Apache Kafka

It has high throughput but low storage.

DB has low throughput but high storage.

Suppose data is getting inserted or fetched every millisecond in DB then it's sure that your DB will be down/crashed.

So, to overcome this problem, we do have Kafka.

Kafka is an open source platform for communication between Producer and Consumer.

It contains Topic , Partition. (KTP-> Kafka, Topic, Partition)

Auto-balancing -> Zookeeper

Let's take an example of Zomato Live Location Sharing of Delivery Driver/ Partner in Customer UI.
  – Here, inserting the data of live location of Delivery Partner in DB every millisecond surely crash your DB.
  – Because DB has low throughput I.e ops(operation per second).
  – Now Apache Kafka comes into picture, it makes the data sharing as Producer and data fetching as Consumer.
  – Here, Delivery Driver/Partner is a Producer and Customer is a consumer.
  – Kafka will provide you Topics which contains Partitions.
  – Suppose we want to divide the data of Live Location of Delivery driver/ partner into North India & South India then we can create two partition in a Topic which will be consumed by a group of Consumers or a consumer or two consumer using Zookeeper (auto-balancing).
  – Here again we have an issue like ok now consumer is able to fetch the data every millisecond but we want to keep records of data in DB. How?
  – Its simple we can do run bulk insert query to insert the data in DB.