Every enterprise application creates data, whether it’s log messages, metrics, user activity, outgoing messages, or something else. How to move all this data becomes nearly as important as the data itself.

Kafka is being used by tens of thousands of organizations, including over a third of the Fortune 500 companies. It’s among the fastest growing open source projects and has spawned an immense ecosystem around it. It’s at the heart of a movement towards managing and processing streams of data

Kafka got its start as an internal infrastructure system we built at LinkedIn. Our observation was simple: there were lots of databases and other systems built to store data, but what was missing in our architecture was something that would help us to handle the continuous flow of data.

We’ve come to think of Kafka as a streaming platform: a system that lets you publish and subscribe to streams of data, store them, and process them, and that is exactly what Apache Kafka is built to be.

Kafka is often compared to Messaging systems like ActiveMQ, Rabbit MQ, IBM’S MQ series, Big data systems like Hadoop, ETL/data integration tools.

1. **3 important things that differentiate from message systems:** distributed cluster and scale elastically. Replicated, Persistent as long as you like and guarantees delivery. Stream processing capabilities and can compute.
2. ***Big data systems like Hadoop. Realtime version of Hadoop.*** *Hadoop lets you store and periodically process file data at a very large scale. Kafka also lets you store and continuously process streams of data, also at a large scale. Continuous, low latency processing. Hadoop and bigdata targeted analytics apps often in data warehousing space but low latency nature of Kafka makes it applicable for kind of core applications that directly power a business. Reacting to events in a business powers the operation of business, feed back into customer experiences and so on.*
3. ***Compared to ETL or data integration tools.*** These tools move data around and of course Kafka does too. Rather than a tool for scraping data out of one system and inserting into another, Kafka is a platform oriented around real-time streams of events. Not only can it connect off-the-shelf applications and data systems, it can also power custom applications built to trigger off the same data streams.