Presentation of Apache Kafka

By Xenias Dimitrios

Agenda

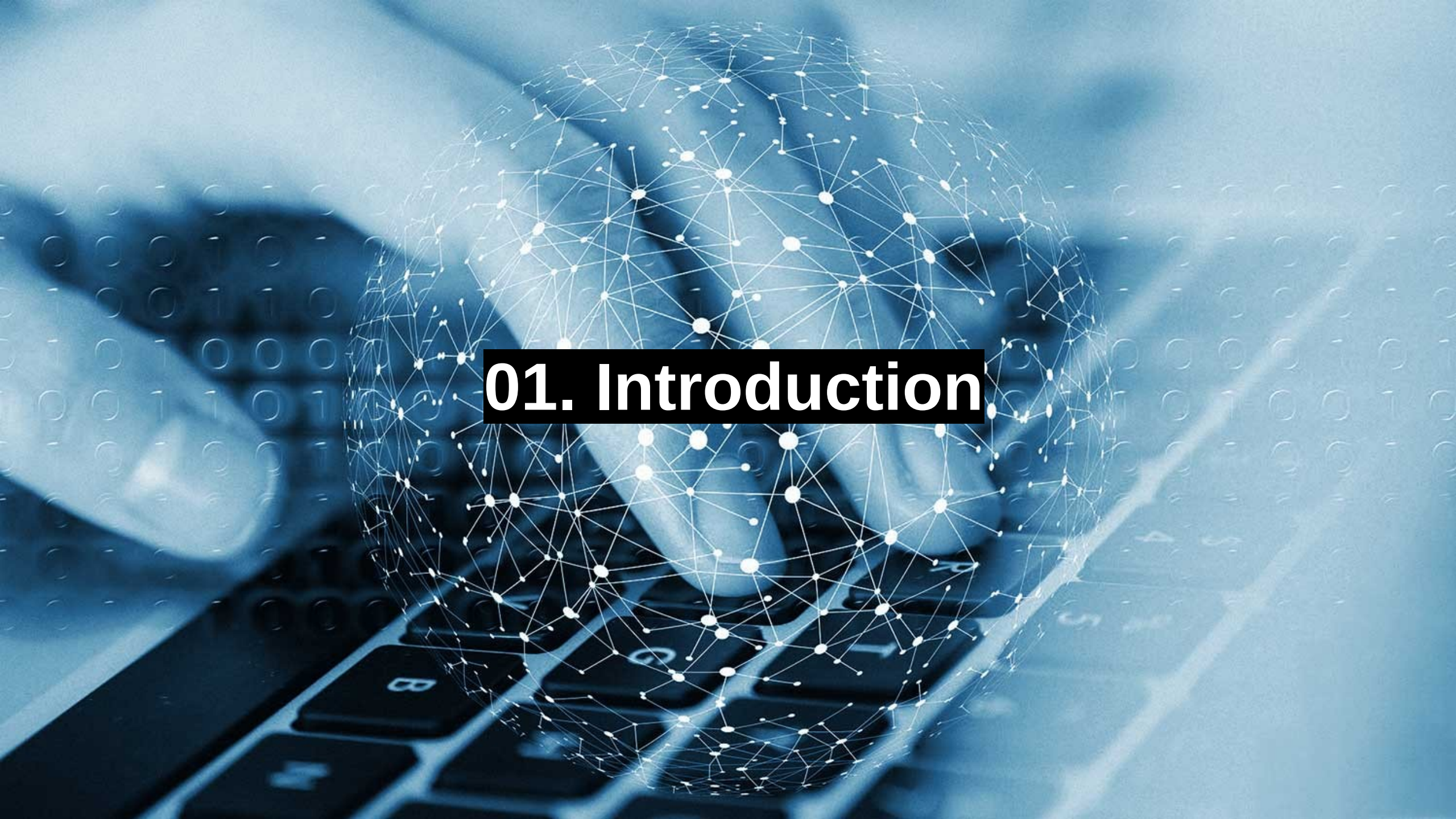
01 Introduction

02 What Kafka is

03 Kafka Components

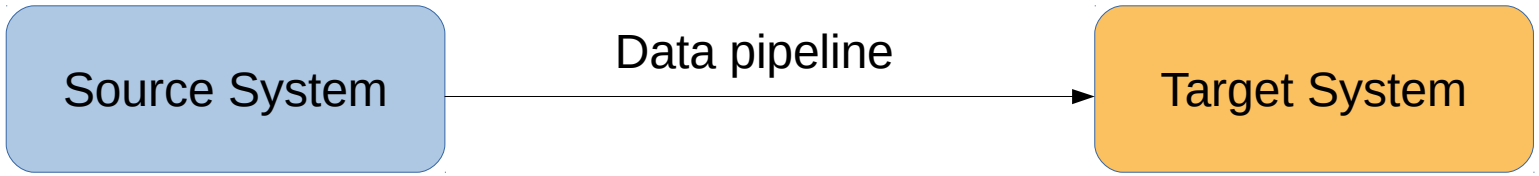
04 Kafka Architecture

By Xenias Dimitrios


01. Introduction

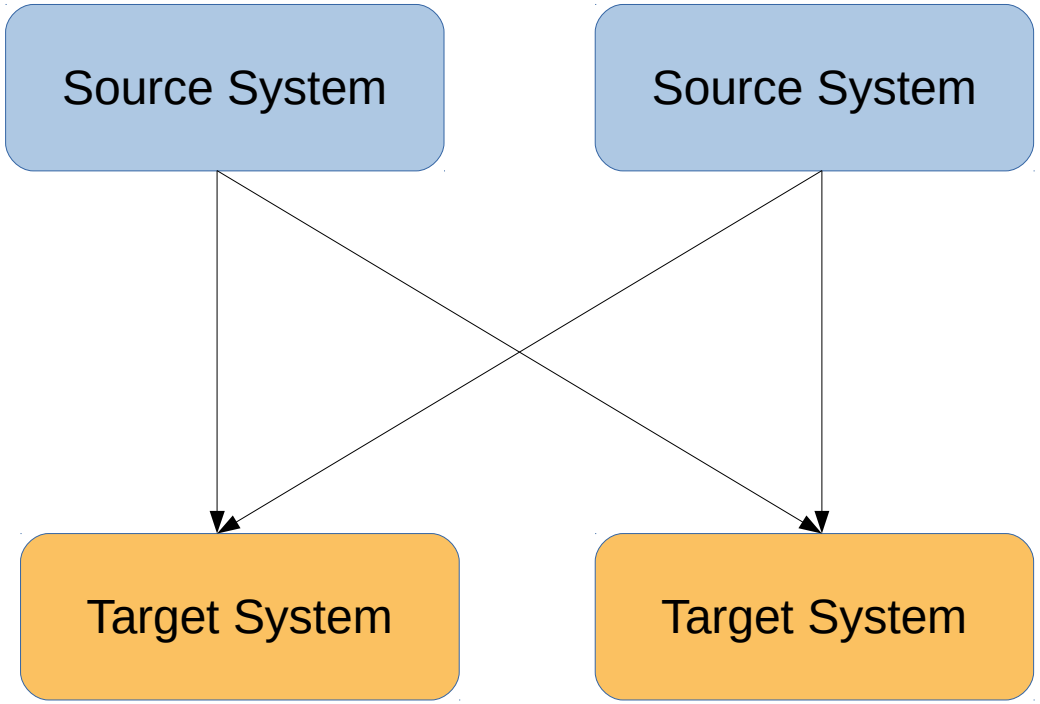
01. Introduction

- How companies started (‘90s-early ‘00s)

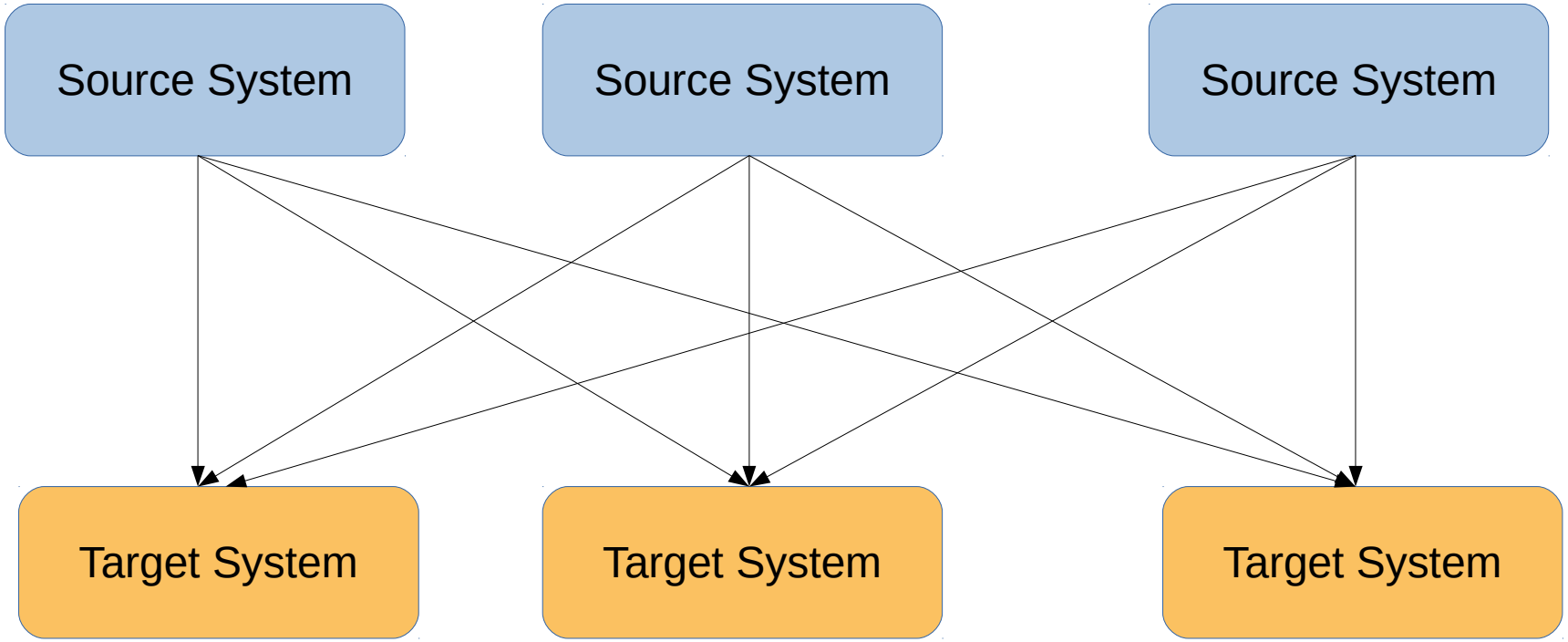


01. Introduction

- After a while



01. Introduction

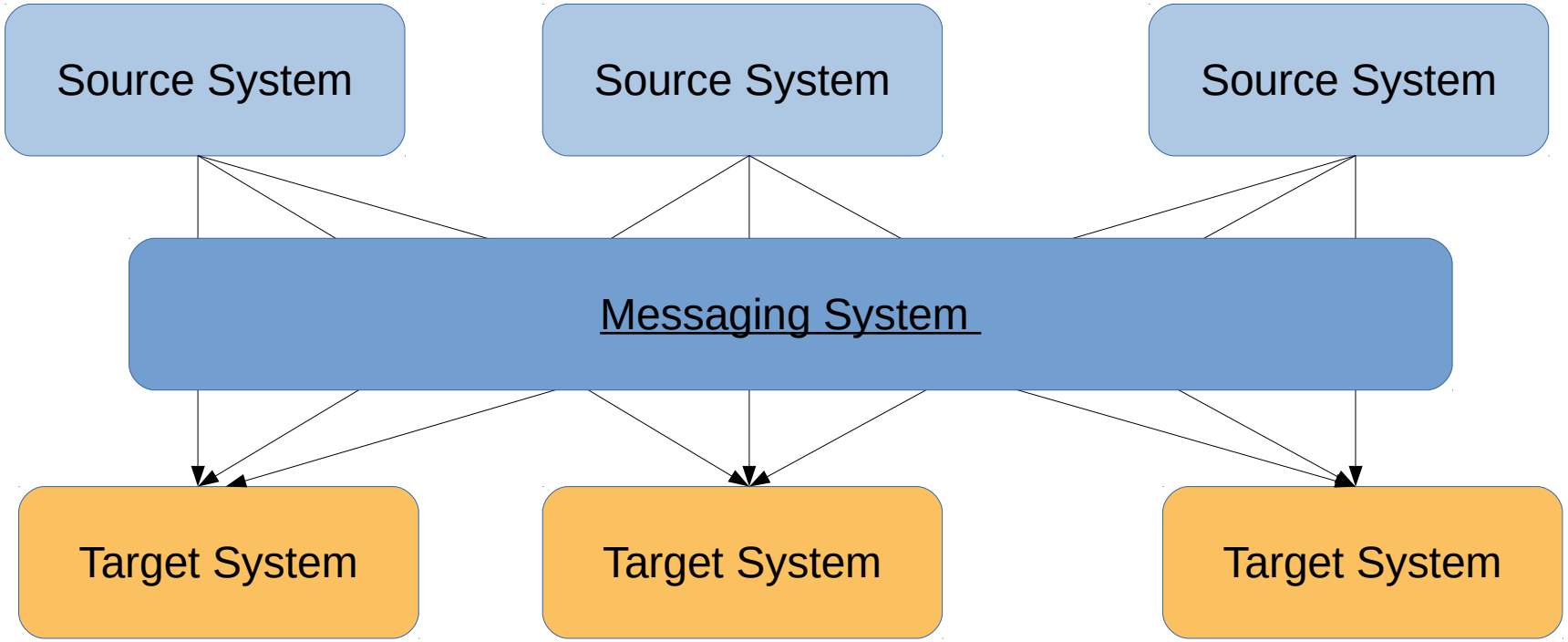


#Integrations = #Source Systems * #Target Systems

By Xenias Dimitrios

01. Introduction

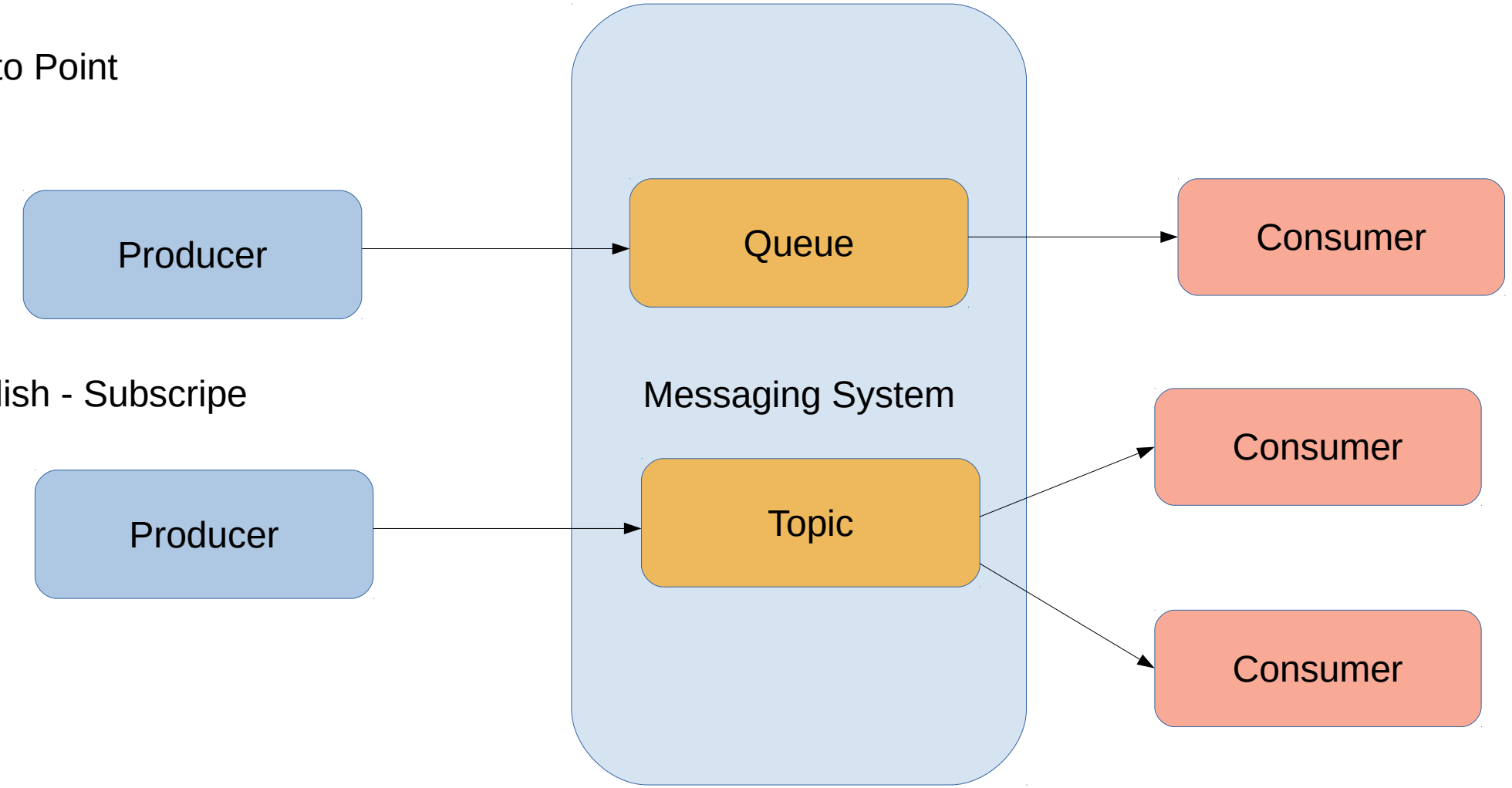
- Solution to Complex Data Pipelines



01. Introduction

- Messaging Systems

1) Point to Point

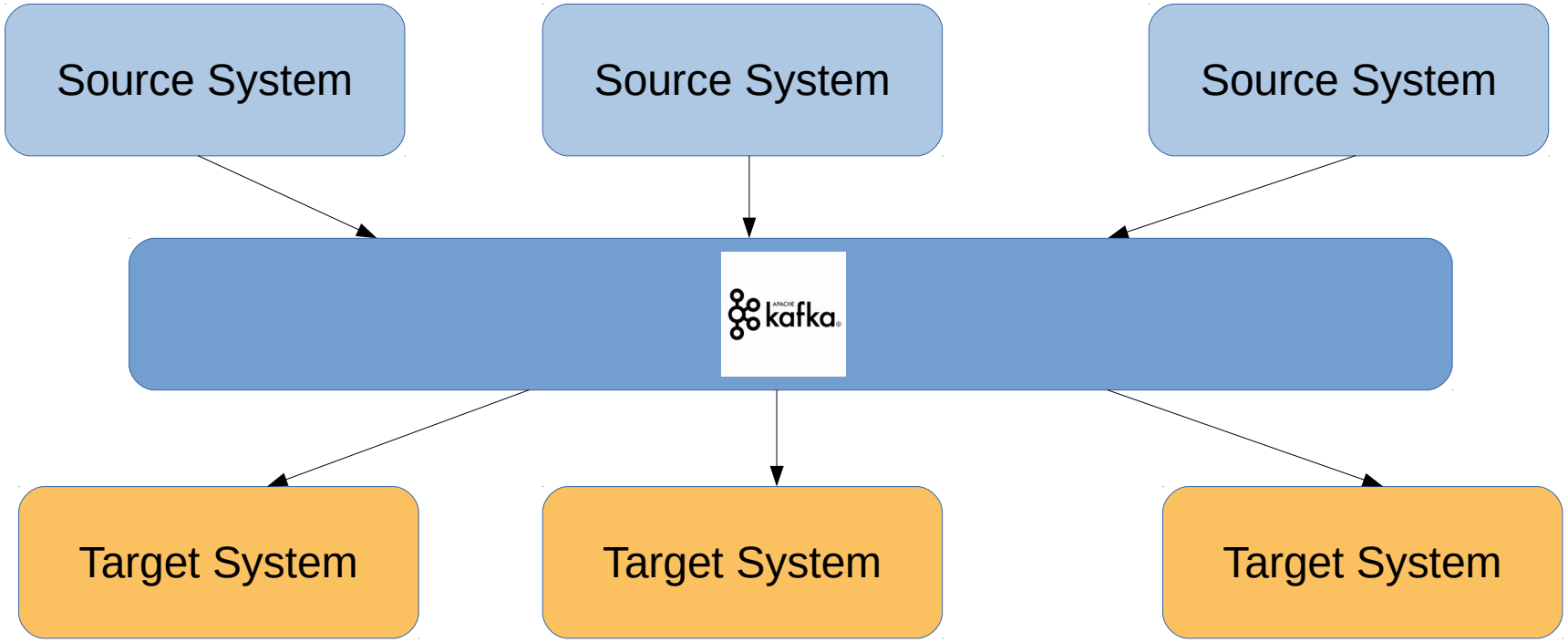


2) Publish - Subscripe

By Xenias Dimitrios

01. Introduction

- Here comes Kafka



02. What Kafka is

02. What Kafka is

- Apache Kafka® is a distributed *publish-subscribe* messaging system as streaming platform.
- It was originally developed at LinkedIn (2011) and later became a part of Apache Project.
- Kafka is fast, scalable, durable, fault-tolerant and distributed by design.
- Kafka is generally used for two broad classes of applications:
  - Building real-time streaming data pipelines that reliably get data between systems or applications
  - Building real-time streaming applications that transform or react to the streams of data

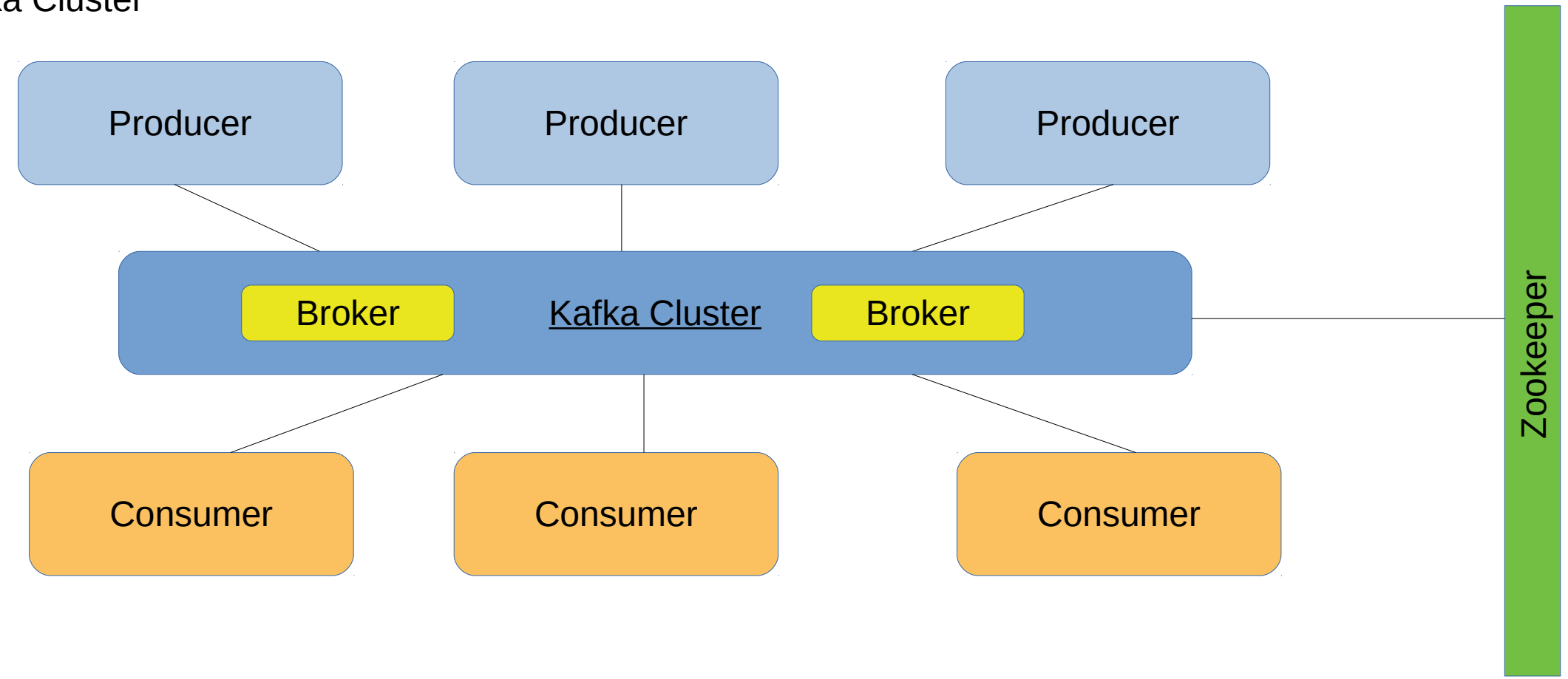
02. What Kafka is

First a few concepts:

- Topic: Is a category to which records are published
- Partition: Topics are broken up into ordered commit logs called partitions
- Producer: Any application who can publish messages to a topic
- Consumer: Any application that subscribes to a topic and consume messages
- Broker: Kafka cluster is a set of server, each of which is called a broker
- ZooKeeper: Is used for managing and coordinating Kafka Cluster

02. What Kafka is

- Kafka Cluster



By Xenias Dimitrios

02. What Kafka is

Kafka Features

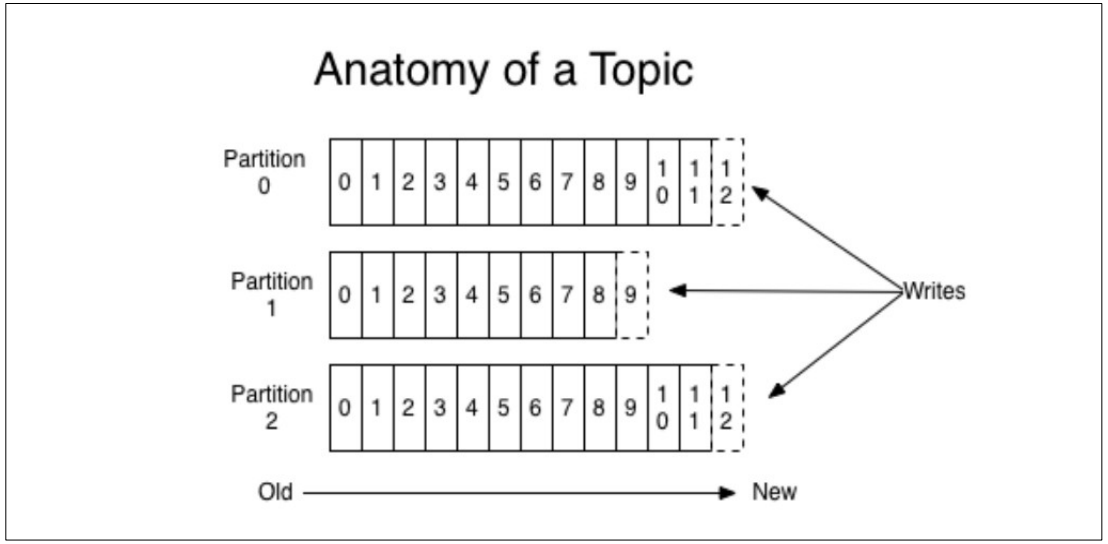
- High Throughput
- Scalability
- Replication and Fault-Tolerant
- Durability

03. Kafka Components

03. Kafka Components

Topics and Partitions

- Topics in Kafka are always multi-subscriber; that is, a topic can have zero, one, or many consumers that subscribe to the data written to it
- For each topic, the Kafka cluster maintains a partitioned log
- Each partition is an ordered, immutable sequence of records that is continually appended to—a structured commit log. The records in the partitions are each assigned a sequential id number called the offset that uniquely identifies each record within the partition.



By Xenias Dimitrios

03. Kafka Components

Topics and Partitions and Replicas

- The partitions of the log are distributed over the servers in the Kafka cluster with each server handling data and requests for a share of the partitions. Each partition is replicated across a configurable number of servers for fault tolerance.

- Each partition has one server which acts as the "leader" and zero or more servers which act as "followers". The leader handles all read and write requests for the partition while the followers passively replicate the leader. If the leader fails, one of the followers will automatically become the new leader. Each server acts as a leader for some of its partitions and a follower for others so load is well balanced within the cluster.

03. Kafka Components

Producer

- The producer is responsible for choosing which record to assign to which partition within the topic
- This can be done in a round-robin fashion simply to balance load or it can be done according to some semantic partition function

03. Kafka Components

Consumer

Low-Level Consumer

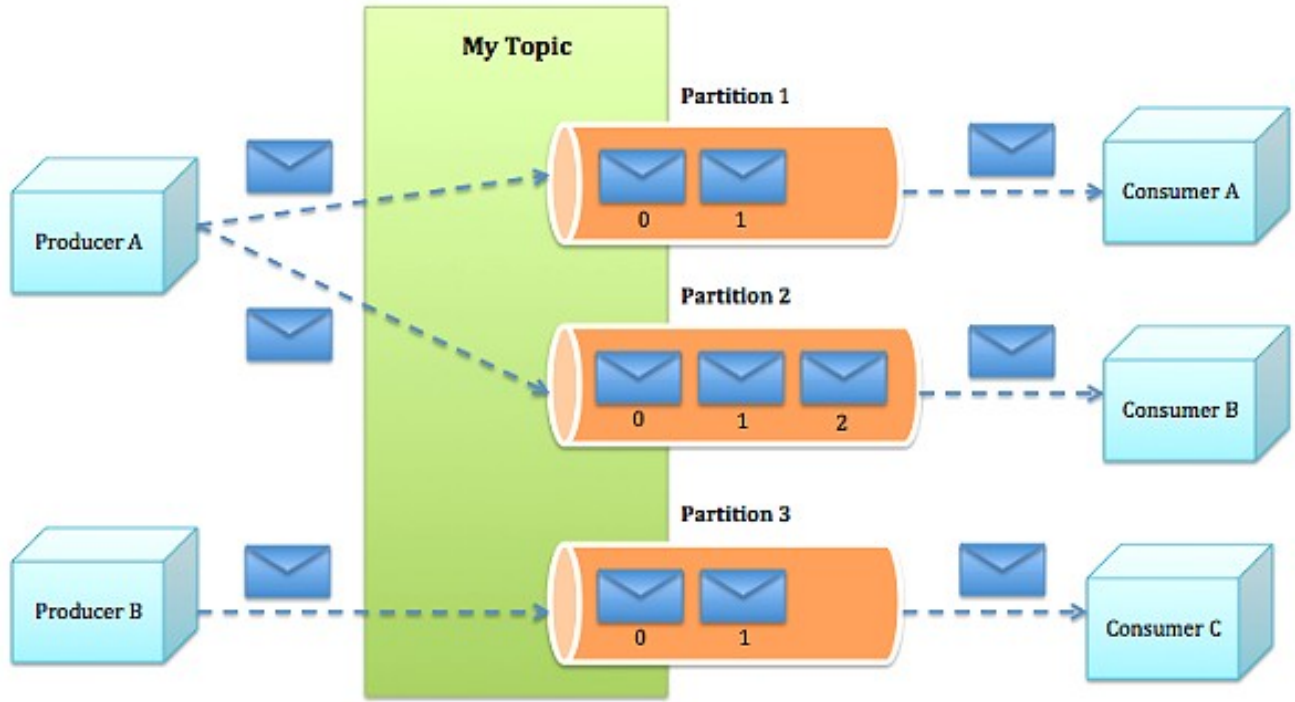
- Topics and partitions are specified as is the offset from which to read, either fixed position, at the beginning or at the end.

High-Level Consumer (Consumer Groups)

- A consumer group consists of several consumer instances
- Each partition of a topic is consumed by one consumer instance within each subscribing consumer group

04. Kafka Architecture

04. Kafka Architecture



Sources

https://kafka.apache.org/intro

https://www.cloudkarafka.com/blog/2016-11-30-part1-kafka-for-beginners-what-is-apache-kafka.html

https://www.youtube.com/watch?v=hyJZP-rgooc

https://www.youtube.com/watch?v=JalUUBKdcA0&t=34s

https://www.youtube.com/watch?time_continue=2448&v=daRykH67_qs&feature=emb_logo


THANK YOU