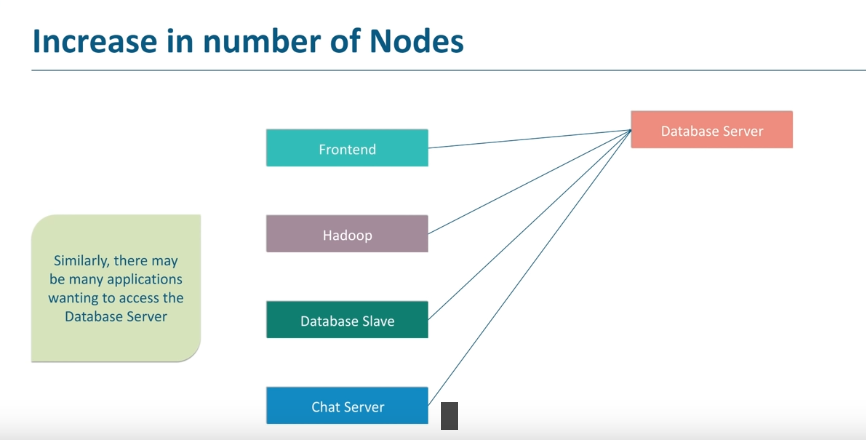
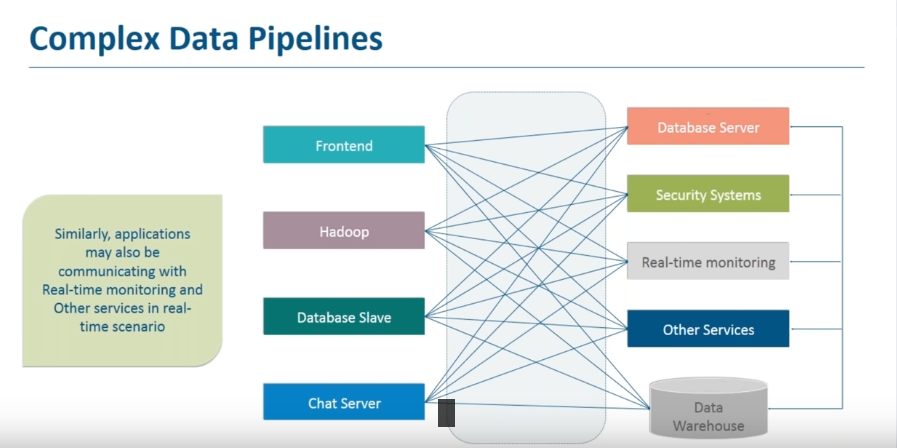
**KAFKA**

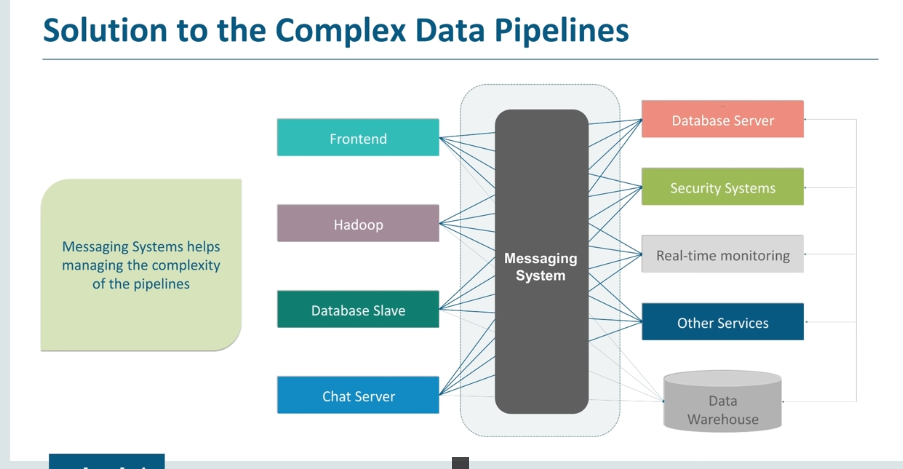
Apache Kafka is an open-source distributed event streaming platform used by thousands of companies for high-performance data pipelines, streaming analytics, data integration, and mission-critical applications.

**Data Pipelines** - Communication is required between two systems in real time scenarios (like chat server and database) which is done using data pipeline.

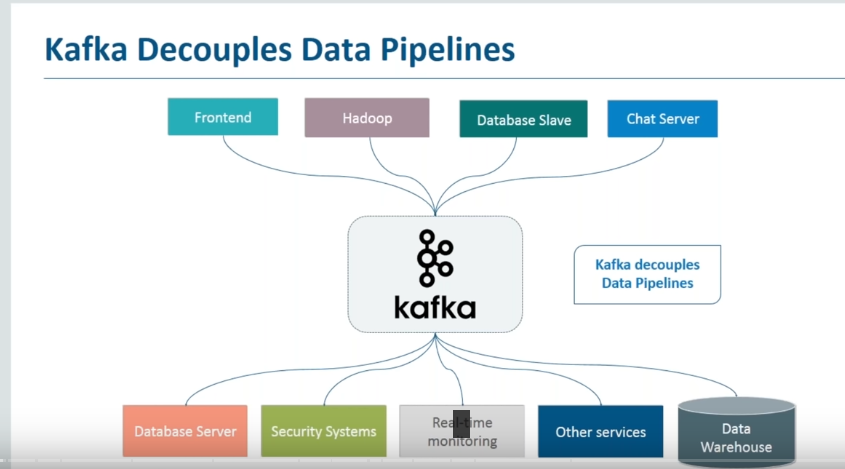




**Need of messaging systems:** Messaging systems reduce complexity of data pipelines and make communication between system simple and manageable.



**Kafka**

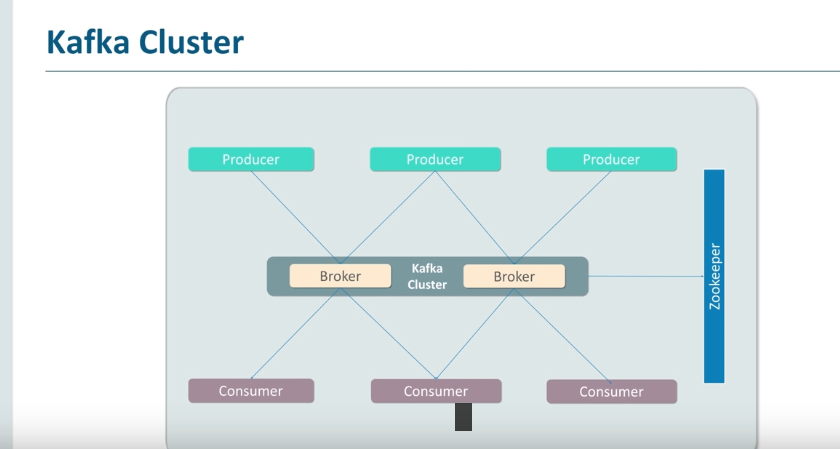
****

**Apache Kafka is a distributed public-subscribed messaging system.**

**Kafka Terminologies**

1. **Producer -** A producer can be any application who can publish messages to a topic.
2. **Consumer -** A consumer can be any application that subscribe to a topic and consume the messages.
3. **Broker -** Kafka cluster is a set of servers, each of which is called a broker.
4. **Topic -** A topic is a category or feed name to which records are published.
5. **Partition -** Topics are broken up into ordered commit logs called partitions.
6. **Zookeeper -** Zookeeper is used for managing and coordinating broker. Like maintaining the meta data.

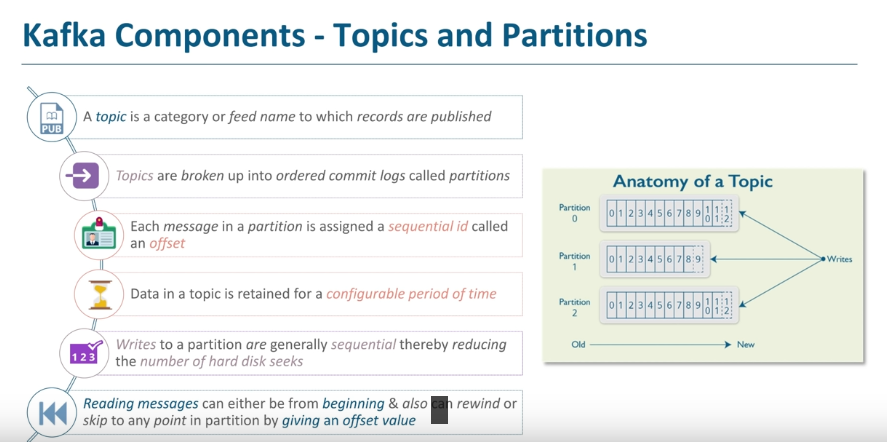
**Kafka Cluster**

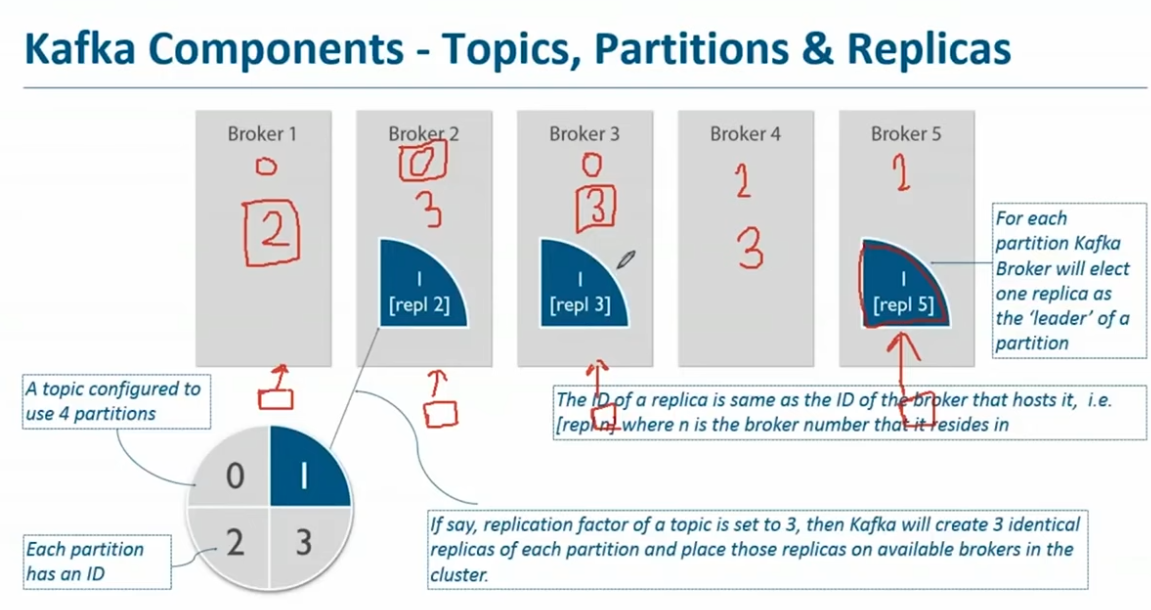


**Kafka Features**

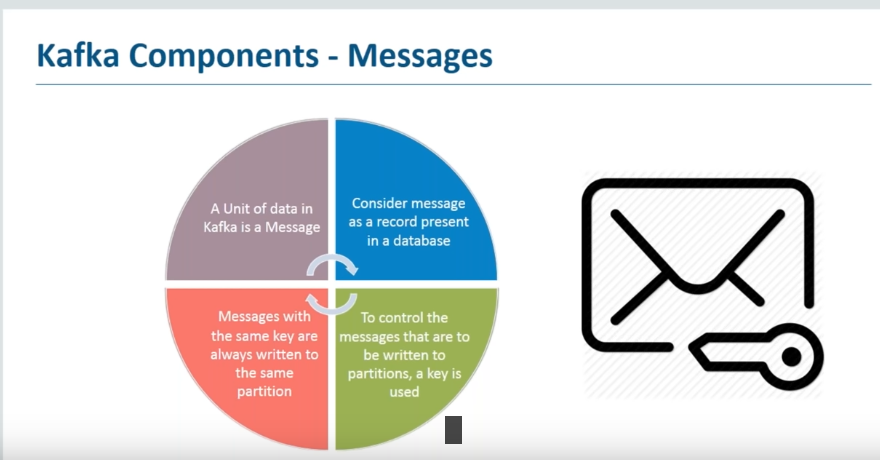
1. High Throughput,
2. Scalability,
3. Data Loss,
4. Streaming Processing,
5. Durability,
6. Replication.

Kafka Components

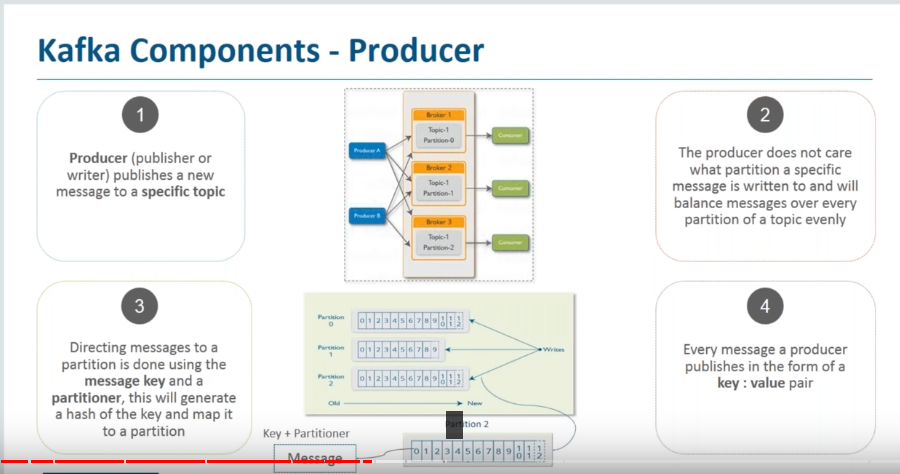




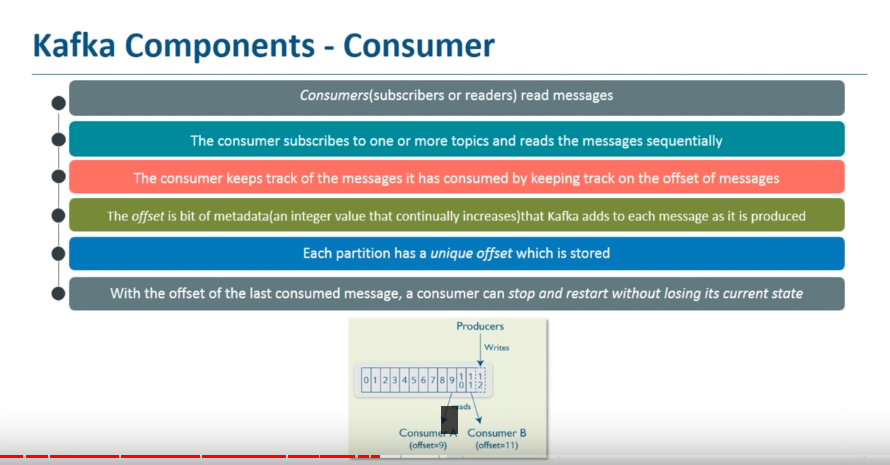
**Kafka components - messages**

****

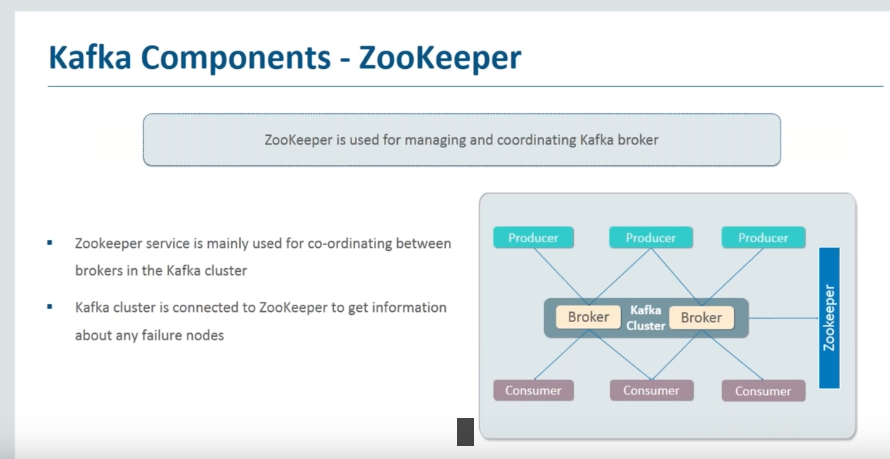
**Kafka Component - Producer**

****

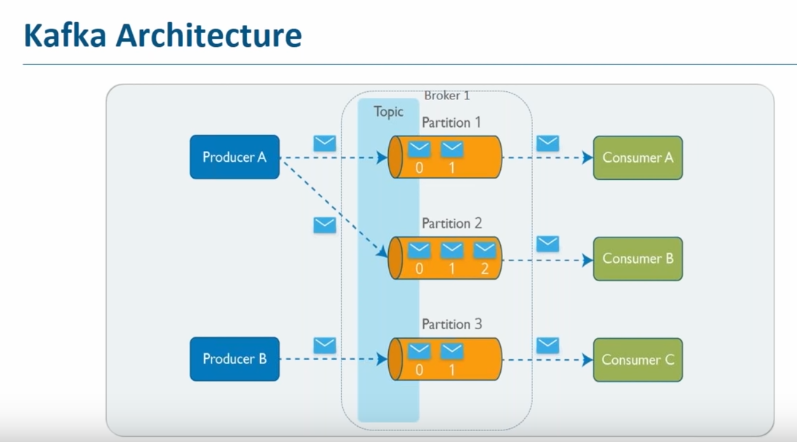
**Kafka Component - Consumer**

****

**Kafka Component - Zookeeper**

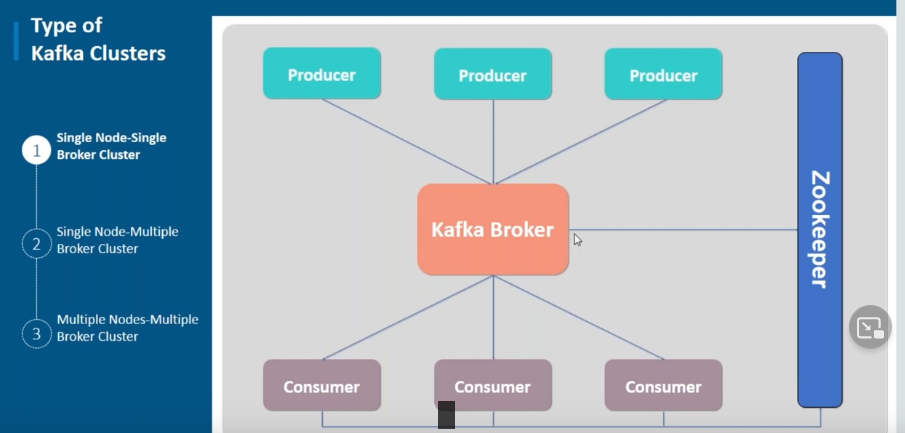


**Kafka Architecture**



**Types of Kafka cluster**

Single node



Multi node - This is used in real world.

