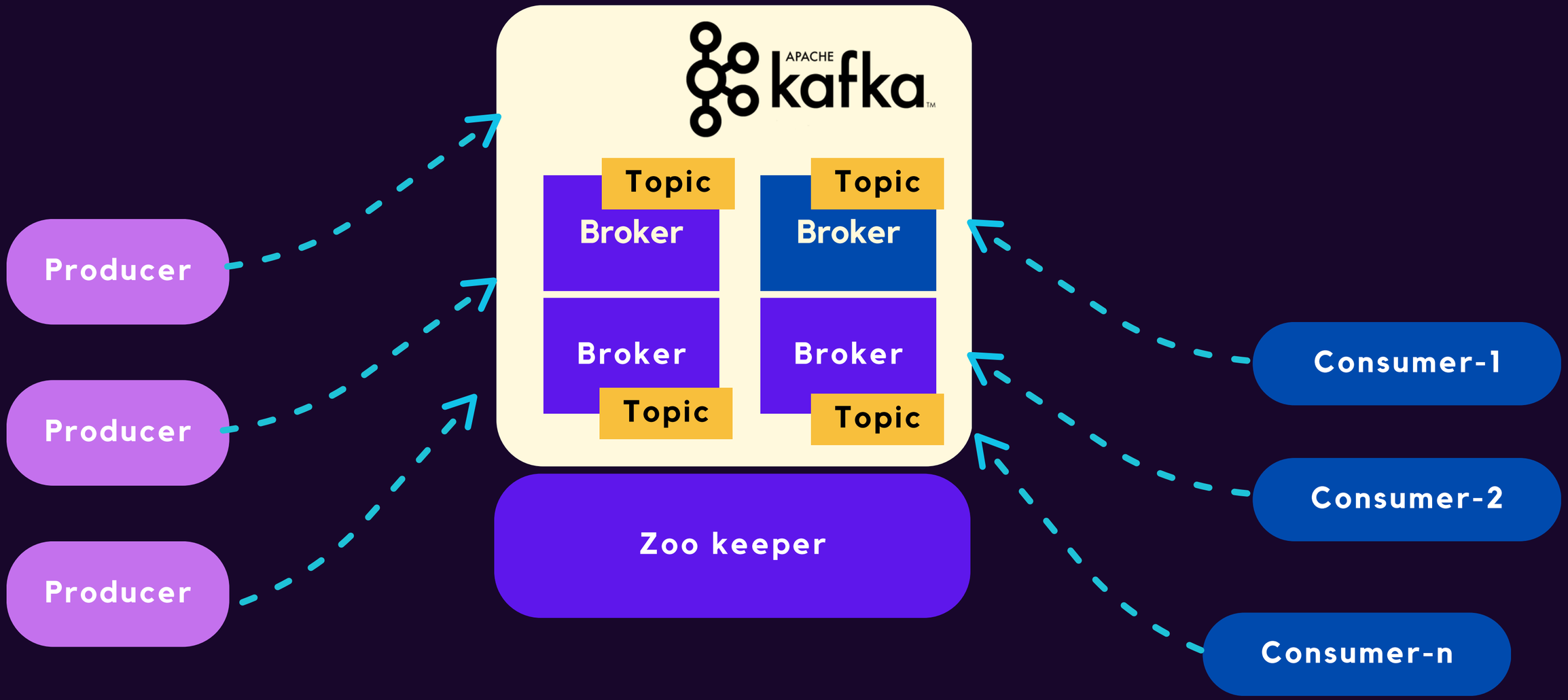
MINDSONAI

HELLO WORLD WITH APACHE KAFKA

https://linkedin.com/rishabhio

JIO MLOPS

Kafka Architecture Simplified



Core Concepts

Topics: Logical Channels for Data

Producers: Send Data to Topics

Consumers: Subscribe to Topics and Process Data

Distributed Architecture

Distributed System: Runs on a Cluster of Servers

Brokers: Kafka Server Instances

Topics Partitioning: Divides Data Across Multiple Brokers (Means a single partition is replicated multiple times)

Producers

Role: Generate and Publish Data to Topics

Flexibility: Can Publish to One or Multiple Topics

Example: Sending Logs, Metrics, or Events in Real-Time

Consumers

Role: Subscribe to Topics and Process Data

Groups: Consumers Can Be Organized into Groups for Parallel Processing

Example: Analyzing User Clicks, Processing Orders

Topics

Topics: Logical Channels for Data in Kafka

Partitioning: Scales Data Across Cluster

Retention Policy: How Long Data Persists in Topics

Fault Tolerance

Replication: Copies of Data Across Brokers

Leader-Follower Model: Ensures Data Redundancy and Availability

Fault Tolerance: No Single Point of Failure

Scalability

Horizontal Scaling: Add More Brokers to Handle More Data

Partitioning: Distributes Load Across Brokers

Elasticity: Scales Seamlessly with Data Growth

Use cases of Kafka Explained!

Real-Time Data Ingestion

Topic: "raw_data"

Publisher: Data producers/sensors/IoT devices

Subscriber: Data processing applications or streaming pipelines

Model Training Data Pipeline

Topic: "training_data"

Publisher: Data sources or data preprocessing pipelines

Subscriber: Model training infrastructure or streaming ML algorithms

Model Deployment and Serving

Topic: "inference_requests", "inference_results"

Publisher: Applications generating inference requests

Subscriber: Model serving infrastructure, applications consuming inference results

Feature Engineering Pipeline

Topic: "raw_data", "transformed_data"

Publisher: Data sources or preprocessing pipelines

Subscriber: Feature engineering pipelines or model training pipelines

Feedback Loops and Online Learning

Topic: "user_feedback", "model_updates"

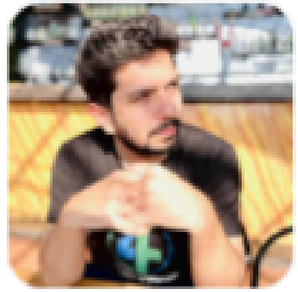
Publisher: Applications collecting user feedback, model evaluation systems

Subscriber: Model retraining pipelines, model serving infrastructure

Demo Time

rishabhio/**Distributed-Queue**

Let's Learn Kafka Distributed Queue by Exploring its functionality.



1 Contributor 0 Issues 0 Stars 0 Forks

rishabhio/Distributed-Queue: Let's Learn Kafka Distributed Queue by Exploring its functionality.

Let's Learn Kafka Distributed Queue by Exploring its functionality. - GitHub - rishabhio/Distributed-Queue: Let's Learn Kafka Distributed