

KAFKA

INTRO



<https://www.linkedin.com/in/vijay-mahawar>

KAFKA

INTRO

What is Kafka?

**Kafka open-source project released
in Github 2010**



<https://www.linkedin.com/in/vijay-mahawar>

KAFKA

INTRO

What is Kafka?

**Kafka was founded by
Jay Kreps
Neha Narkhede
Jun Rao**

**Engineers from Confluent and
Linkedin.**



KAFKA

INTRO

What is Kafka?

‘Kafka’ name originated from Jay Kreps’ interest in literature texts from college by Franz Kafka



KAFKA

INTRO

Why Kafka?

It was built to be scalable messaging system that could meet the needs of both the monitoring and tracking systems.



KAFKA

INTRO

Why Kafka?

It was built to manage real-time streams with large volumes for event-driven architecture.



KAFKA

INTRO

Why Kafka?

It is described as ‘Distributed commit log’ or more recently as a “distributed streaming platform”.



KAFKA

INTRO

KAFKA Terms

Message

Message is Unit of data

Similar to Row/Record in Database



KAFKA

INTRO

KAFKA Terms

Schemas

Structure for Message content

Similar to JSON, XMLTYPE or XSD



KAFKA

INTRO

KAFKA Terms

Topics

Message are categorized
Tables or folder in filesystems



KAFKA

INTRO

KAFKA Terms

Partitions

**Topics further broken down into
partitions**

**Table Partitions/Table Sub-
Partitions**



KAFKA

INTRO

KAFKA Terms

Stream

Data from single topic which is moving.

Similar to Oracle Streams in Change Data Capture used for Replication



KAFKA

INTRO

KAFKA Terms

Producer

**Publisher/Writer or creator of
message in topic**

**Imagine as source in ETL job or
source in UTL_FILE**



KAFKA

INTRO

KAFKA Terms

Consumer

**Subscriber or reader of message in
topic**

**Imagine as target in ETL job or
target in UTL_FILE**



KAFKA

INTRO

KAFKA Terms

Broker

A single Kafka server that receives, stores and transfers the message from producer to consumer.

Instance (set of memory/processes) in Oracle, that acts as intermediary between datafiles/control files/logfiles and users/clients



KAFKA

INTRO

KAFKA Terms

Cluster

Group of Brokers. One is leader broker and other brokers are followers

Similar to Oracle RAC – Oracle Real Application cluster with Multiple instances



KAFKA

INTRO

KAFKA Terms

Zookeeper

centralized service that maintains naming & config data. Provides synchronization b/w brokers, producers and consumers

Similar to gv\$ gv\$session which contains connection metadata or interconnect in Oracle RAC to communicate/sync between different clusters



KAFKA

INTRO

Kafka Terms	Meaning	Similar term in Database
Message	Unit of data	Row or Record
Schemas	Structure for Message content	JSON, XMLTYPE or XSD
Topics	Message are categorized	Tables or folder in filesystems
Partitions	Topics further broken down into partitions	Table Partitions/Table Sub-Partitions
Stream	Data from single topic which is moving.	Similar to Oracle Streams in Change Data Capture used for Replication
Producer/Writer/Source/Publisher	Publisher or creator of message in topic	Imagine as <i>source</i> in ETL job or <i>source</i> in UTL_FILE
Consumer/Reader/Sink/Subscriber	Subscriber or reader of message in topic	Imagine as <i>target</i> in ETL job or <i>target</i> in UTL_FILE
Broker	A single Kafka server that receives, stores and transfers the message from producer to consumer.	<i>Instance</i> (set of memory/processes) in Oracle, that acts as intermediary between datafiles/control files/logfiles and users/clients. One instance in Oracle RAC – Oracle real application cluster
Cluster	Group of Brokers. One is leader broker and other brokers are followers	Similar to Oracle RAC – Oracle Real Application cluster with Multiple instances
Zookeeper	Stores configuration information about the cluster and consumer client's details	Similar to global dictionary views (gv\$) like gv\$session which contains connection metadata or interconnect in Oracle RAC to communicate/sync between different clusters



KAFKA

INTRO

Thanks

Have A Nice Day



<https://www.linkedin.com/in/vijay-mahawar>