

Apache Kafka

Kafka Introduction

Apache Kafka® is *a distributed streaming platform*. What exactly does that mean?

A streaming platform has three key capabilities:

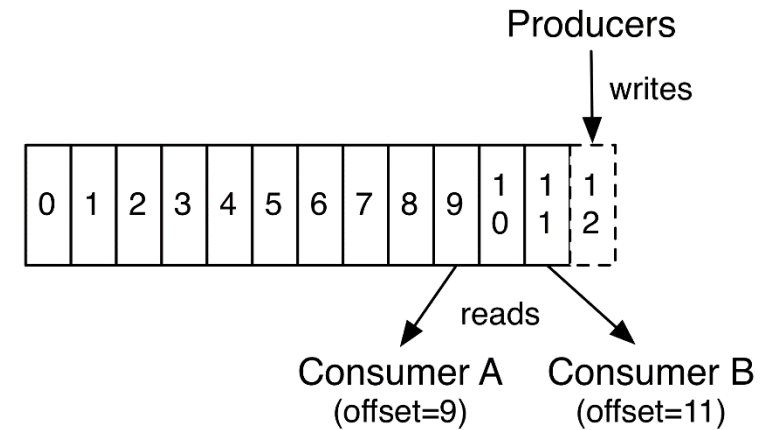
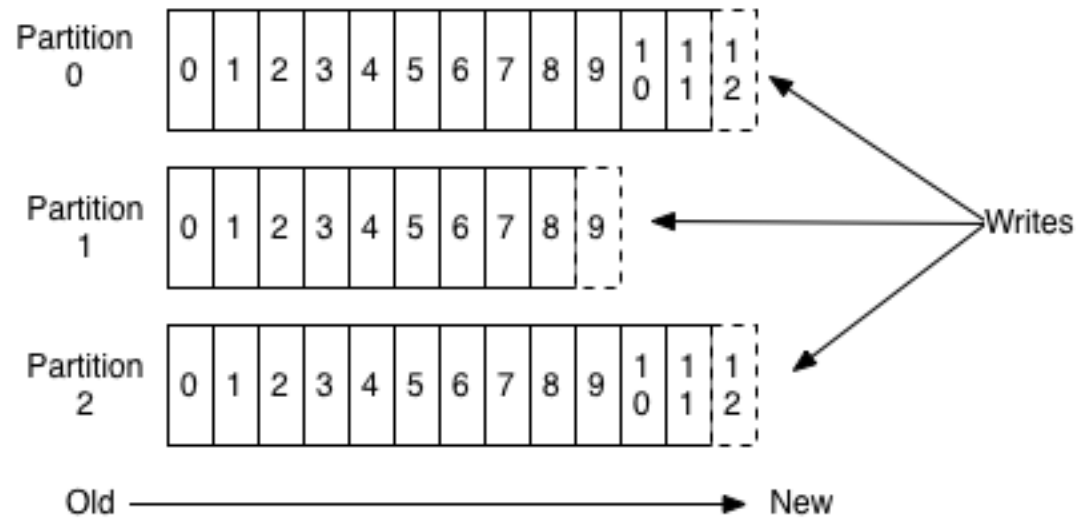
- Publish and subscribe to streams of records, similar to a message queue or enterprise messaging system.
- Store streams of records in a fault-tolerant durable way.
- Process streams of records as they occur.

First a few concepts:

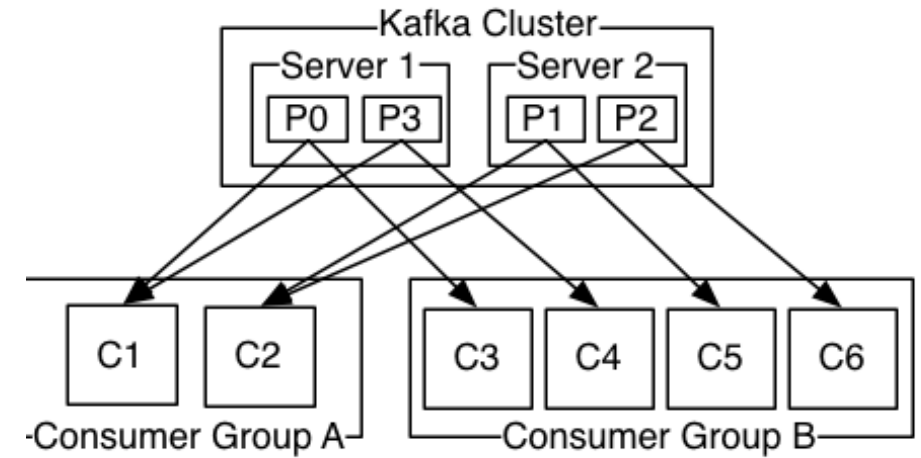
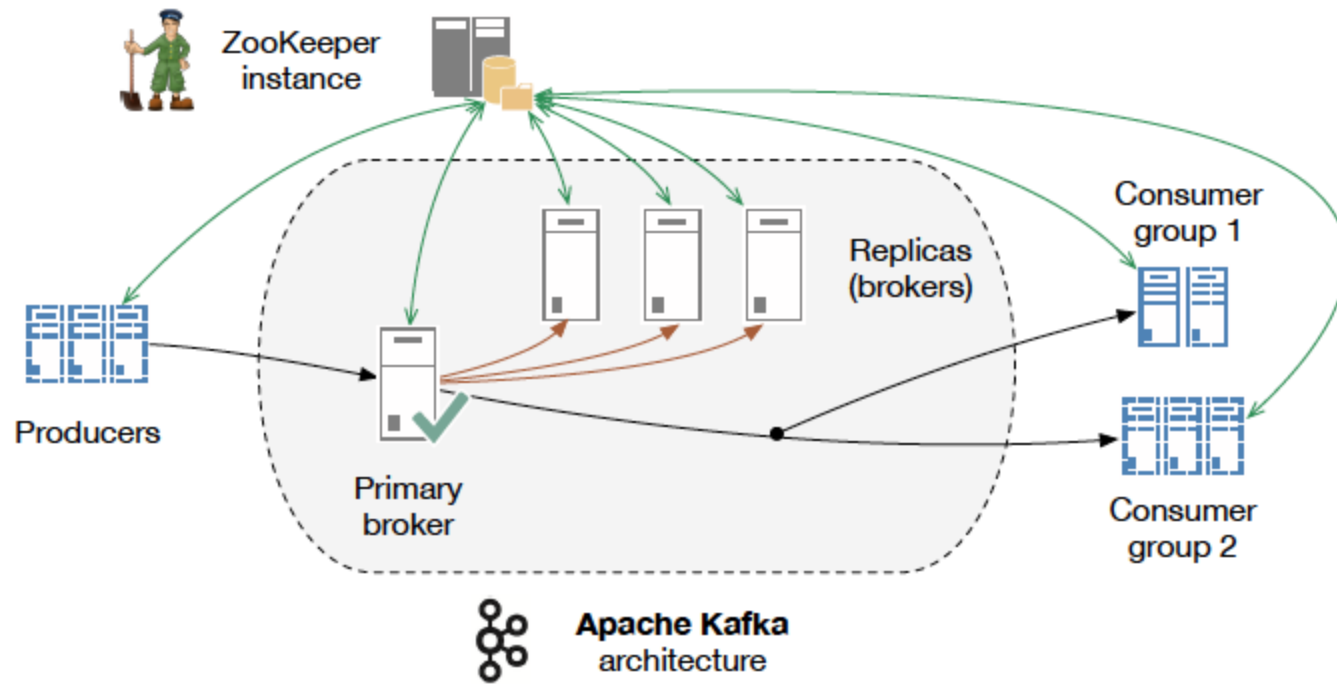
- Kafka is run as a cluster on one or more servers that can span multiple datacenters.
- The Kafka cluster stores streams of *records* in categories called *topics*.
- Each record consists of a key, a value, and a timestamp.

Kafka Topic

Anatomy of a Topic



Zookeeper Leading Kafka



Topic Creation

```
> bin\windows\kafka-topics.bat --create  
    --zookeeper localhost:2181  
    --replication-factor 1  
    --partitions 1  
    --topic my-topic
```

List Topics

```
> bin\windows\kafka-topics.bat --list  
    --zookeeper localhost:2181
```

Kafka Console Consumer

```
> bin\windows\kafka-console-consumer.bat  
    --bootstrap-server localhost:9092  
    --topic my-topic - -from-beginning
```