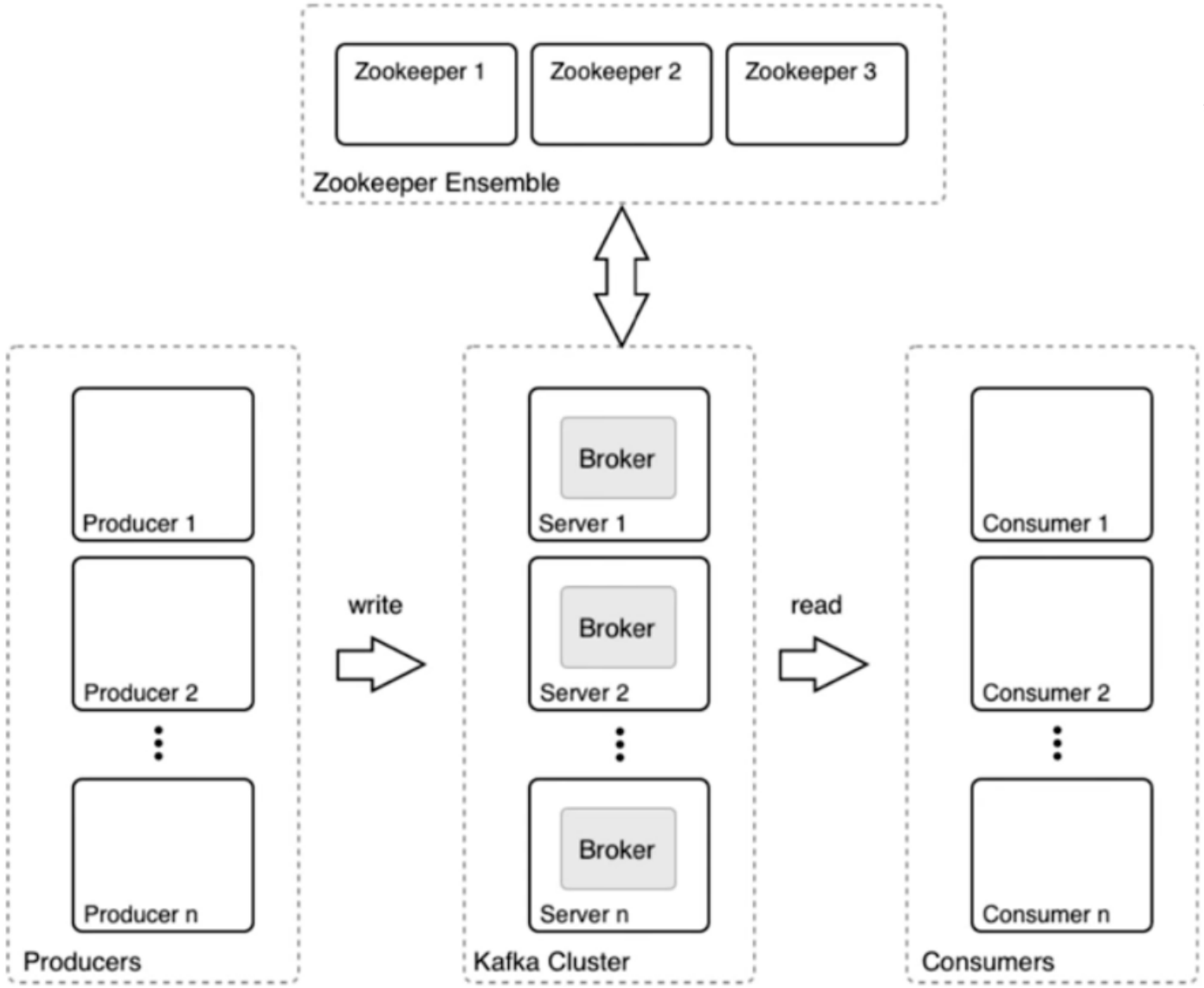
❑ نقش Consumer در آپاچی کافکا (Apache Kafka)



تحلیل کلان داده

مروری بر اسپارک

معماری آپاچی کافکا (Apache Kafka)

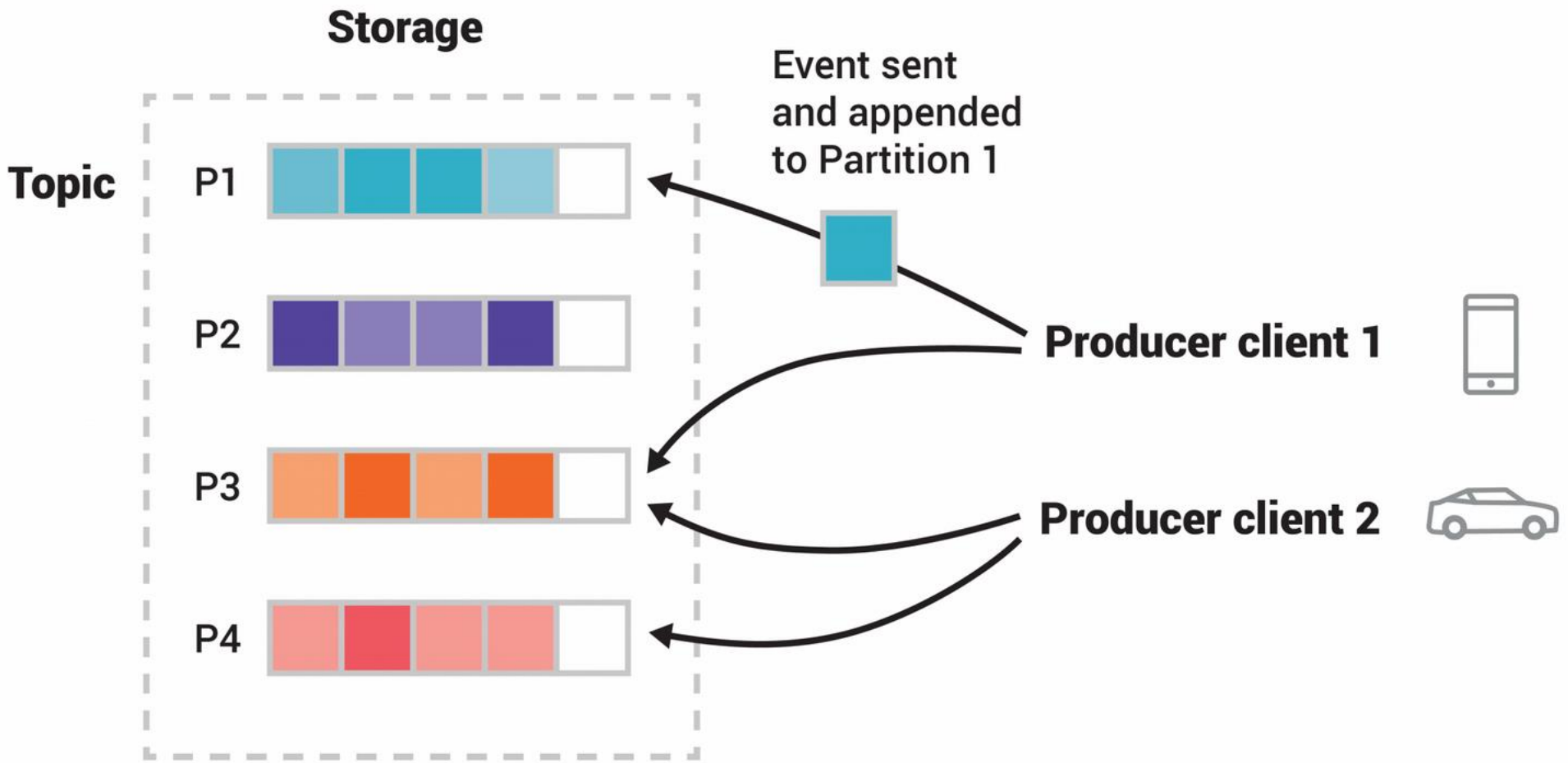


تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

مروری بر اسپارک

- نقش Topic در آپاچی کافکا (Apache Kafka)



تحلیل کلان داده

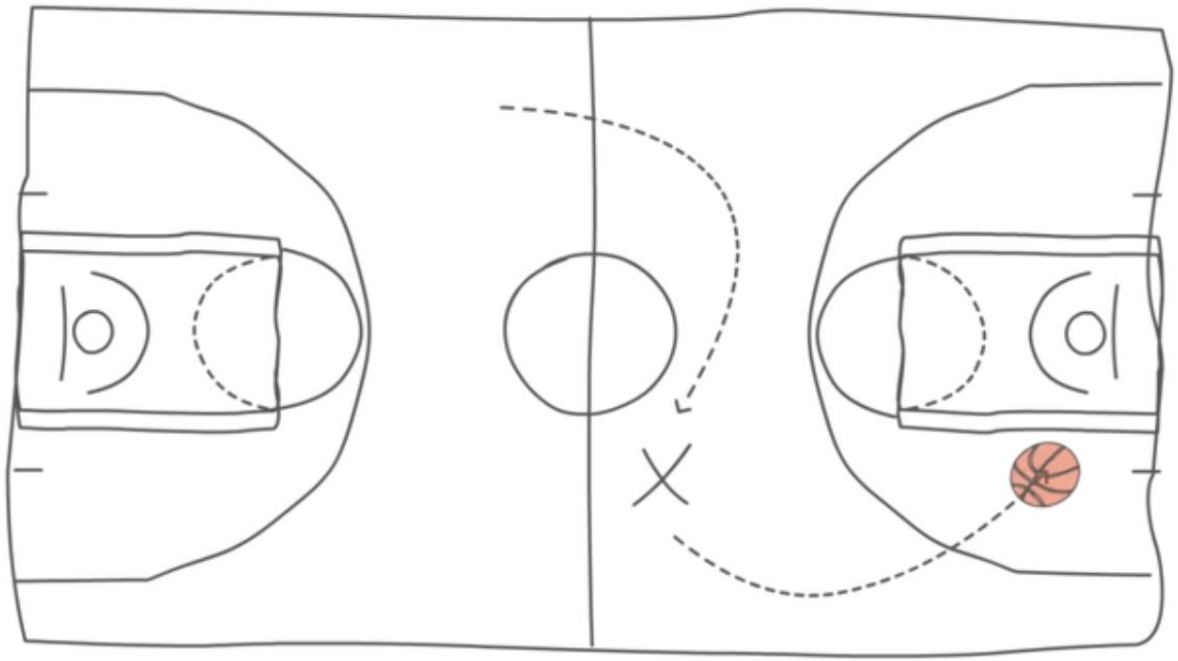
معرفی آپاچی کافکا (سناریو Live Score مسابقات بسکتبال)

تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

معرفی آپاچی کافکا

❑ سناریوی برنامه اعلان نتایج زنده بازی های بسکتبال

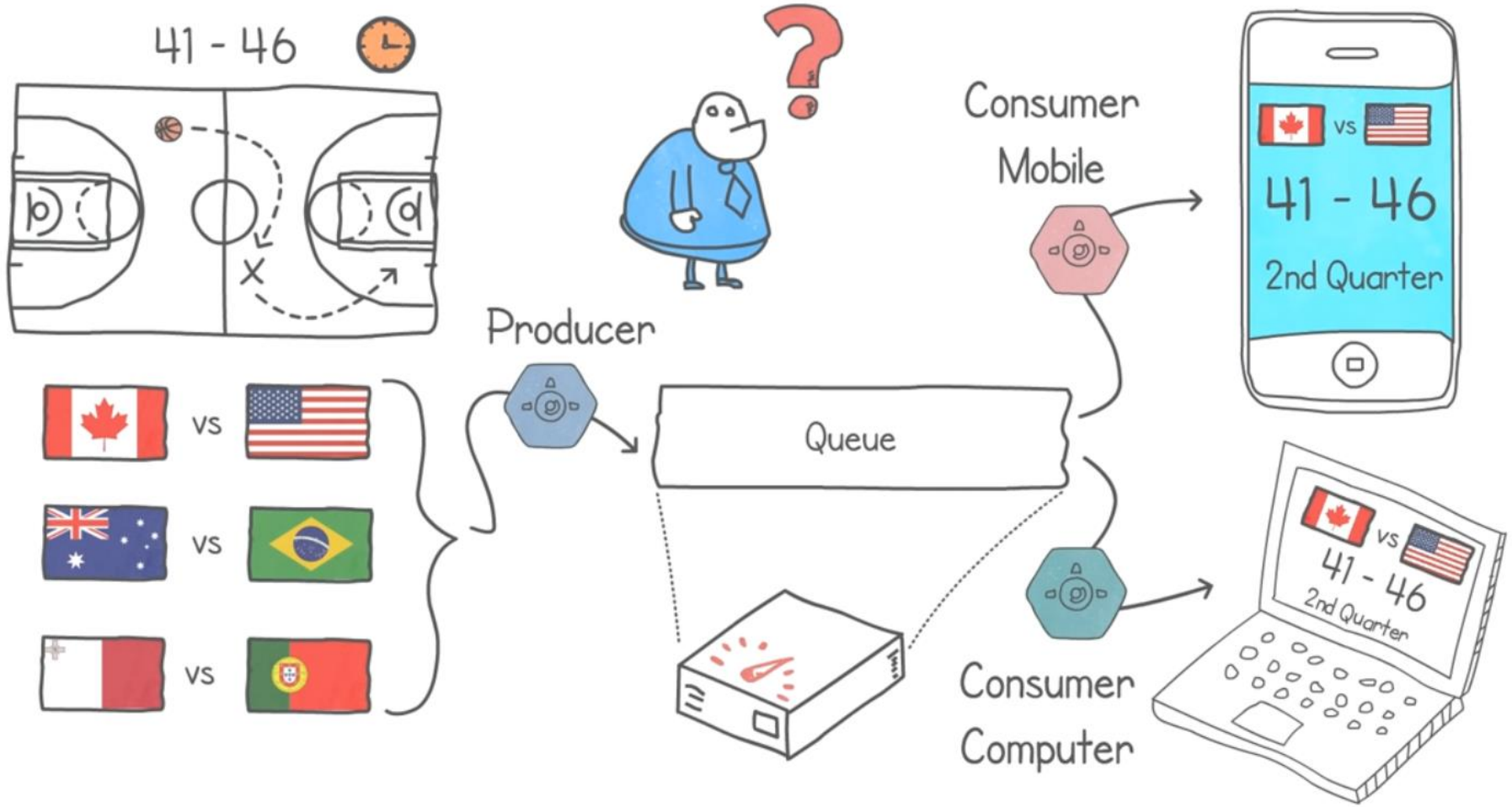


تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

معرفی آپاچی کافکا

❑ سناریوی برنامه اعلان نتایج زنده بازی های بسکتبال

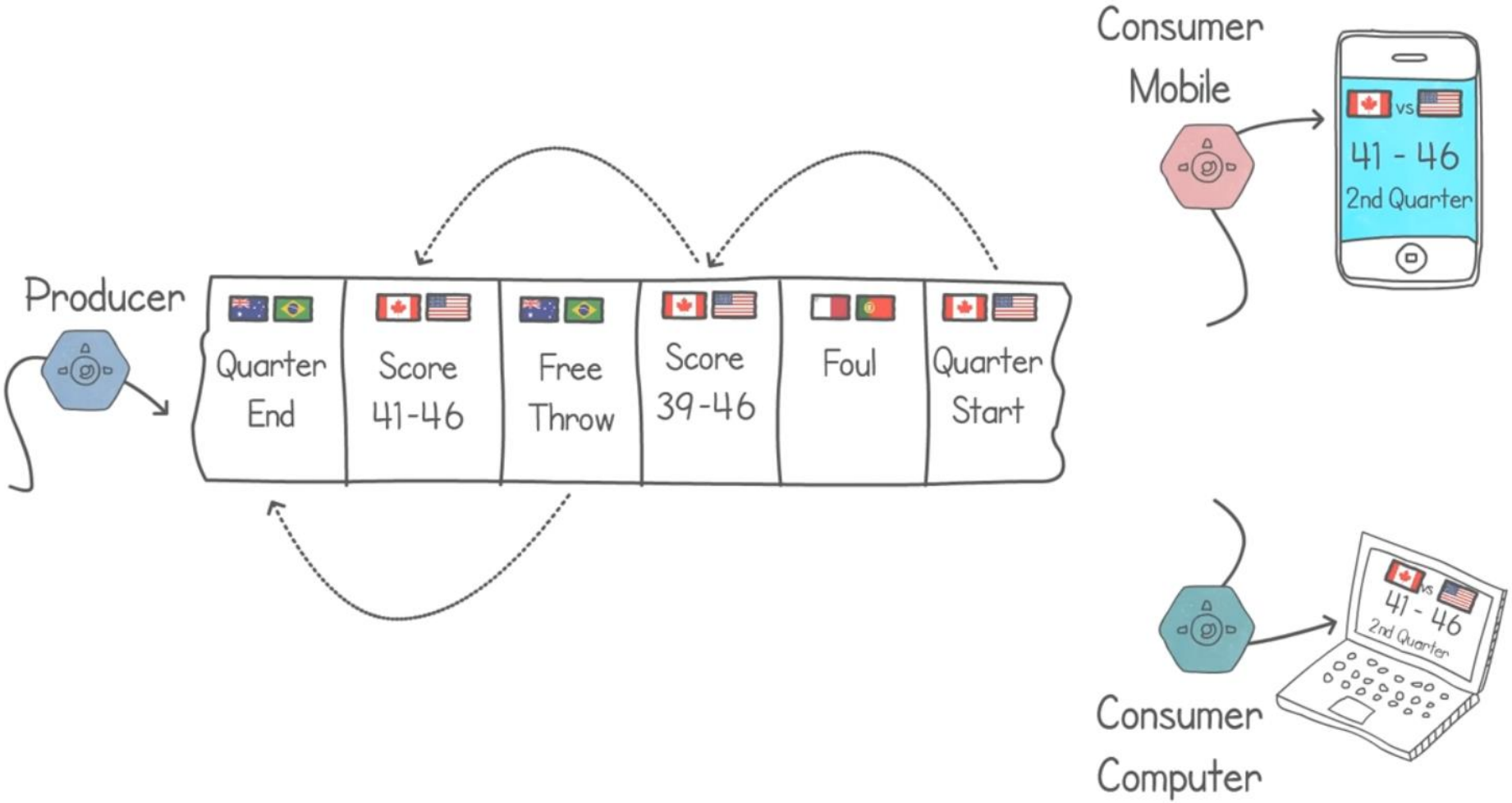


تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

معرفی آپاچی کافکا

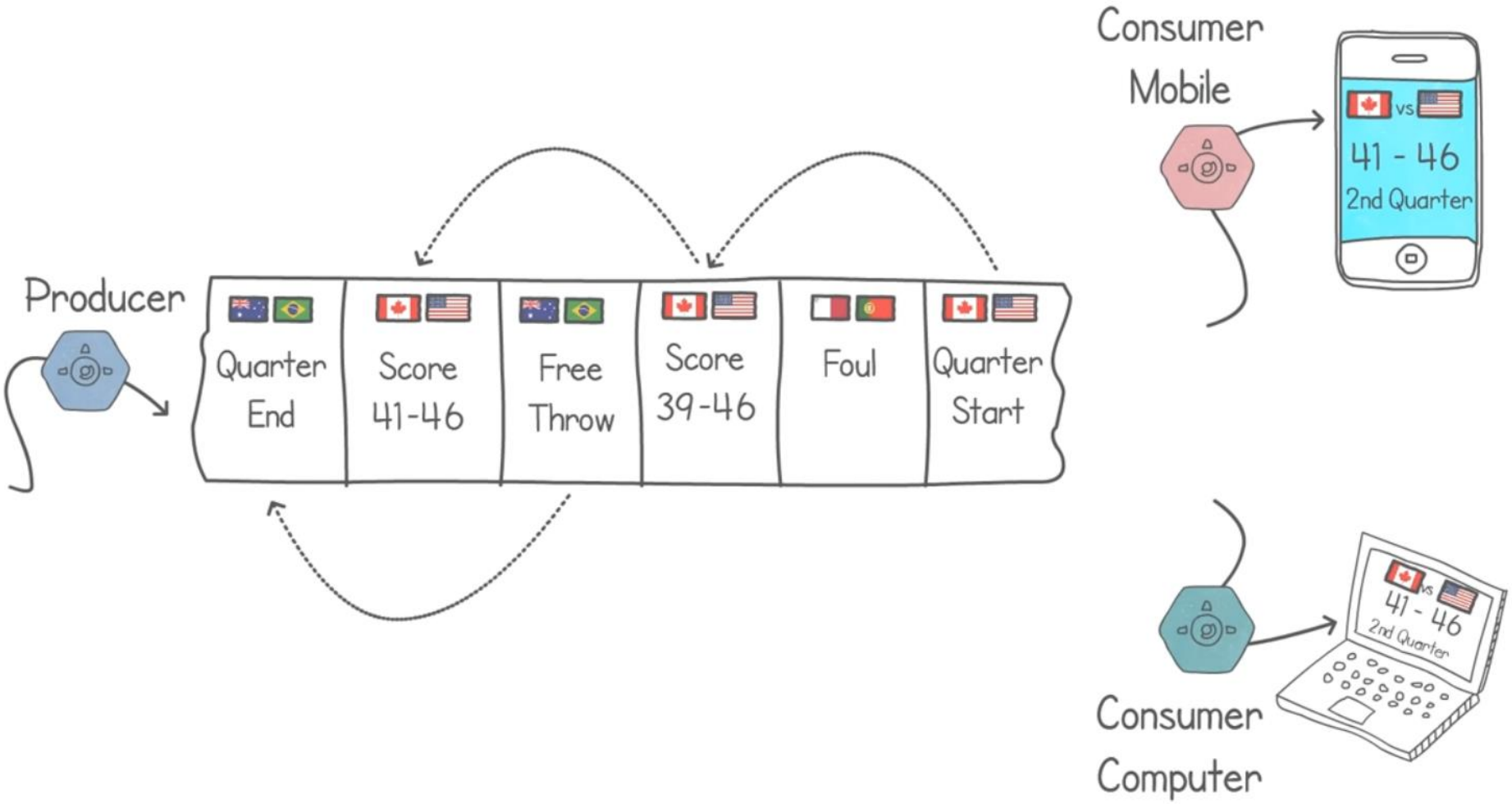
❑ سناریوی برنامه اعلان نتایج زنده بازی های بسکتبال



تحلیل کلان داده

معرفی آپاچی کافکا

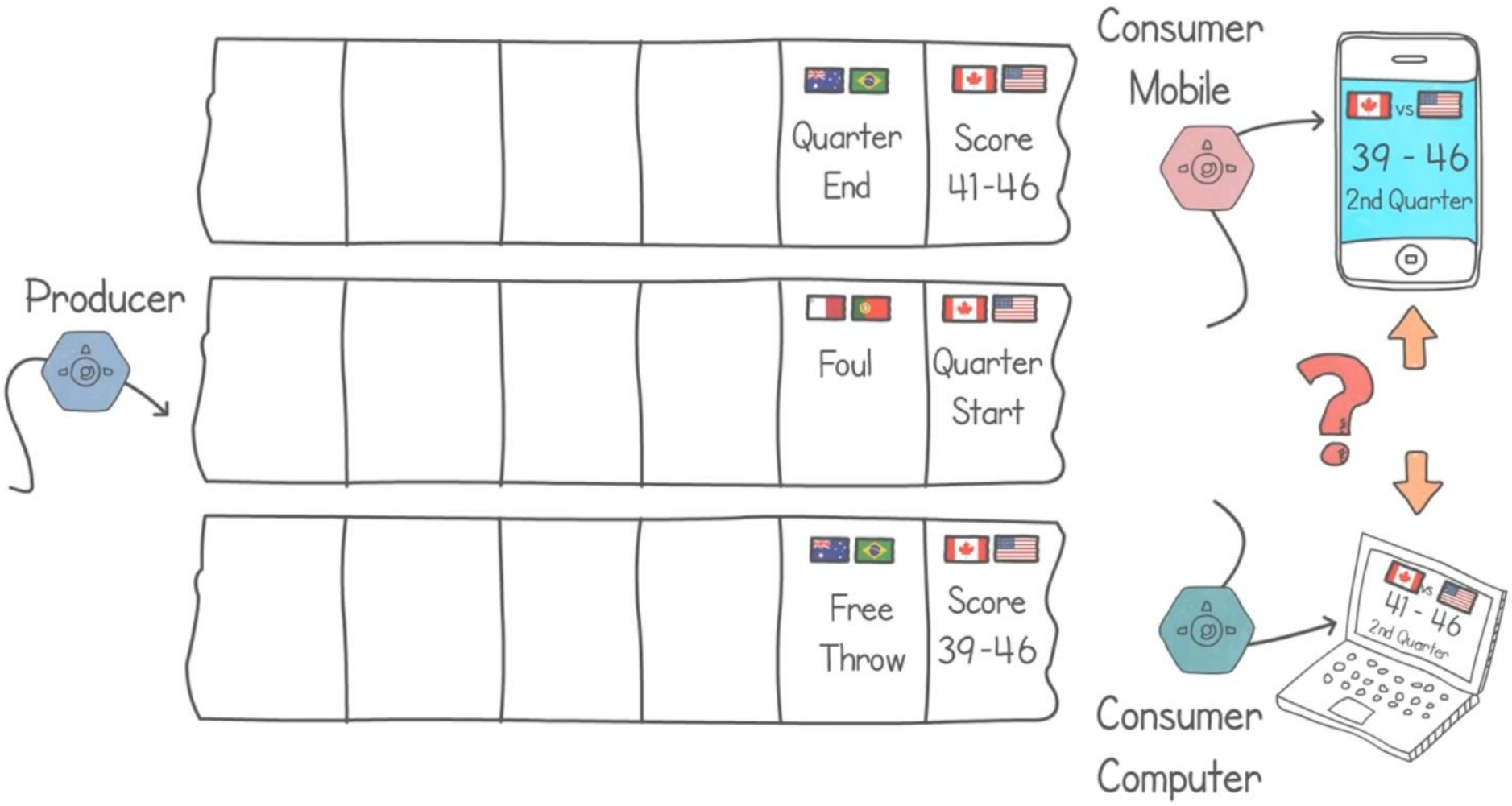
❑ سناریوی برنامه اعلان نتایج زنده بازی های بسکتبال



تحلیل کلان داده

معرفی آپاچی کافکا

❑ سناریوی برنامه اعلان نتایج زنده بازی های بسکتبال

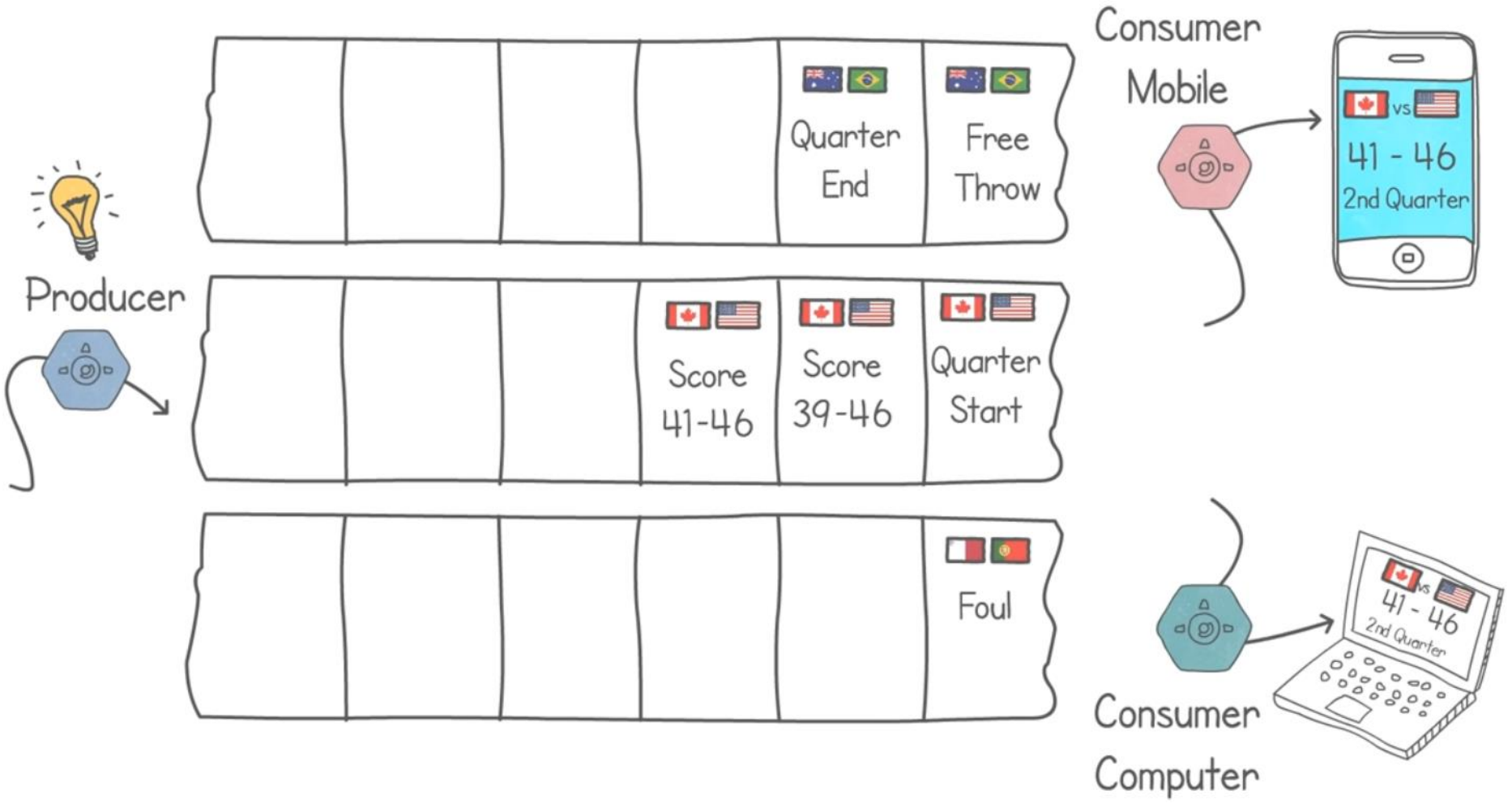


تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

معرفی آپاچی کافکا

❑ سناریوی برنامه اعلان نتایج زنده بازی های بسکتبال

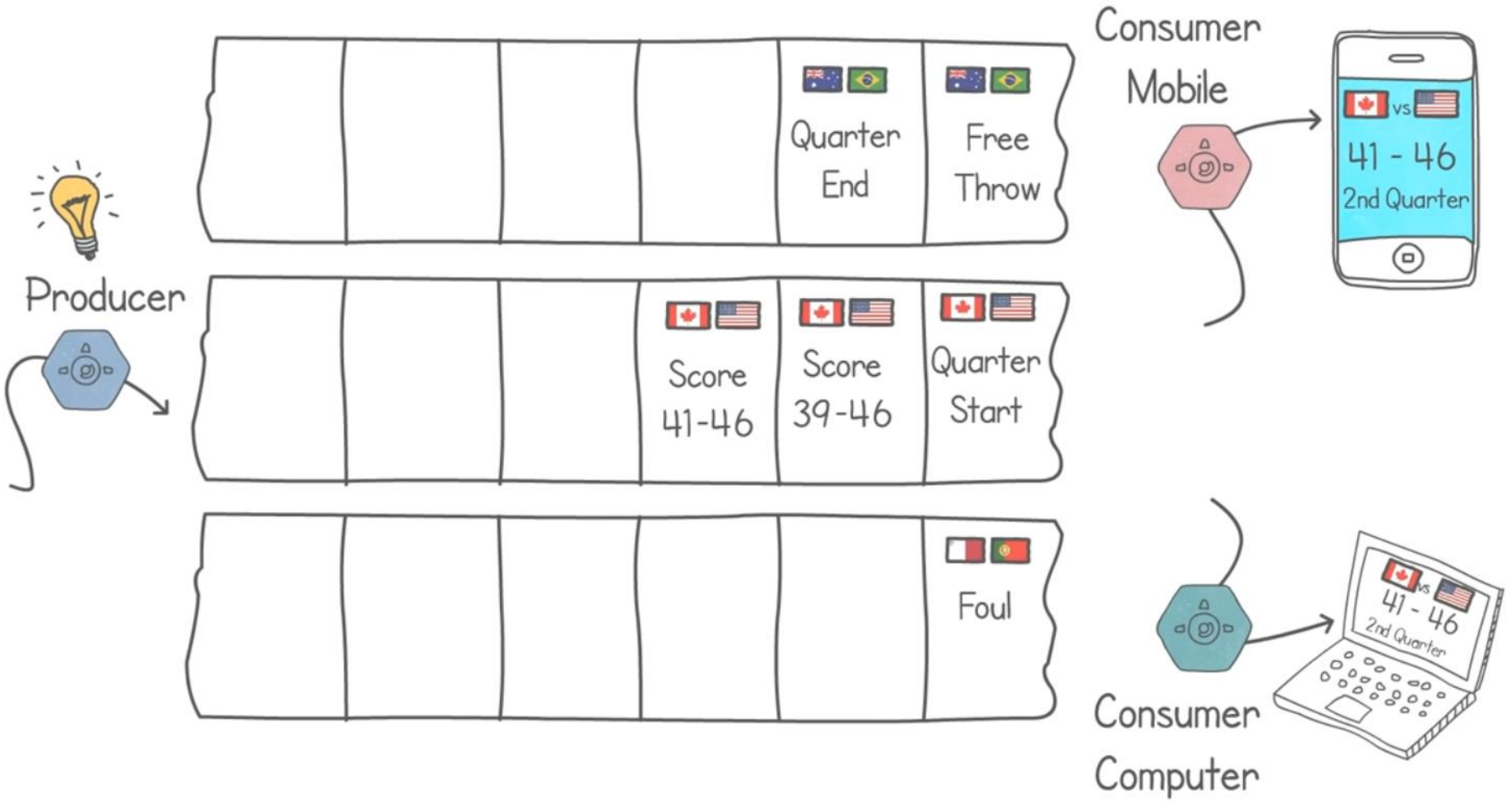


تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

معرفی آپاچی کافکا

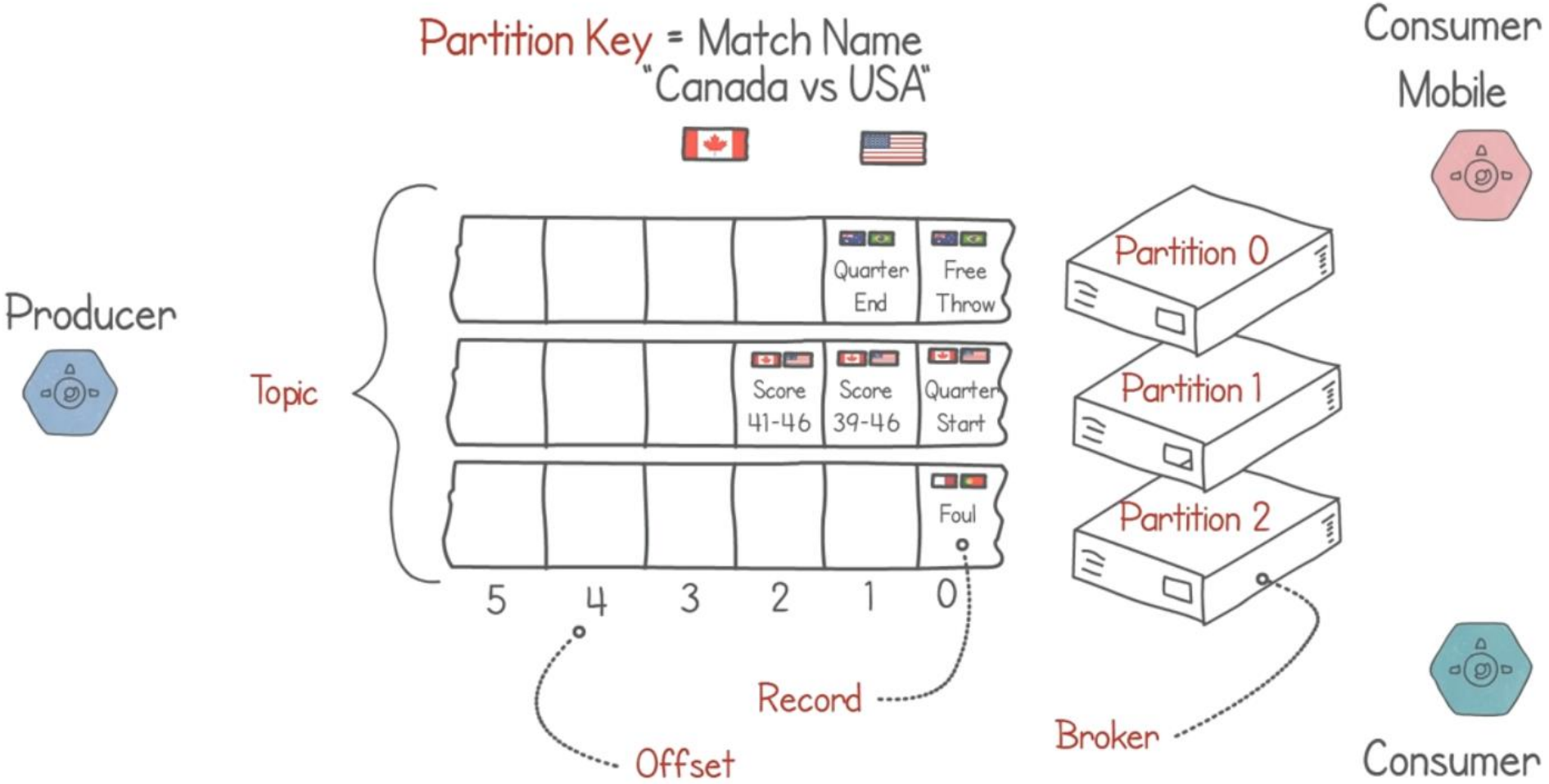
❑ سناریوی برنامه اعلان نتایج زنده بازی های بسکتبال



تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

معرفی آپاچی کافکا



تولید محتوا: سهیل تهرانی پور

daychegroup

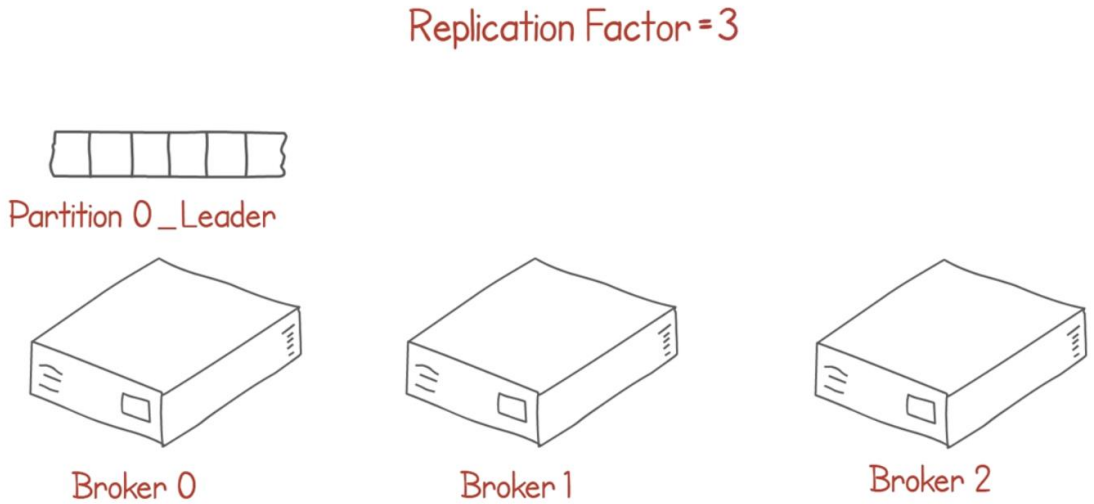
daychegroup

dayche.com | گروه دایکه

تحلیل کلان داده

معرفی آپاچی کافکا

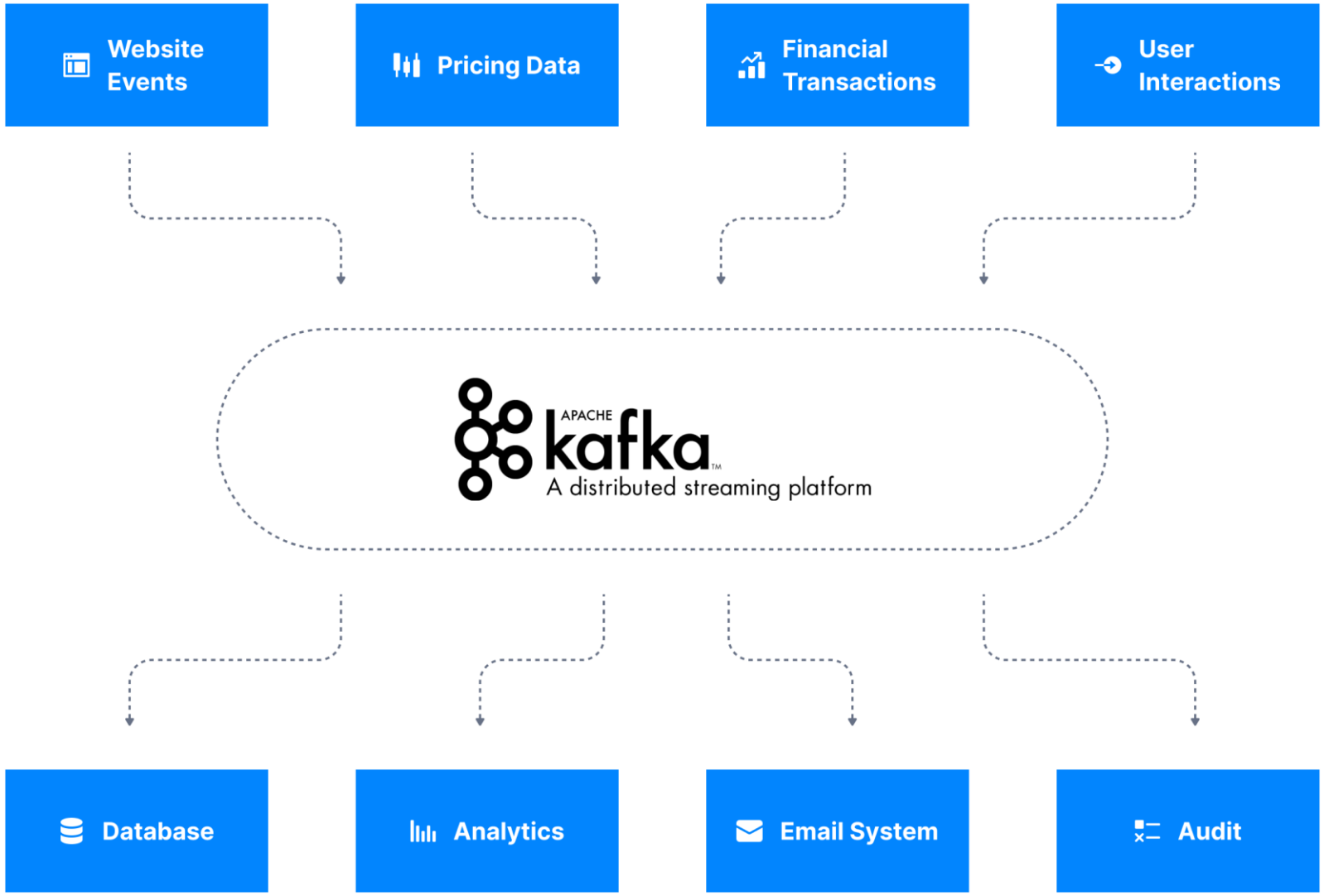
FAULT TOLERANT
&
DURABLE



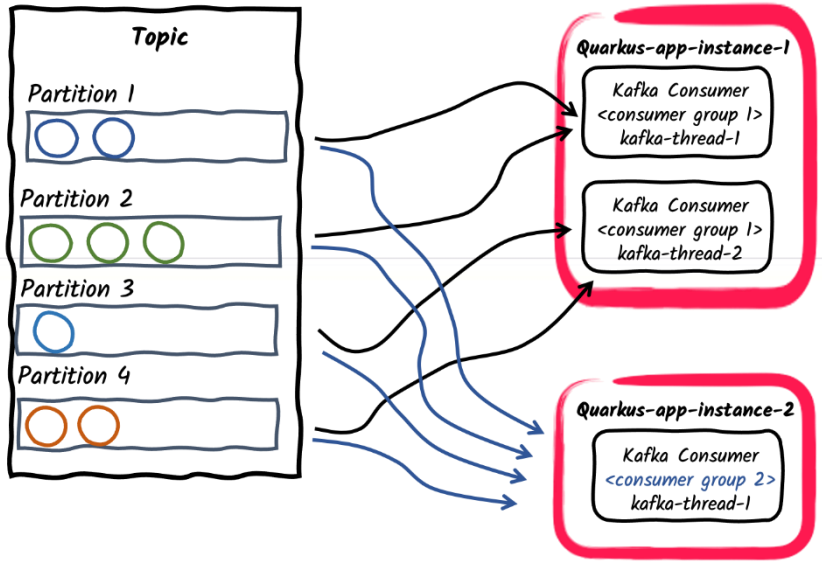
تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه

تحلیل کلان داده

معرفی آپاچی کافکا



تولید محتوا: سهیل تهرانی پور
daychegroup
daychegroup
dayche.com | گروه دایکه



Messaging

Kafka works well as a replacement for a more traditional message broker. Message brokers are used for a variety of reasons (to decouple processing from data producers, to buffer unprocessed messages, etc). In comparison to most messaging systems Kafka has better throughput, built-in partitioning, replication, and fault-tolerance which makes it a good solution for large scale message processing applications.

In our experience messaging uses are often comparatively low-throughput, but may require low end-to-end latency and often depend on the strong durability guarantees Kafka provides.

In this domain Kafka is comparable to traditional messaging systems such as ActiveMQ or RabbitMQ.