Strimzi:
Data Streaming on Kubernetes with Apache Kafka

Jakub Scholz   Yaodong Yang

https://strimzi.io

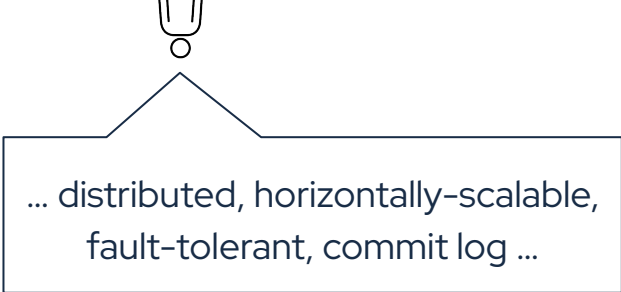**STRIMZI**

CLOUD NATIVE
COMPUTING FOUNDATION

What is Strimzi

- Open Source project / community (Apache License 2.0)

- CNCF Incubating project

- Focuses on Apache Kafka on Kubernetes

  - Based on the operator pattern

  - Provides operators for running and managing Apache Kafka and its components

  - Additional tools to make Apache Kafka easier to use on Kubernetes
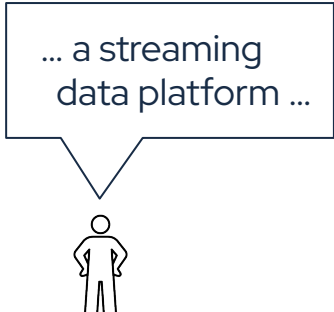
STRIMZI

Ecosystem

... distributed, horizontally-scalable, fault-tolerant, commit log ...

Apache Kafka

- Leading distributed message log and data streaming platform

- Open Source project
  - Originally developed by LinkedIn
  - Part of the Apache Software Foundation
  - Licensed under Apache License 2.0

... a publish/subscribe messaging system ...

... a streaming data platform ...

3

STRIMZI

Kubernetes-native Kafka

- Follows the operator pattern

    - Uses CRDs to extend Kubernetes API and define the Kafka resources

    - Encodes the Kafka operations knowledge into the operator code

    - Automates the installation, operations and monitoring tasks

- Examples where Strimzi operators help most

    - Handling the Kafka discovery protocol, Upgrades, Security, Scaling

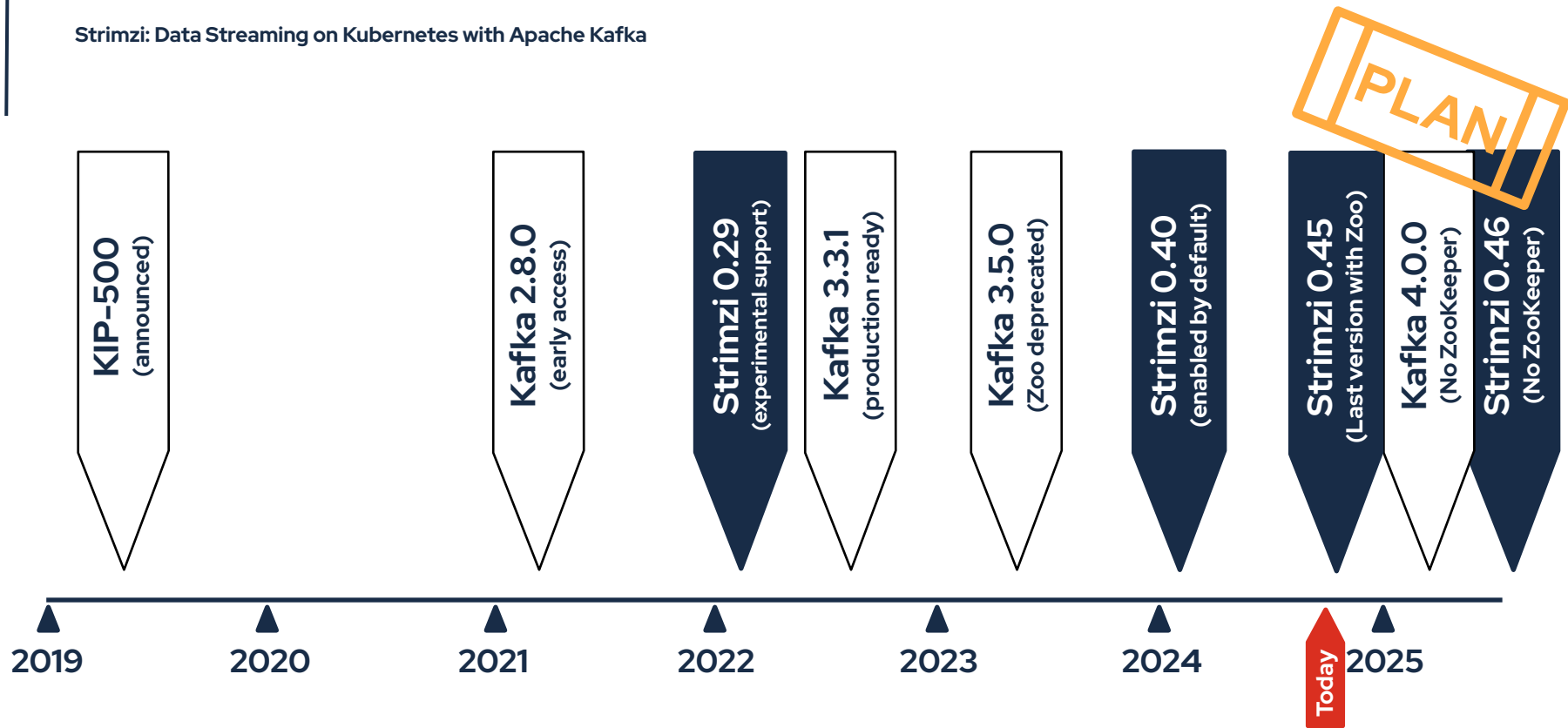**STRIMZI**

Kubernetes-native Kafka

- Support for all Apache Kafka components & more

  ○ ZooKeeper, Kafka brokers, Kafka Connect, Kafka Mirror Maker (1 and 2)

  ○ Our own HTTP Bridge

  ○ Cruise Control for *balancing* the Kafka cluster

  ○ The components can be used independently or together

- Makes it easier to use Kafka

  ○ Manages topics, users or connectors through the operator pattern as well
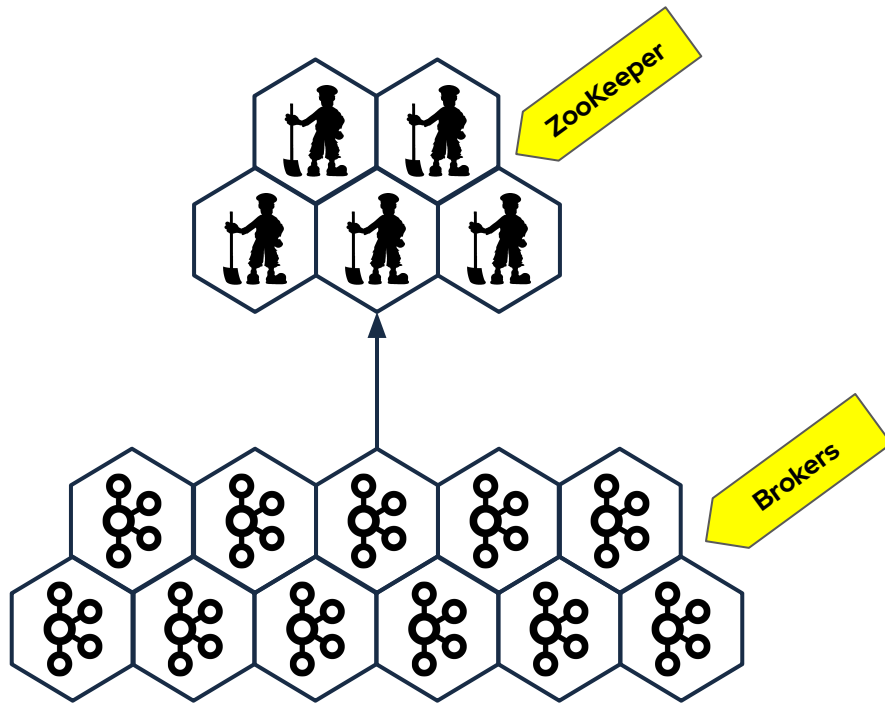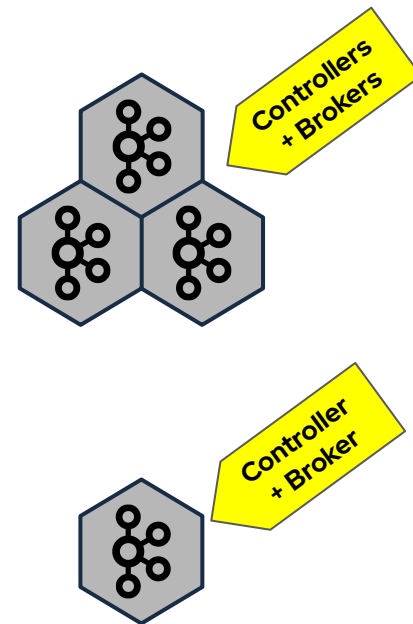
STRIMZI
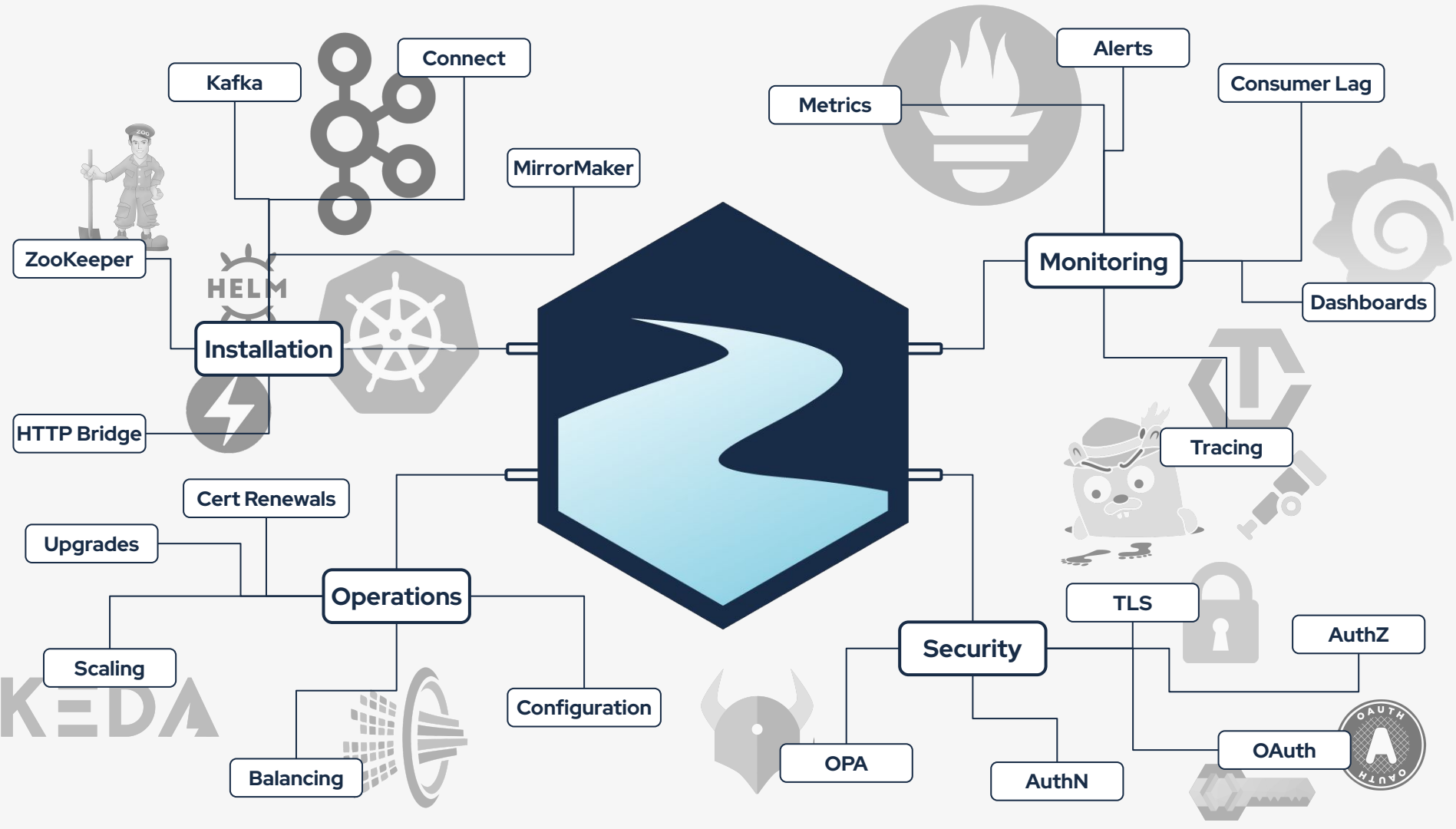
KRaft

KRaft

- Kafka is removing its dependency on ZooKeeper

  - Replaced with Kafka's own implementation based on the Raft protocol

  - Changes the APIs and architecture of the Kafka clusters

  - Lot of effort is dedicated to supporting the new KRaft mode and a smooth transition

  - Coming to an end after 5 years!

STRIMZI

Strimzi: Data Streaming on Kubernetes with Apache Kafka

PLAN

KIP-500 (announced)

Kafka 2.8.0 (early access)

Strimzi 0.29 (experimental support)

Kafka 3.3.1 (production ready)

Kafka 3.5.0 (Zoo deprecated)

Strimzi 0.40 (enabled by default)

Strimzi 0.45 (Last version with Zoo)

Kafka 4.0.0 (No ZooKeeper)

Strimzi 0.46 (No ZooKeeper)

2019   2020   2021   2022   2023   2024   Today   2025

9
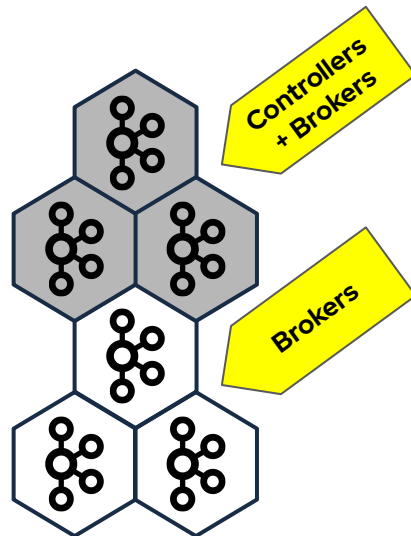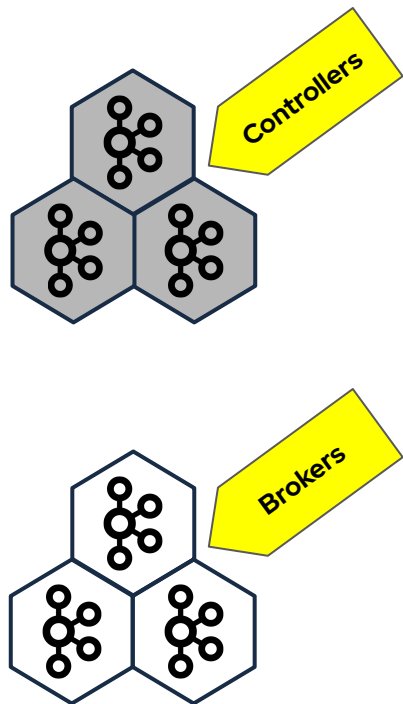
STRIMZI

Timeline

- Strimzi 0.45.0 (next release, end of 2024)

  - Supports Apache Kafka 3.8.0 and 3.9.0

  - **Last version with ZooKeeper support**

  - We plan to provide "extended support"

  - **ZooKeeper-based Kafka clusters need to be migrated to KRaft**

STRIMZI

Timeline

- Strimzi 0.46.0 (early 2025)
  - Supports Apache Kafka 3.9.0 and 4.0.0
  - **Only KRaft-based Kafka clusters supported**
  - Mirror Maker 1 support will be removed as well
  - **ZooKeeper-based Kafka clusters need to be migrated to KRaft before upgrading to Strimzi 0.46.0**

Migration

- Existing ZooKeeper based clusters can be migrated to KRaft

  - Migration to KRaft is driven by users through annotations

  - Cannot be fully automatic because of new configurations, architectures, …

  - Migration has to happen before upgrading to Strimzi 0.46.0

- KRaft and migration are already supported in Strimzi => you can use it today!

Learn more about KRaft

https://strimzi.io/kraft/

STRIMZI

Title

- Point one
  - Subpoint
- Point two
  - Subpoint

STRIMZI

Tiered Storage

STRIMZI

Tiered Storage

- Move data from expensive local disk to remote storage

    - Significantly reduce storage cost

    - Broader use cases adoption

    - Better scalability and reliability

    - Reduced operation overhead
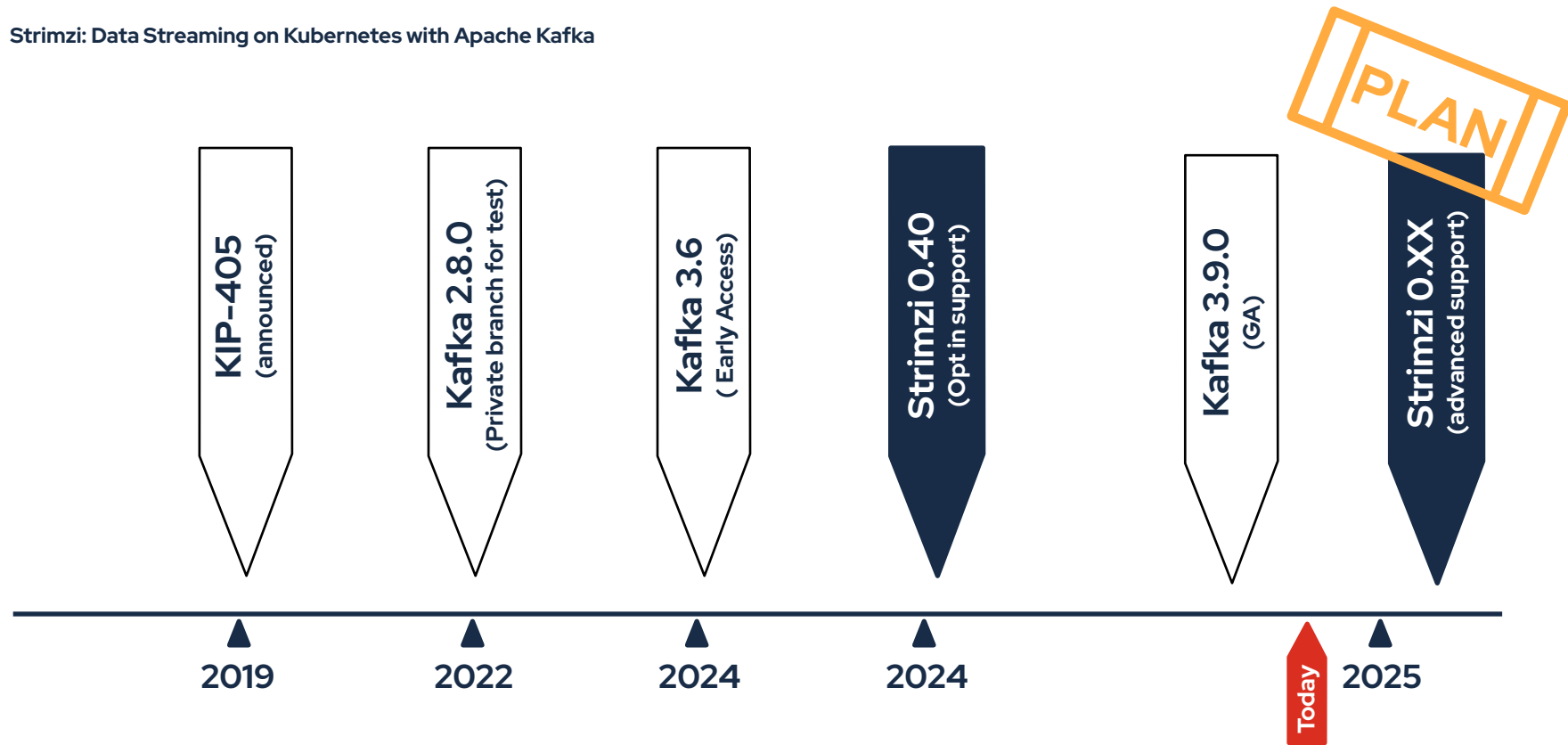
STRIMZI

Tiered Storage

- Native support for tiered storage configuration

- Defaulting to use TopicBasedRLMM class for metadata

- Custom Plugin library required for RLM class implementation

STRIMZI

Tiered Storage

```
kafka:
  tieredStorage:
    type: custom
    remoteStorageManager:
      className: com.example.kafka.tiered.storage.s3.S3RemoteStorageManager
      classPath: /opt/kafka/plugins/tiered-storage-s3/*
      config:
        # A map with String keys and String values. Key fields will be automatically
        # prefixed with `rsm.config.` and appended to Kafka broker config.
        storage.bucket.name: my-bucket
  config:
    ...
    rlmm.config.remote.log.metadata.topic.replication.factor: 1
```

STRIMZI

Tiered Storage

- Limitation
  - Compacted topics not supported
  - Disabling tiered storage supported only in KRaft mode
- Plugins options
  - Aiven's open source plugins: https://github.com/Aiven-Open/tiered-storage-for-apache-kafka
    - Amazon S3, Google Cloud Storage, Azure Blob Storage and file storage
  - Custom implementation

STRIMZI

Strimzi: Data Streaming on Kubernetes with Apache Kafka

PLAN

KIP-405 (announced)	Kafka 2.8.0 (Private branch for test)	Kafka 3.6 (Early Access)	Strimzi 0.40 (Opt in support)	Kafka 3.9.0 (GA)	Strimzi 0.XX (advanced support)
2019	2022	2024	2024	Today 2025	

STRIMZI

Auto-rebalancing

STRIMZI

Auto-rebalancing

- Automatically triggers a rebalance when brokers are added or removed
    - Based on Cruise Control integration
    - Makes it easier to scale the Kafka cluster up or down
    - Moves partition replicas to new brokers after they were added
    - Moves partition replicas from brokers before they are removed

STRIMZI

Future plans

STRIMZI

Future plans

- **v1 CRD APIs and Strimzi 1.0.0**

- Better support for external certificate management tools

- Kafka cluster self-healing

- Built-in Gateway API support for exposing Kafka clusters

- Kafka clusters stretched across multiple Kubernetes clusters

- and more ...

STRIMZI

v1 APIs and Strimzi 1.0.0

- Once ZooKeeper is removed
  - We will start working on the new `v1` API (CRDs)
  - The main change is expected to be restructuring after ZooKeeper removal
- Strimzi 1.0.0 should follow after the `v1` API is rolled out

STRIMZI

Better CA abstraction

- Better integration with external providers of TLS certificates
    - Pluggable architecture which can be more easily extended
    - Better support for common tools such as Cert Manager
    - Makes it easier to integrate into user's own certificate management processes and systems
- Possibly disabling TLS completely?

STRIMZI

Kafka Cluster self-healing

- Strimzi already integrates with Cruise Control for cluster rebalancing

    - Users interact with it through the `KafkaRebalance` resources

- We plan to integrate Cruise Control self-healing feature as well

    - Uses anomaly detection to find out about partition, disk, or broker failures

    - Where possible uses self-healing to fix the problem

Gateway API support

- Gateway API aims at replacing Ingress

  - Should provide better compatibility between providers

- Strimzi-based Kafka clusters can be already exposed using a Gateway API

  - Requires manual configuration of the Gateway API resources, advertised hosts etc.

- We plan to provide built-in support for Gateway API to make it easier to use!

STRIMZI

Stretch clusters

- Different parts of the Kafka cluster running on different Kubernetes clusters
    - Kafka is latency sensitive, so running brokers on different continents might never be possible
        - But should work fine with co-located Kubernetes cluster
    - Makes it easier to migrate between clusters
    - Suitable for *Metropolitan Area Networks / Metro DR*

STRIMZI

Join us

https://strimzi.io/join-us/

STRIMZI

Other talks

STRIMZI

Other talks

- Strimzi and the future of Apache Kafka on Kubernetes
  - Project lightning talks, Tue 12th November

- Elastic Data Streaming: Autoscaling Apache Kafka
  - Data Processing + Storage, Thu 14th November

STRIMZI

https://strimzi.io/

https://github.com/strimzi

@strimziio

https://youtube.com/c/Strimzi

https://www.linkedin.com/company/strimzi

STRIMZI

Demo

- External access
- Broker re-configuration
- Cluster rebalancing

STRIMZI