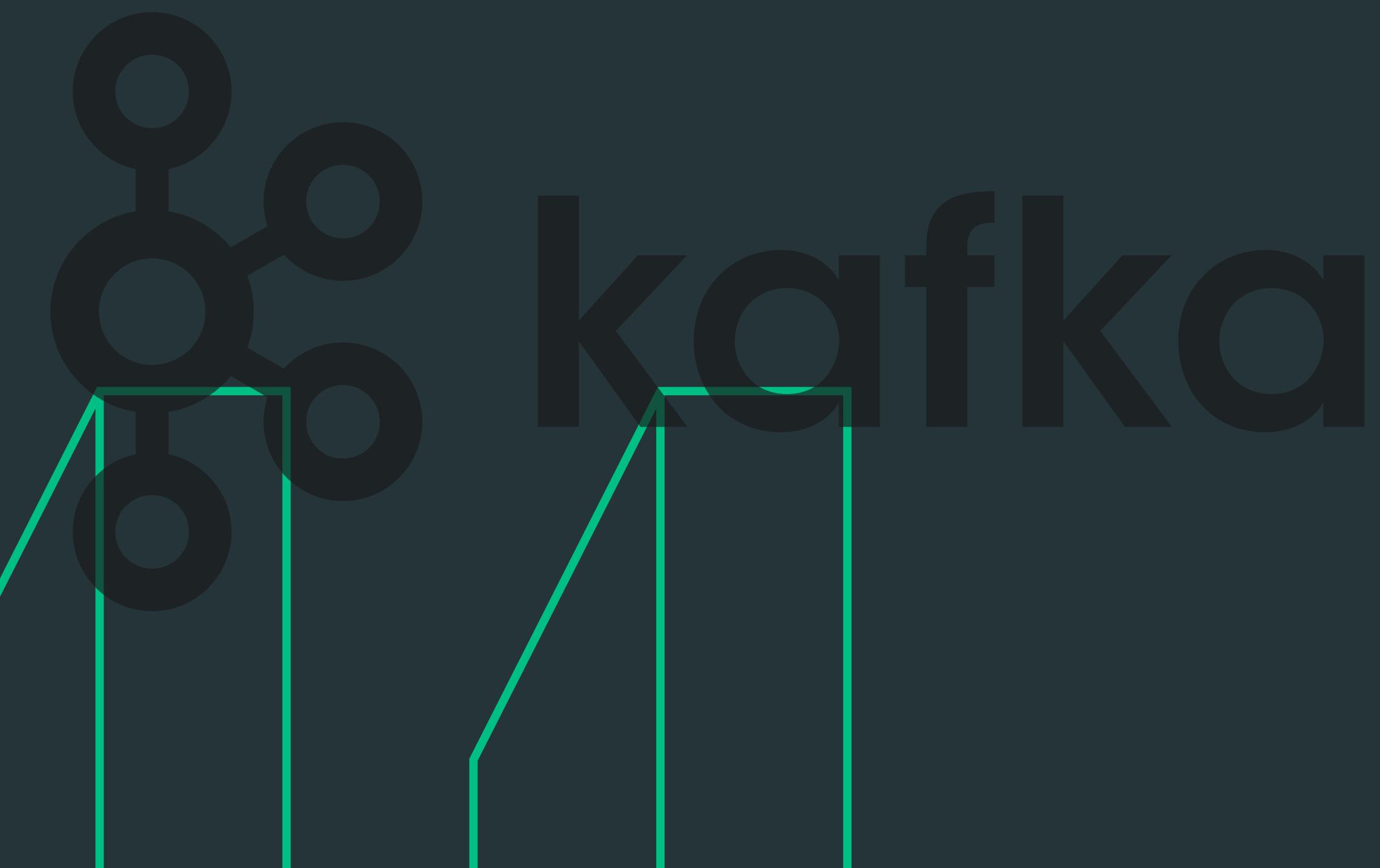


' Data Engineering
101 - Kafka

KAFKA: Concepts



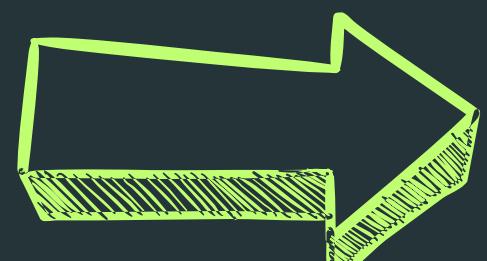
KAFKA BROKER

A KAFKA BROKER IS A SERVER THAT RUNS THE KAFKA SOFTWARE AND IS RESPONSIBLE FOR STORING AND SERVING DATA. BROKERS RECEIVE MESSAGES FROM PRODUCERS, ASSIGN OFFSETS TO MESSAGES, AND STORE THEM ON DISK.

EXAMPLE: IN A KAFKA CLUSTER, MULTIPLE BROKERS WORK TOGETHER TO ENSURE DATA IS RELIABLY STORED AND SERVED.



Shwetank Singh
GritSetGrow - GSGLearn.com



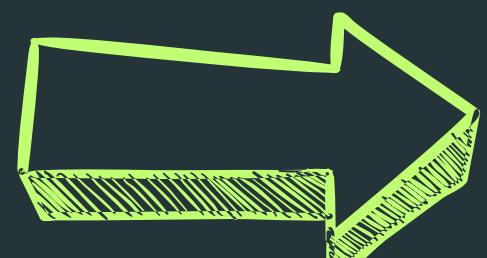
TOPICS

TOPICS ARE LOGICAL CHANNELS TO WHICH MESSAGES ARE SENT BY PRODUCERS AND FROM WHICH MESSAGES ARE READ BY CONSUMERS. A TOPIC IS DIVIDED INTO MULTIPLE PARTITIONS TO ALLOW PARALLEL PROCESSING.

EXAMPLE: A "USER_ACTIVITY" TOPIC MIGHT BE DIVIDED INTO SEVERAL PARTITIONS TO HANDLE HIGH MESSAGE VOLUME.



Shwetank Singh
GritSetGrow - GSGLearn.com



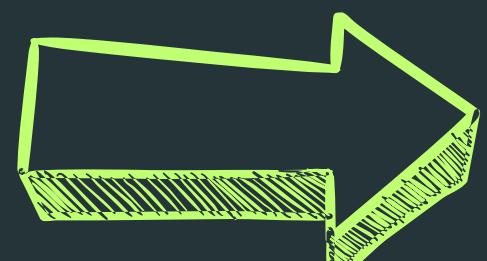
PARTITIONS

PARTITIONS ARE SUBDIVISIONS OF TOPICS. EACH PARTITION IS AN ORDERED, IMMUTABLE SEQUENCE OF MESSAGES THAT IS CONTINUALLY APPENDED TO. PARTITIONS ENABLE KAFKA TO SCALE HORIZONTALLY AND MAINTAIN MESSAGE ORDER.

EXAMPLE: PARTITION 0 OF THE "USER_ACTIVITY" TOPIC STORES MESSAGES FOR A SPECIFIC SUBSET OF USERS.



Shwetank Singh
GritSetGrow - GSGLearn.com



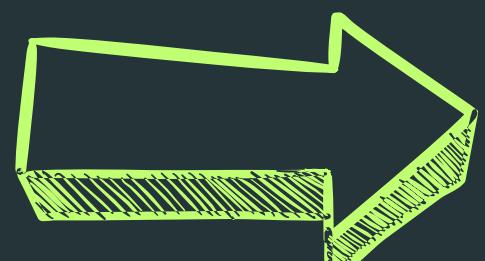
PRODUCERS

PRODUCERS ARE CLIENTS THAT SEND MESSAGES TO KAFKA TOPICS. THEY CAN SEND MESSAGES TO SPECIFIC PARTITIONS BASED ON A PARTITIONING STRATEGY OR DISTRIBUTE THEM EVENLY ACROSS ALL PARTITIONS.

EXAMPLE: A WEB APPLICATION THAT LOGS USER ACTIVITY SENDS THESE LOGS TO A KAFKA TOPIC AS MESSAGES.



Shwetank Singh
GritSetGrow - GSGLearn.com



CONSUMERS

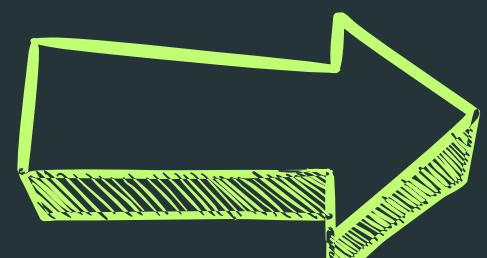
CONSUMERS ARE CLIENTS THAT READ MESSAGES FROM KAFKA TOPICS.

CONSUMERS CAN OPERATE INDIVIDUALLY OR AS PART OF A CONSUMER GROUP, WHICH ALLOWS FOR PARALLEL PROCESSING OF MESSAGES.

EXAMPLE: AN ANALYTICS SERVICE READS USER ACTIVITY LOGS FROM A KAFKA TOPIC TO GENERATE REPORTS.



Shwetank Singh
GritSetGrow - GSGLearn.com



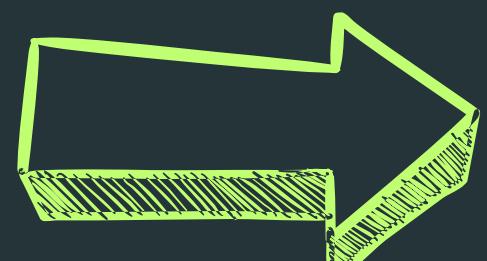
Kafka CONSUMER GROUPS

CONSUMER GROUPS ALLOW MULTIPLE CONSUMERS TO COLLABORATE ON PROCESSING MESSAGES FROM A TOPIC. EACH PARTITION IN A TOPIC IS ASSIGNED TO ONLY ONE CONSUMER WITHIN A GROUP AT A TIME, ENSURING PARALLEL PROCESSING AND LOAD BALANCING.

EXAMPLE: THREE CONSUMERS IN A GROUP PROCESS MESSAGES FROM SIX PARTITIONS.



Shwetank Singh
GritSetGrow - GSGLearn.com



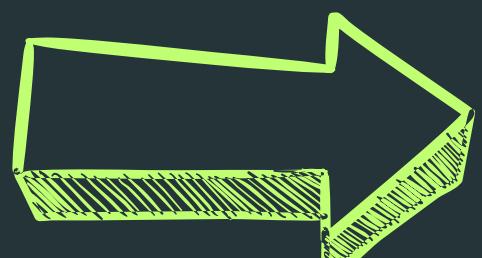
OFFSETS

OFFSETS ARE UNIQUE IDENTIFIERS ASSIGNED TO EACH MESSAGE WITHIN A PARTITION. CONSUMERS USE OFFSETS TO TRACK WHICH MESSAGES HAVE BEEN READ.

EXAMPLE: A CONSUMER READS MESSAGES UP TO OFFSET 105 AND RESUMES FROM OFFSET 106 AFTER A RESTART.



Shwetank Singh
GritSetGrow - GSGLearn.com



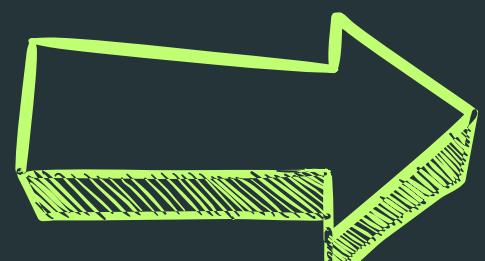
KAFKA CLUSTER

A KAFKA CLUSTER IS COMPOSED OF MULTIPLE BROKERS THAT WORK TOGETHER. CLUSTERS PROVIDE FAULT TOLERANCE AND HIGH AVAILABILITY.

EXAMPLE: A CLUSTER WITH THREE BROKERS CAN CONTINUE OPERATING IF ONE BROKER FAILS.



Shwetank Singh
GritSetGrow - GSGLearn.com



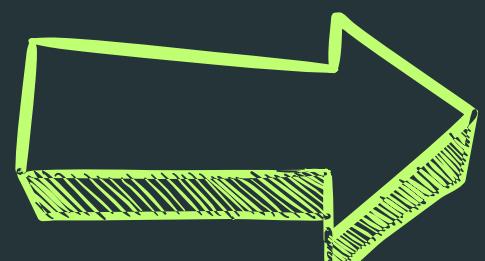
REPLICATION

KAFKA REPLICATES PARTITIONS ACROSS MULTIPLE BROKERS TO ENSURE FAULT TOLERANCE. EACH PARTITION HAS A LEADER AND SEVERAL FOLLOWERS. THE LEADER HANDLES ALL READS AND WRITES, WHILE FOLLOWERS REPLICATE THE DATA.

EXAMPLE: PARTITION 0 HAS ONE LEADER AND TWO FOLLOWERS ACROSS THREE BROKERS.



Shwetank Singh
GritSetGrow - GSGLearn.com



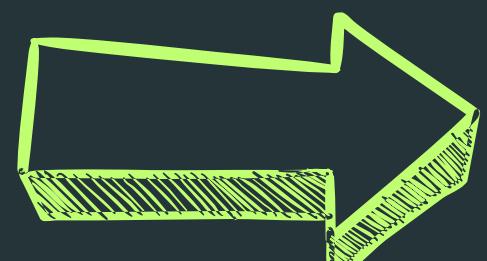
ZOOKEEPER

ZOOKEEPER IS USED FOR DISTRIBUTED COORDINATION AND METADATA MANAGEMENT IN KAFKA. IT MANAGES BROKER METADATA, LEADER ELECTION, AND CONFIGURATION.

EXAMPLE: ZOOKEEPER ENSURES A NEW LEADER IS ELECTED IF THE CURRENT LEADER BROKER FAILS.



Shwetank Singh
GritSetGrow - GSGLearn.com



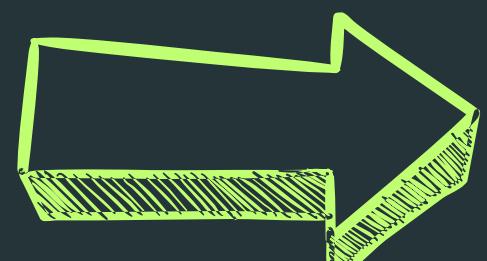
PRODUCERS AND ACKS

PRODUCERS SEND MESSAGES TO BROKERS AND CAN CONFIGURE ACKNOWLEDGMENT SETTINGS (ACKS) TO ENSURE RELIABLE MESSAGE DELIVERY.

EXAMPLE: A PRODUCER CONFIGURES ACKS TO WAIT FOR CONFIRMATION FROM ALL REPLICAS BEFORE CONSIDERING A MESSAGE SENT.



Shwetank Singh
GritSetGrow - GSGLearn.com



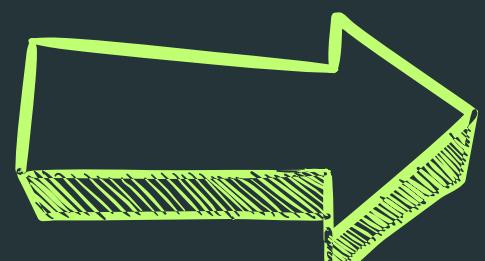
RETENTION POLICY

KAFKA TOPICS CAN HAVE RETENTION POLICIES THAT DETERMINE HOW LONG MESSAGES ARE STORED. POLICIES CAN BE TIME-BASED OR SIZE-BASED.

EXAMPLE: A TOPIC IS CONFIGURED TO RETAIN MESSAGES FOR 7 DAYS, AFTER WHICH THEY ARE DELETED.



Shwetank Singh
GritSetGrow - GSGLearn.com

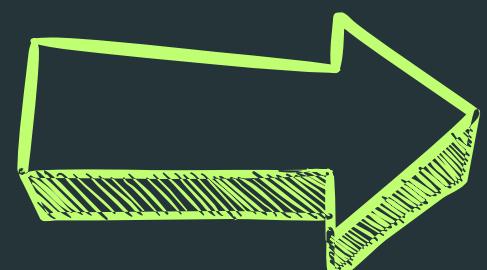


LOG COMPACTION

LOG COMPACTION ENSURES THAT ONLY THE LATEST MESSAGE FOR EACH KEY IS RETAINED IN A TOPIC, USEFUL FOR MAINTAINING THE LATEST STATE.
EXAMPLE: A LOG-COMPACTED TOPIC RETAINS ONLY THE LATEST UPDATE FOR EACH USER PROFILE.



Shwetank Singh
GritSetGrow - GSGLearn.com



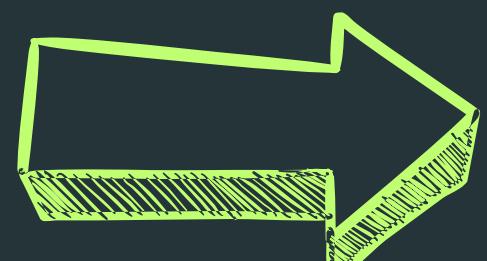
KAFKA CONNECT

KAFKA CONNECT IS A FRAMEWORK FOR INTEGRATING KAFKA WITH OTHER DATA SYSTEMS. IT PROVIDES CONNECTORS TO MOVE DATA IN AND OUT OF KAFKA.

EXAMPLE: USING KAFKA CONNECT TO SYNC DATA BETWEEN A MYSQL DATABASE AND A KAFKA TOPIC.



Shwetank Singh
GritSetGrow - GSGLearn.com



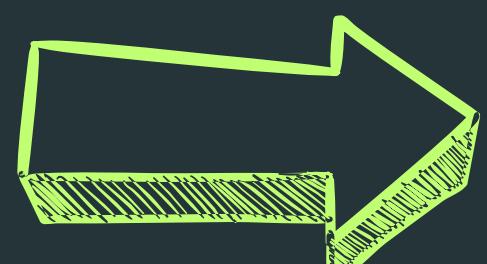
KAFKA STREAMS

KAFKA STREAMS IS A LIBRARY FOR BUILDING STREAM PROCESSING APPLICATIONS ON TOP OF KAFKA. IT ALLOWS PROCESSING AND TRANSFORMING DATA IN REAL TIME.

EXAMPLE: AN APPLICATION USING KAFKA STREAMS AGGREGATES CLICKSTREAM DATA TO GENERATE REAL-TIME METRICS.



Shwetank Singh
GritSetGrow - GSGLearn.com



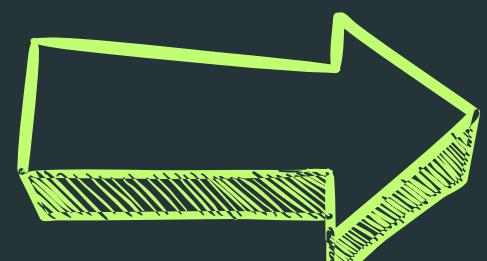
MIRRORMAKER

MIRRORMAKER IS A TOOL FOR REPLICATING DATA BETWEEN KAFKA CLUSTERS, OFTEN USED FOR CROSS-DATACENTER REPPLICATION.

EXAMPLE: USING MIRRORMAKER TO REPLICATE MESSAGES FROM A PRIMARY DATACENTER TO A BACKUP DATACENTER.



Shwetank Singh
GritSetGrow - GSGLearn.com



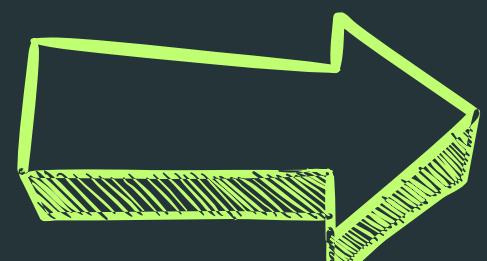
KAFKA API

KAFKA PROVIDES APIS FOR PRODUCING, CONSUMING, AND MANAGING DATA, INCLUDING PRODUCER API, CONSUMER API, AND ADMIN API.

EXAMPLE: USING THE PRODUCER API TO SEND MESSAGES FROM A JAVA APPLICATION TO A KAFKA TOPIC.



Shwetank Singh
GritSetGrow - GSGLearn.com



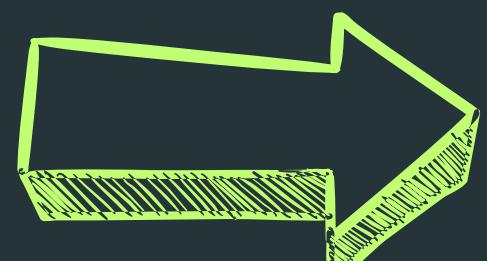
SECURITY

KAFKA SUPPORTS VARIOUS SECURITY FEATURES, INCLUDING SSL ENCRYPTION, SASL AUTHENTICATION, AND ACLS FOR AUTHORIZATION.

EXAMPLE: CONFIGURING SSL TO ENCRYPT DATA IN TRANSIT AND SASL FOR CLIENT AUTHENTICATION.



Shwetank Singh
GritSetGrow - GSGLearn.com



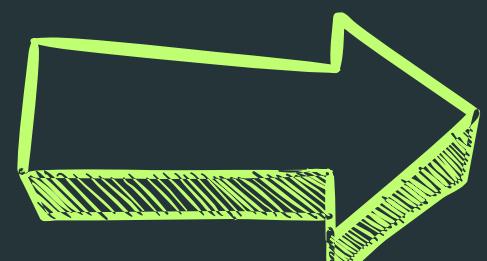
ADMINCLIENT API

THE ADMINCLIENT API ALLOWS PROGRAMMATIC MANAGEMENT OF KAFKA TOPICS, BROKERS, AND CONFIGURATIONS.

EXAMPLE: USING ADMINCLIENT TO CREATE A NEW TOPIC AND CONFIGURE ITS RETENTION POLICY.



Shwetank Singh
GritSetGrow - GSGLearn.com



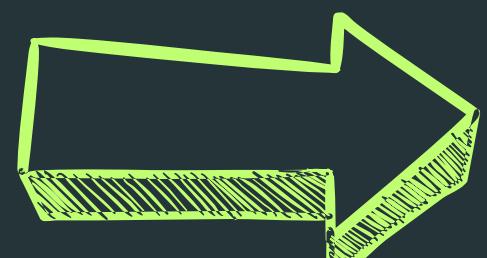
MONITORING AND METRICS

KAFKA PROVIDES METRICS FOR MONITORING CLUSTER HEALTH AND PERFORMANCE. TOOLS LIKE PROMETHEUS AND GRAFANA CAN BE USED TO VISUALIZE THESE METRICS.

EXAMPLE: MONITORING CONSUMER LAG AND BROKER HEALTH USING PROMETHEUS AND GRAFANA DASHBOARDS.



Shwetank Singh
GritSetGrow - GSGLearn.com



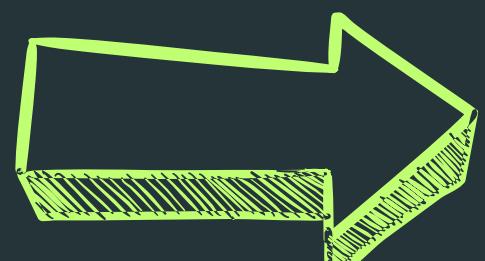
MESSAGE DELIVERY SEMANTICS

KAFKA SUPPORTS THREE TYPES OF MESSAGE DELIVERY SEMANTICS: AT MOST ONCE, AT LEAST ONCE, AND EXACTLY ONCE.

EXAMPLE: CONFIGURING A PRODUCER FOR EXACTLY-ONCE DELIVERY TO ENSURE NO MESSAGE IS LOST OR DUPLICATED.



Shwetank Singh
GritSetGrow - GSGLearn.com



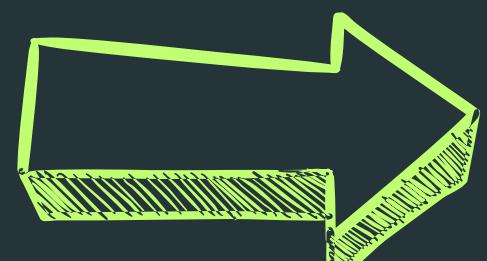
STATEFUL PROCESSING

KAFKA STREAMS SUPPORTS STATEFUL PROCESSING, ALLOWING APPLICATIONS TO MAINTAIN STATE ACROSS MESSAGES USING STATE STORES.

EXAMPLE: A STREAM PROCESSING APPLICATION THAT MAINTAINS A RUNNING COUNT OF EVENTS OVER A WINDOW OF TIME.



Shwetank Singh
GritSetGrow - GSGLearn.com



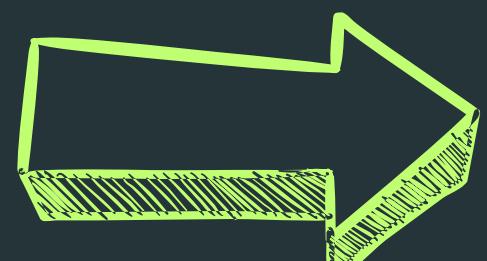
WINDOWED OPERATIONS

KAFKA STREAMS PROVIDES SUPPORT FOR WINDOWED OPERATIONS, ENABLING TIME-BASED AGGREGATIONS AND TRANSFORMATIONS.

EXAMPLE: CALCULATING THE AVERAGE NUMBER OF USER CLICKS PER MINUTE USING WINDOWED OPERATIONS.



Shwetank Singh
GritSetGrow - GSGLearn.com



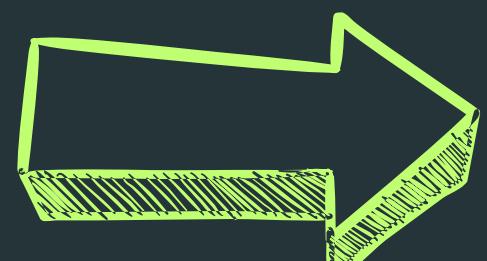
KSQL

KSQL IS A SQL-LIKE INTERFACE FOR STREAM PROCESSING IN KAFKA, SIMPLIFYING THE CREATION OF STREAM PROCESSING APPLICATIONS.

EXAMPLE: USING KSQL TO FILTER, AGGREGATE, AND TRANSFORM STREAMS OF DATA IN REAL TIME.



Shwetank Singh
GritSetGrow - GSGLearn.com



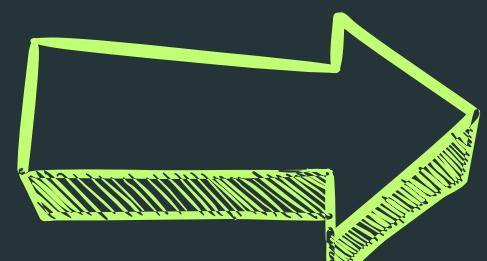
KAFKA ECOSYSTEM

KAFKA'S ECOSYSTEM INCLUDES VARIOUS TOOLS AND FRAMEWORKS FOR COMPREHENSIVE DATA PROCESSING, SUCH AS KAFKA CONNECT, KAFKA STREAMS, AND KSQL.

EXAMPLE: INTEGRATING KAFKA WITH A RELATIONAL DATABASE USING KAFKA CONNECT AND PROCESSING THE DATA WITH KAFKA STREAMS.



Shwetank Singh
GritSetGrow - GSGLearn.com



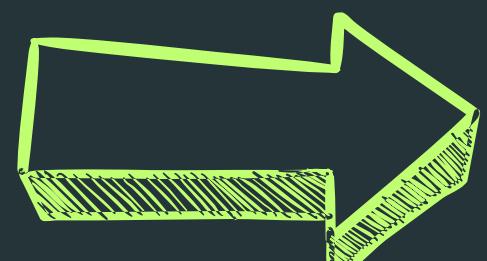
PUBLISH/SUBSCRIBE MESSAGING

PUB/SUB SYSTEMS ALLOW DECOUPLING OF MESSAGE PRODUCERS AND CONSUMERS. KAFKA ACTS AS A BROKER FACILITATING THIS.

EXAMPLE: AN APPLICATION PUBLISHES USER ACTIVITY LOGS WHICH CAN BE CONSUMED BY ANALYTICS SERVICES.



Shwetank Singh
GritSetGrow - GSGLearn.com



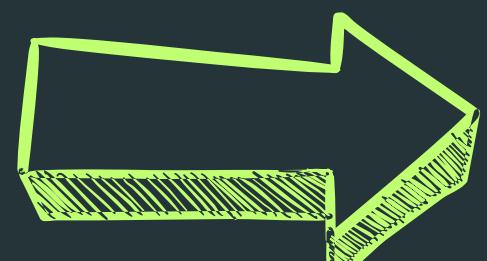
MESSAGE AND BATCHES

MESSAGES ARE THE BASIC UNIT OF DATA IN KAFKA, STORED AS BYTE ARRAYS. MESSAGES ARE WRITTEN IN BATCHES FOR EFFICIENCY.

EXAMPLE: A BATCH OF LOG MESSAGES SENT FROM AN APPLICATION.



Shwetank Singh
GritSetGrow - GSGLearn.com



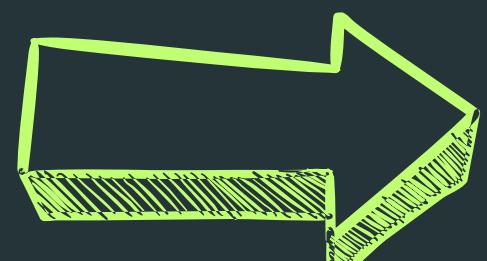
SCHEMAS

SCHEMAS DEFINE THE STRUCTURE OF MESSAGES, ENSURING CONSISTENCY. APACHE AVRO IS A COMMON SERIALIZATION FRAMEWORK USED WITH KAFKA.

EXAMPLE: AVRO SCHEMA FOR USER PROFILE DATA.



Shwetank Singh
GritSetGrow - GSGLearn.com



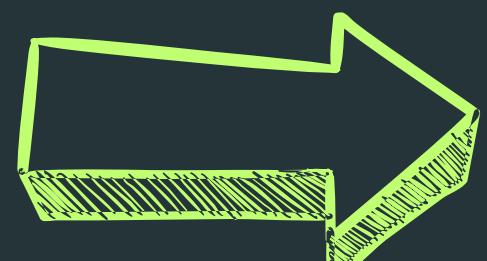
TOPICS AND PARTITIONS

TOPICS ARE CATEGORIES TO WHICH MESSAGES ARE PUBLISHED. TOPICS ARE DIVIDED INTO PARTITIONS FOR SCALABILITY AND REDUNDANCY.

EXAMPLE: A "USER_ACTIVITY" TOPIC WITH PARTITIONS.



Shwetank Singh
GritSetGrow - GSGLearn.com



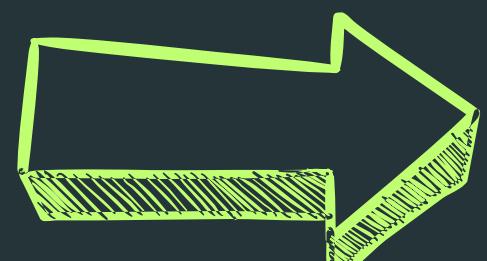
PRODUCERS AND CONSUMERS

PRODUCERS CREATE AND SEND MESSAGES TO KAFKA TOPICS.
CONSUMERS READ MESSAGES FROM TOPICS.

EXAMPLE: A MICROSERVICE PRODUCING ORDER DATA AND ANOTHER CONSUMING FOR PROCESSING.



Shwetank Singh
GritSetGrow - GSGLearn.com



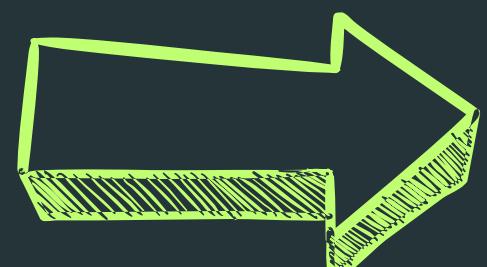
BROKERS AND CLUSTERS

A BROKER IS A KAFKA SERVER THAT STORES DATA AND SERVES CLIENTS. MULTIPLE BROKERS FORM A KAFKA CLUSTER, PROVIDING FAULT TOLERANCE AND SCALABILITY.

EXAMPLE: A KAFKA CLUSTER WITH THREE BROKERS.



Shwetank Singh
GritSetGrow - GSGLearn.com



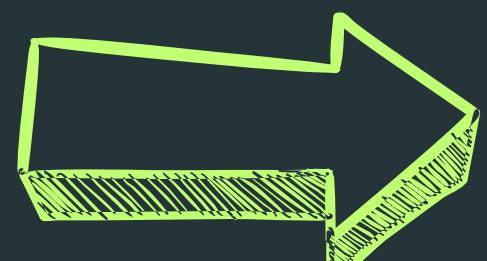
DISK-BASED RETENTION

KAFKA RETAINS MESSAGES ON DISK FOR A CONFIGURED PERIOD, ALLOWING CONSUMERS TO READ AT THEIR PACE.

EXAMPLE: RETAINING LOGS FOR 7 DAYS.



Shwetank Singh
GritSetGrow - GSGLearn.com



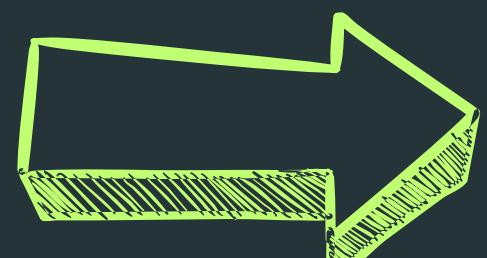
MULTIPLE PRODUCERS AND CONSUMERS

KAFKA SUPPORTS MULTIPLE PRODUCERS AND CONSUMERS FOR THE SAME TOPIC, ENABLING FLEXIBLE DATA PIPELINES.

EXAMPLE: MULTIPLE SENSORS PRODUCING DATA TO A SINGLE TOPIC, MULTIPLE ANALYTICS SERVICES CONSUMING IT.



Shwetank Singh
GritSetGrow - GSGLearn.com



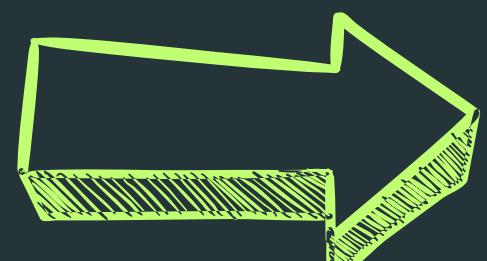
HIGH THROUGHPUT

KAFKA CAN HANDLE LARGE VOLUMES OF MESSAGES EFFICIENTLY DUE TO ITS ARCHITECTURE.

EXAMPLE: PROCESSING MILLIONS OF LOG ENTRIES PER SECOND.



Shwetank Singh
GritSetGrow - GSGLearn.com



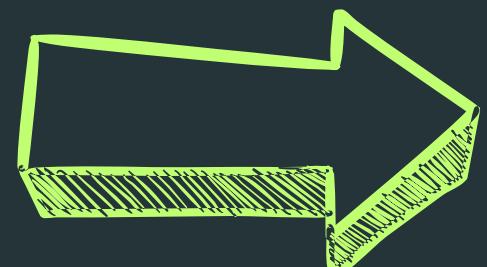
STREAM PROCESSING

KAFKA SUPPORTS REAL-TIME
PROCESSING OF STREAMS OF DATA
USING TOOLS LIKE KAFKA STREAMS.

EXAMPLE: REAL-TIME ANALYTICS ON
INCOMING TRANSACTION DATA.



Shwetank Singh
GritSetGrow - GSGLearn.com



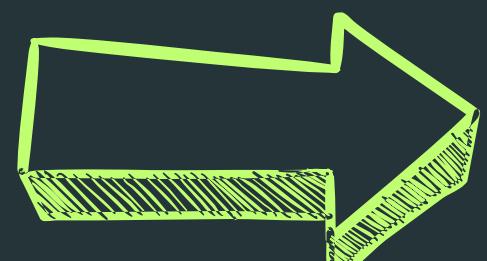
KAFKA CONNECT

**KAFKA CONNECT SIMPLIFIES THE
INTEGRATION OF KAFKA WITH OTHER
DATA SYSTEMS.**

**EXAMPLE: USING KAFKA CONNECT TO
SYNC DATA BETWEEN A DATABASE
AND A KAFKA TOPIC.**



Shwetank Singh
GritSetGrow - GSGLearn.com



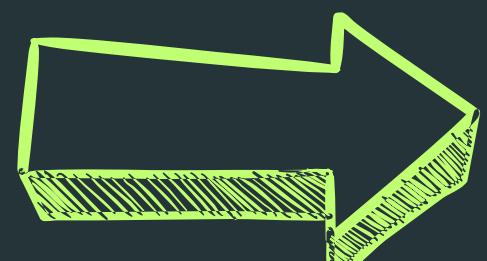
KAFKA STREAMS API

KAFKA STREAMS API ALLOWS BUILDING STREAM PROCESSING APPLICATIONS WITH KAFKA.

EXAMPLE: AN APPLICATION THAT AGGREGATES USER CLICKSTREAM DATA IN REAL-TIME.



Shwetank Singh
GritSetGrow - GSGLearn.com



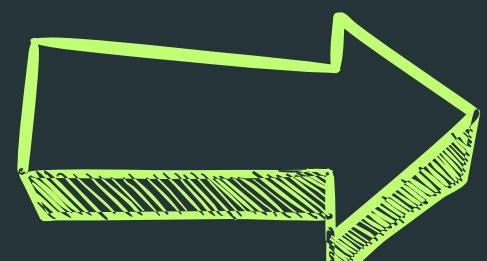
LOG COMPACTION

KAFKA CAN RETAIN ONLY THE LATEST MESSAGE PER KEY IN A LOG-COMPACTED TOPIC, USEFUL FOR CHANGELOG DATA.

EXAMPLE: KEEPING ONLY THE LATEST UPDATE TO USER PROFILES.



Shwetank Singh
GritSetGrow - GSGLearn.com



EXACTLY ONCE SEMANTICS

KAFKA ENSURES THAT MESSAGES ARE PROCESSED EXACTLY ONCE, EVEN IN DISTRIBUTED SYSTEMS.

EXAMPLE: FINANCIAL TRANSACTIONS PROCESSED WITHOUT DUPLICATES.



Shwetank Singh
GritSetGrow - GSGLearn.com



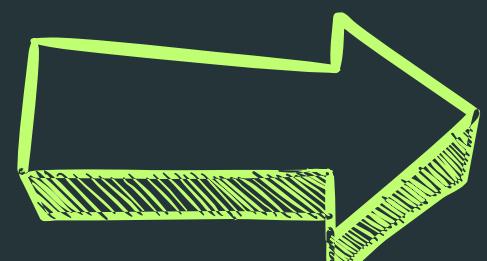
IDEMPOTENT PRODUCER

PRODUCERS CAN SAFELY RETRY
SENDING MESSAGES WITHOUT
DUPLICATING THEM.

EXAMPLE: SENDING A PAYMENT
CONFIRMATION MESSAGE WITH
GUARANTEED SINGLE DELIVERY.



Shwetank Singh
GritSetGrow - GSGLearn.com



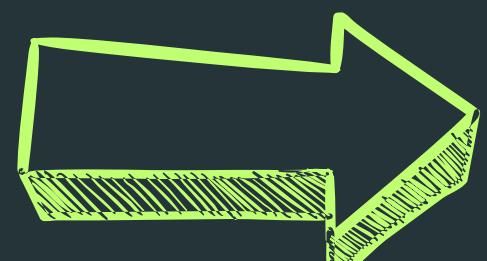
TRANSACTIONS

KAFKA SUPPORTS ATOMIC WRITES
ACROSS MULTIPLE PARTITIONS AND
TOPICS USING TRANSACTIONS.

EXAMPLE: ENSURING THAT A SERIES
OF RELATED MESSAGES ARE EITHER
ALL WRITTEN OR NONE ARE.



Shwetank Singh
GritSetGrow - GSGLearn.com



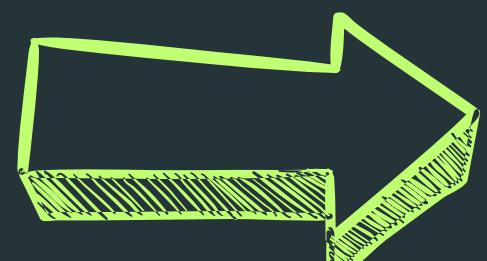
MIRRORMAKER

TOOL FOR REPLICATING KAFKA TOPICS
ACROSS CLUSTERS, USEFUL FOR
DISASTER RECOVERY AND MULTI-
DATACENTER SETUPS.

EXAMPLE: MIRRORING PRODUCTION
DATA TO A BACKUP DATACENTER.



Shwetank Singh
GritSetGrow - GSGLearn.com



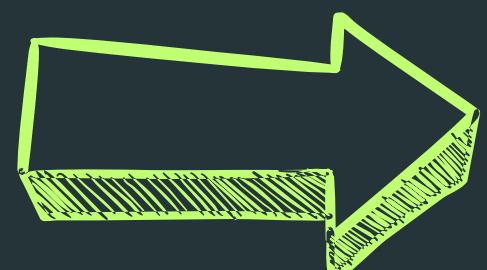
SECURITY

KAFKA SUPPORTS AUTHENTICATION,
AUTHORIZATION, AND ENCRYPTION
TO SECURE DATA.

EXAMPLE: USING SSL FOR
ENCRYPTING DATA IN TRANSIT.



Shwetank Singh
GritSetGrow - GSGLearn.com



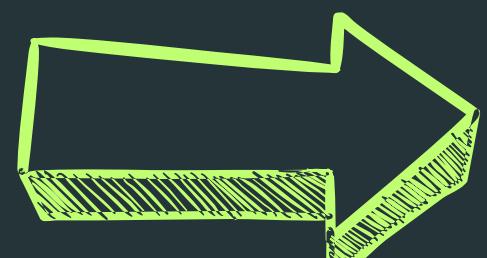
KAFKA ADMINCLIENT

**ADMINCLIENT API ALLOWS
PROGRAMMATIC MANAGEMENT OF
KAFKA.**

**EXAMPLE: CREATING TOPICS,
ALTERING CONFIGURATIONS
PROGRAMMATICALLY.**



Shwetank Singh
GritSetGrow - GSGLearn.com



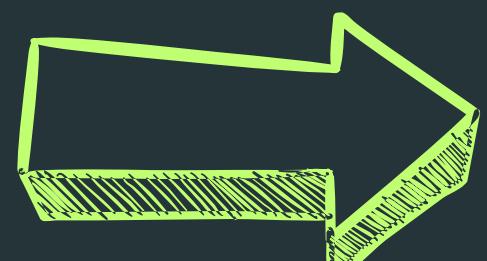
MONITORING AND METRICS

KAFKA PROVIDES METRICS AND MONITORING TOOLS TO TRACK CLUSTER PERFORMANCE.

EXAMPLE: MONITORING CONSUMER LAG AND BROKER HEALTH.



Shwetank Singh
GritSetGrow - GSGLearn.com



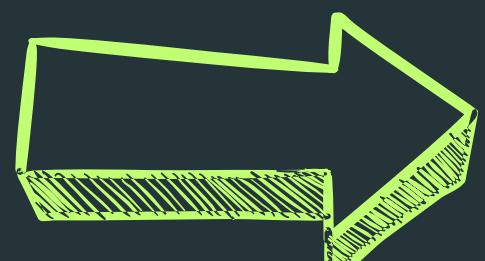
SERIALIZATION AND DESERIALIZATION

KAFKA REQUIRES SERIALIZATION OF DATA FOR TRANSMISSION, WITH SUPPORT FOR VARIOUS FORMATS LIKE AVRO, JSON.

EXAMPLE: SERIALIZING USER DATA TO AVRO FORMAT BEFORE SENDING TO KAFKA.



Shwetank Singh
GritSetGrow - GSGLearn.com



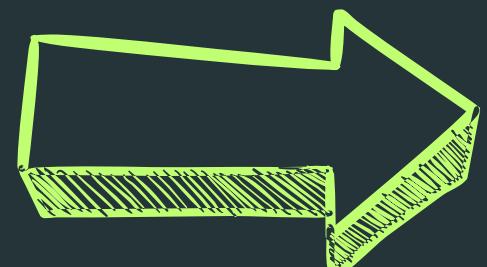
MESSAGE ORDERING

KAFKA MAINTAINS THE ORDER OF MESSAGES WITHIN A PARTITION, IMPORTANT FOR CONSISTENCY.

EXAMPLE: ENSURING ORDER OF TRANSACTION LOGS.



Shwetank Singh
GritSetGrow - GSGLearn.com



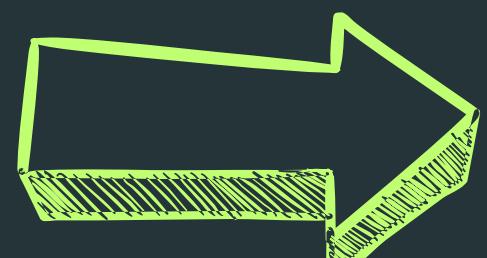
Kafka CONSUMER GROUP

**CONSUMERS CAN JOIN GROUPS TO
BALANCE LOAD AND ENSURE EACH
MESSAGE IS PROCESSED ONCE.**

**EXAMPLE: MULTIPLE CONSUMERS
PROCESSING A HIGH-VOLUME TOPIC
COLLABORATIVELY.**



Shwetank Singh
GritSetGrow - GSGLearn.com



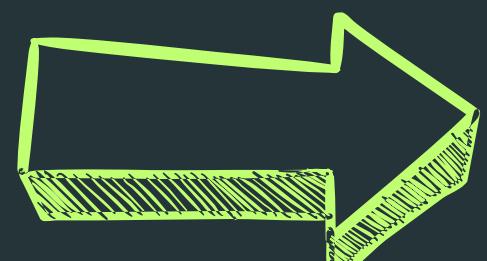
OFFSET MANAGEMENT

KAFKA TRACKS THE OFFSET OF MESSAGES TO MANAGE CONSUMER PROGRESS.

EXAMPLE: STORING OFFSETS IN KAFKA TO RESUME PROCESSING AFTER A RESTART.



Shwetank Singh
GritSetGrow - GSGLearn.com



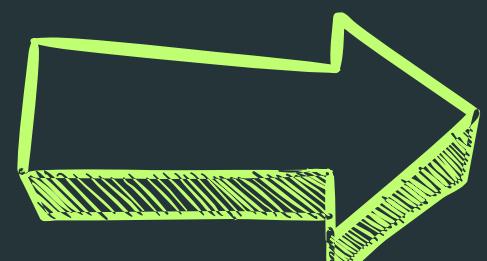
TOPIC REPPLICATION

KAFKA REPLICATES PARTITIONS
ACROSS MULTIPLE BROKERS FOR
FAULT TOLERANCE.

EXAMPLE: A PARTITION REPLICATED
ACROSS THREE BROKERS TO HANDLE
BROKER FAILURE.



Shwetank Singh
GritSetGrow - GSGLearn.com



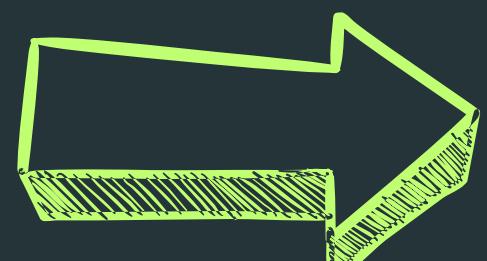
MESSAGE COMPRESSION

KAFKA SUPPORTS COMPRESSING MESSAGES TO SAVE BANDWIDTH AND STORAGE.

EXAMPLE: COMPRESSING LOG MESSAGES BEFORE SENDING TO KAFKA.



Shwetank Singh
GritSetGrow - GSGLearn.com



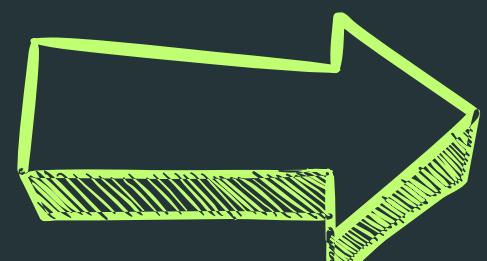
ZOOKEEPER

**KAFKA USES ZOOKEEPER FOR
DISTRIBUTED COORDINATION AND
METADATA MANAGEMENT.**

**EXAMPLE: ZOOKEEPER MANAGING
BROKER METADATA AND LEADER
ELECTION.**



Shwetank Singh
GritSetGrow - GSGLearn.com



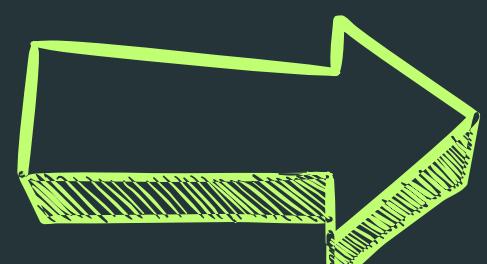
KAFKA BROKER CONFIGURATION

BROKERS CAN BE CONFIGURED FOR
PERFORMANCE, RETENTION POLICIES,
AND MORE.

EXAMPLE: CONFIGURING A BROKER TO
RETAIN MESSAGES FOR 30 DAYS.



Shwetank Singh
GritSetGrow - GSGLearn.com



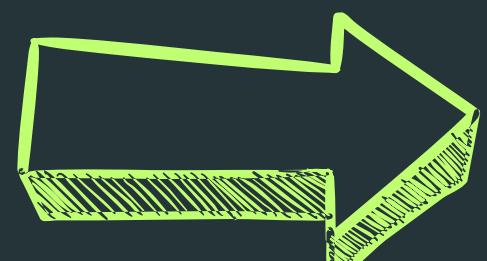
PRODUCER CONFIGURATION

PRODUCERS HAVE CONFIGURABLE
PARAMETERS FOR MESSAGE DELIVERY,
RETRIES, AND MORE.

EXAMPLE: SETTING PRODUCER
RETRIES TO HANDLE TRANSIENT
FAILURES.



Shwetank Singh
GritSetGrow - GSGLearn.com



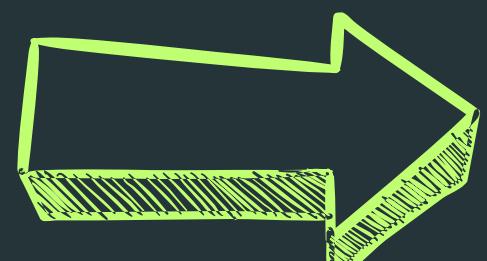
Kafka CONSUMER CONFIGURATION

CONSUMERS CAN BE CONFIGURED FOR
FETCH SIZES, TIMEOUT SETTINGS, AND
MORE.

EXAMPLE: CONFIGURING CONSUMER
FETCH SIZE FOR OPTIMAL
PERFORMANCE.



Shwetank Singh
GritSetGrow - GSGLearn.com



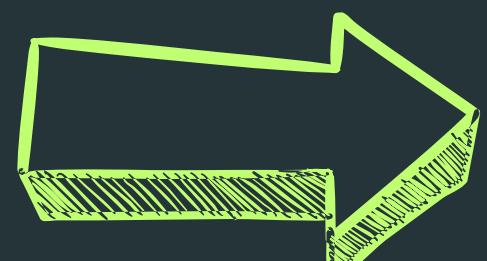
TOPIC MANAGEMENT

**TOPICS CAN BE CREATED, DELETED,
AND MANAGED PROGRAMMATICALLY
OR VIA CLI.**

**EXAMPLE: CREATING A NEW TOPIC
FOR STORING EVENT LOGS.**



Shwetank Singh
GritSetGrow - GSGLearn.com



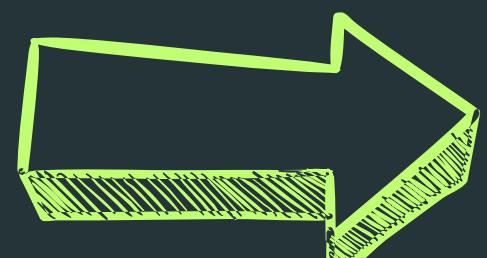
QUOTAS AND THROTTLING

KAFKA SUPPORTS SETTING QUOTAS TO CONTROL RESOURCE USAGE BY CLIENTS.

EXAMPLE: THROTTLING A HIGH-VOLUME PRODUCER TO PREVENT OVERWHELMING THE CLUSTER.



Shwetank Singh
GritSetGrow - GSGLearn.com



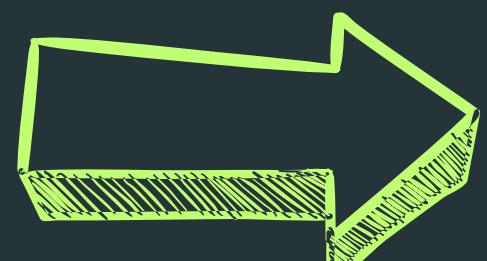
REBALANCE PROTOCOL

KAFKA HANDLES REBALANCING OF CONSUMERS WITHIN A GROUP TO MAINTAIN LOAD BALANCE.

EXAMPLE: REBALANCING PARTITIONS WHEN A NEW CONSUMER JOINS THE GROUP.



Shwetank Singh
GritSetGrow - GSGLearn.com



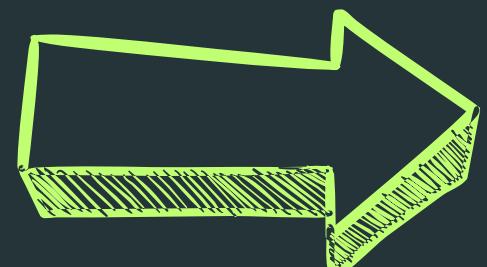
KAFKA REST PROXY

PROVIDES A RESTFUL INTERFACE TO
INTERACT WITH KAFKA CLUSTERS.

EXAMPLE: SENDING MESSAGES TO
KAFKA USING HTTP REQUESTS.



Shwetank Singh
GritSetGrow - GSGLearn.com



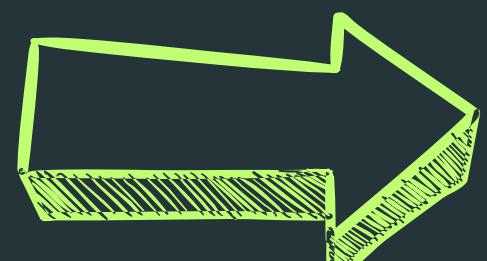
KAFKA API

KAFKA PROVIDES APIS FOR PRODUCING, CONSUMING, AND MANAGING DATA.

EXAMPLE: USING THE KAFKA PRODUCER API TO SEND MESSAGES FROM A JAVA APPLICATION.



Shwetank Singh
GritSetGrow - GSGLearn.com



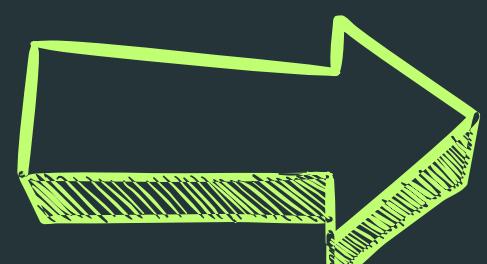
SCHEMA REGISTRY

**CONFLUENT SCHEMA REGISTRY
MANAGES AND ENFORCES SCHEMAS
FOR KAFKA MESSAGES.**

**EXAMPLE: ENSURING ALL MESSAGES
IN A TOPIC FOLLOW A PREDEFINED
SCHEMA.**



Shwetank Singh
GritSetGrow - GSGLearn.com



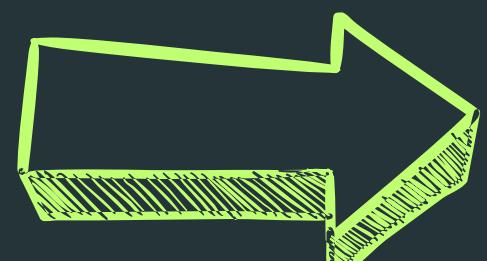
KAFKA STREAMS DSL

A HIGH-LEVEL API FOR STREAM
PROCESSING IN KAFKA.

EXAMPLE: USING KAFKA STREAMS DSL
TO FILTER AND TRANSFORM A STREAM
OF EVENTS.



Shwetank Singh
GritSetGrow - GSGLearn.com



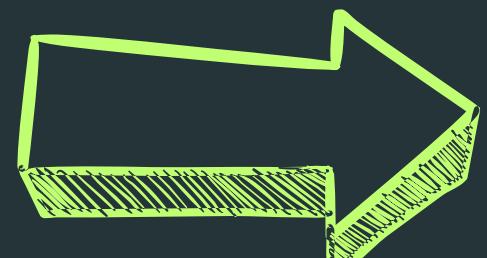
FAULT TOLERANCE

KAFKA'S DESIGN ENSURES HIGH AVAILABILITY AND FAULT TOLERANCE.

EXAMPLE: AUTOMATIC FAILOVER TO REPLICAS WHEN A BROKER FAILS.



Shwetank Singh
GritSetGrow - GSGLearn.com



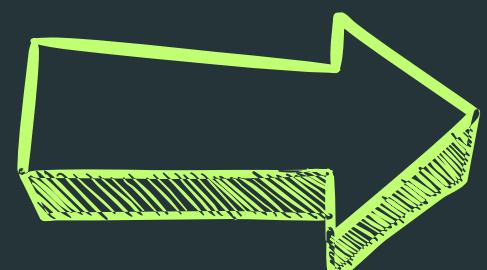
REAL-TIME ANALYTICS

KAFKA SUPPORTS REAL-TIME DATA ANALYTICS AND PROCESSING.

EXAMPLE: REAL-TIME DASHBOARD UPDATING WITH LIVE METRICS FROM KAFKA.



Shwetank Singh
GritSetGrow - GSGLearn.com



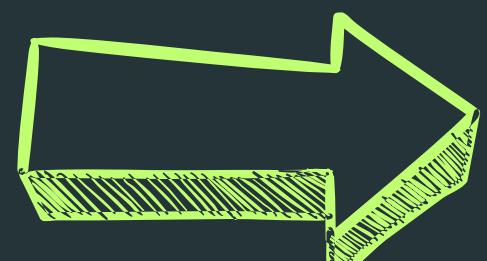
ETL PIPELINES

KAFKA CAN BE USED TO BUILD
EFFICIENT ETL PIPELINES FOR DATA
INTEGRATION.

EXAMPLE: EXTRACTING DATA FROM
DATABASES, TRANSFORMING IT, AND
LOADING IT INTO A DATA WAREHOUSE
VIA KAFKA.



Shwetank Singh
GritSetGrow - GSGLearn.com



KAFKA UPGRADES

**KAFKA SUPPORTS ROLLING UPDATES
TO MINIMIZE DOWNTIME.**

**EXAMPLE: UPGRADING KAFKA
BROKERS WITHOUT DISRUPTING
MESSAGE FLOW.**



Shwetank Singh
GritSetGrow - GSGLearn.com



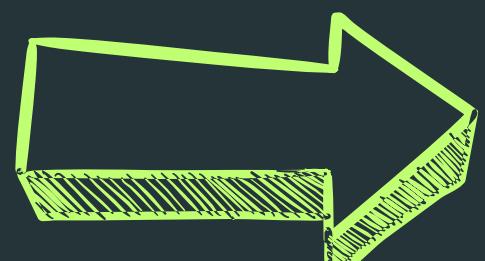
MESSAGE TIMESTAMPING

KAFKA MESSAGES CAN HAVE
TIMESTAMPS FOR TIME-BASED
PROCESSING.

EXAMPLE: USING TIMESTAMPS FOR
EVENT TIME PROCESSING IN KAFKA
STREAMS.



Shwetank Singh
GritSetGrow - GSGLearn.com



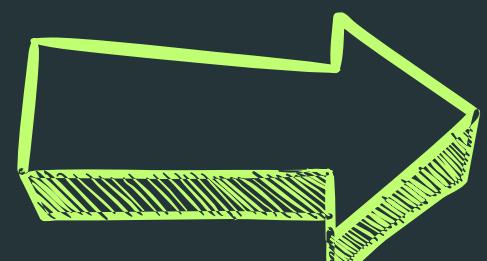
STATE STORES

KAFKA STREAMS ALLOWS
MAINTAINING STATEFUL PROCESSING
WITH STATE STORES.

EXAMPLE: COUNTING OCCURRENCES
OF EVENTS OVER A WINDOW OF TIME
USING STATE STORES.



Shwetank Singh
GritSetGrow - GSGLearn.com



WINDOWED OPERATIONS

KAFKA STREAMS SUPPORTS
WINDOWED OPERATIONS FOR
AGGREGATIONS OVER TIME WINDOWS.

EXAMPLE: CALCULATING THE SUM OF
TRANSACTIONS EVERY MINUTE.



Shwetank Singh
GritSetGrow - GSGLearn.com



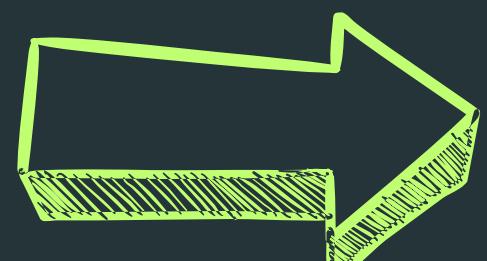
KSQL

KSQL IS A SQL-LIKE INTERFACE FOR STREAM PROCESSING WITH KAFKA.

EXAMPLE: USING KSQL TO PERFORM REAL-TIME FILTERING AND AGGREGATIONS ON KAFKA TOPICS.



Shwetank Singh
GritSetGrow - GSGLearn.com



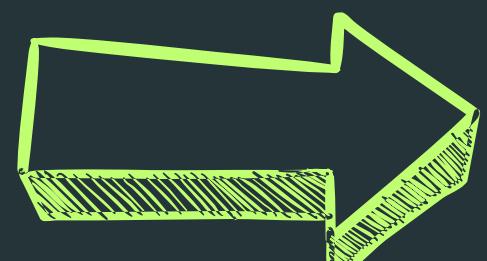
KAFKA ECOSYSTEM

KAFKA'S ECOSYSTEM INCLUDES TOOLS LIKE CONNECT, STREAMS, KSQL, AND MORE FOR COMPREHENSIVE DATA PROCESSING.

EXAMPLE: USING KAFKA CONNECT TO INTEGRATE WITH DATABASES, KAFKA STREAMS FOR PROCESSING, AND KSQL FOR QUERYING STREAMS.



Shwetank Singh
GritSetGrow - GSGLearn.com



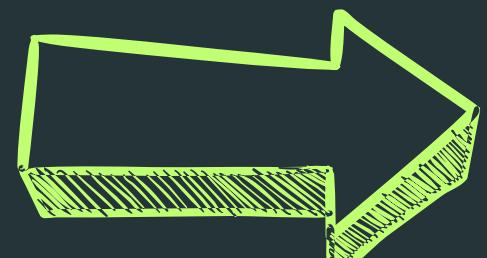
KAFKA CONNECTORS

PRE-BUILT CONNECTORS FOR
INTEGRATING KAFKA WITH VARIOUS
DATA SOURCES AND SINKS.

EXAMPLE: USING A JDBC CONNECTOR
TO SYNC DATA BETWEEN A DATABASE
AND KAFKA.



Shwetank Singh
GritSetGrow - GSGLearn.com



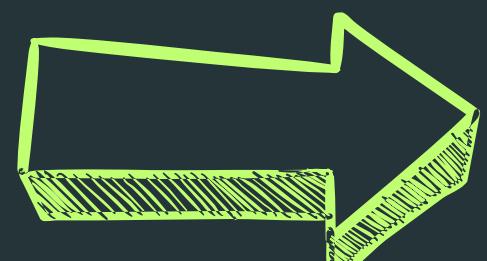
KAFKA CLUSTER MANAGEMENT

TOOLS AND PRACTICES FOR
MANAGING KAFKA CLUSTERS
EFFICIENTLY.

EXAMPLE: USING TOOLS LIKE KAFKA
MANAGER FOR MONITORING AND
MANAGING CLUSTER HEALTH.



Shwetank Singh
GritSetGrow - GSGLearn.com



TIERED STORAGE

KAFKA'S TIERED STORAGE ALLOWS OFFLOADING OLDER DATA TO CHEAPER STORAGE.

EXAMPLE: STORING OLDER KAFKA TOPIC DATA IN S3 TO REDUCE ON-PREM STORAGE COSTS.



Shwetank Singh
GritSetGrow - GSGLearn.com

