

Kubernetes as Enabler for a Data Analytics Platform

Data, Analytics & AI Group

Sammy Ramareddy



Agenda

1. Intro

2. Firehose 2.0

3. Kubernetes



Intro

Data Analytics Pipeline



Data Sourcing

Find the relevant data, find out how it can be sourced, what is the data quality, ...



Compute/Transform Data

Combine one or more data sources and get insights

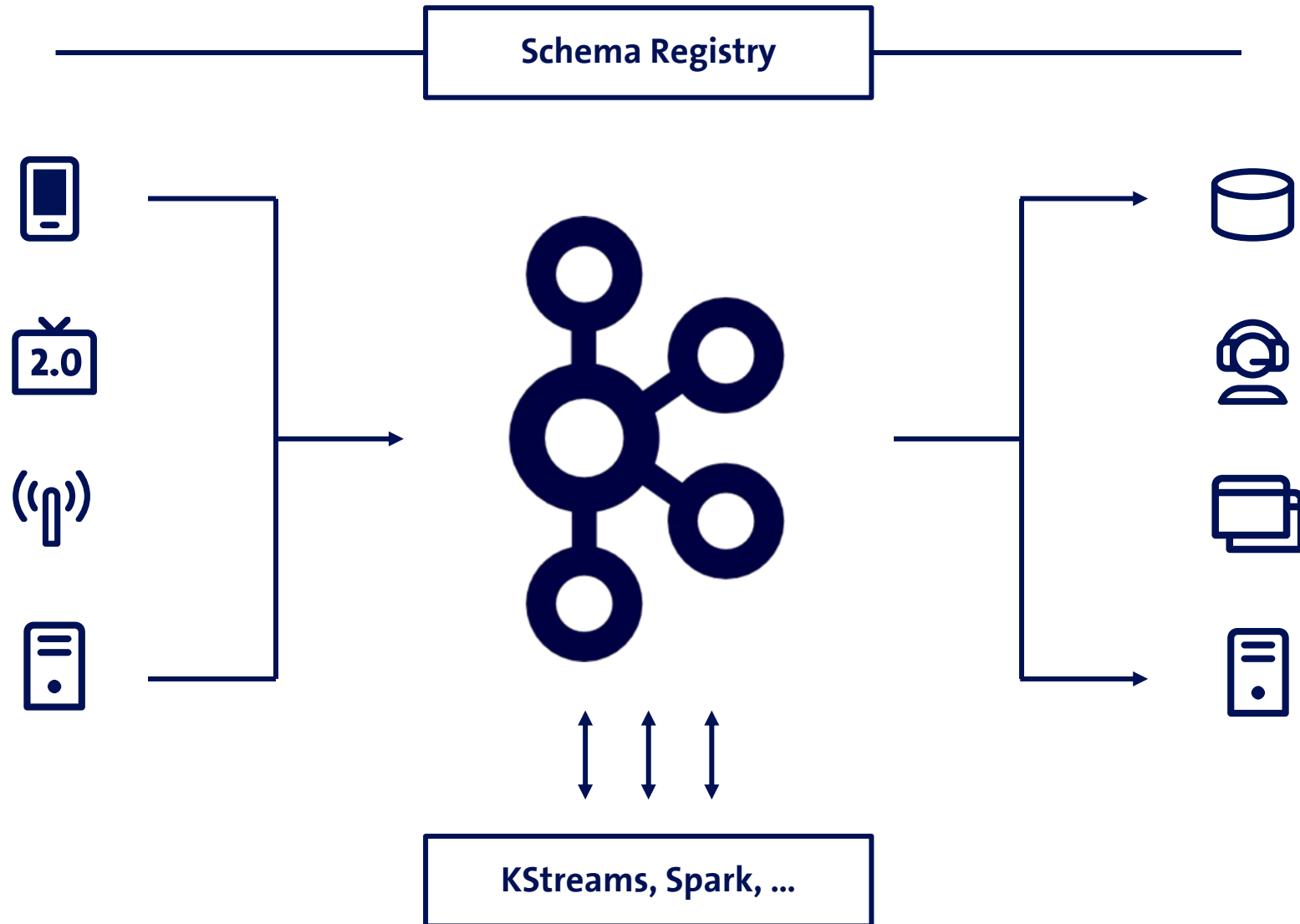


Present Computed Data

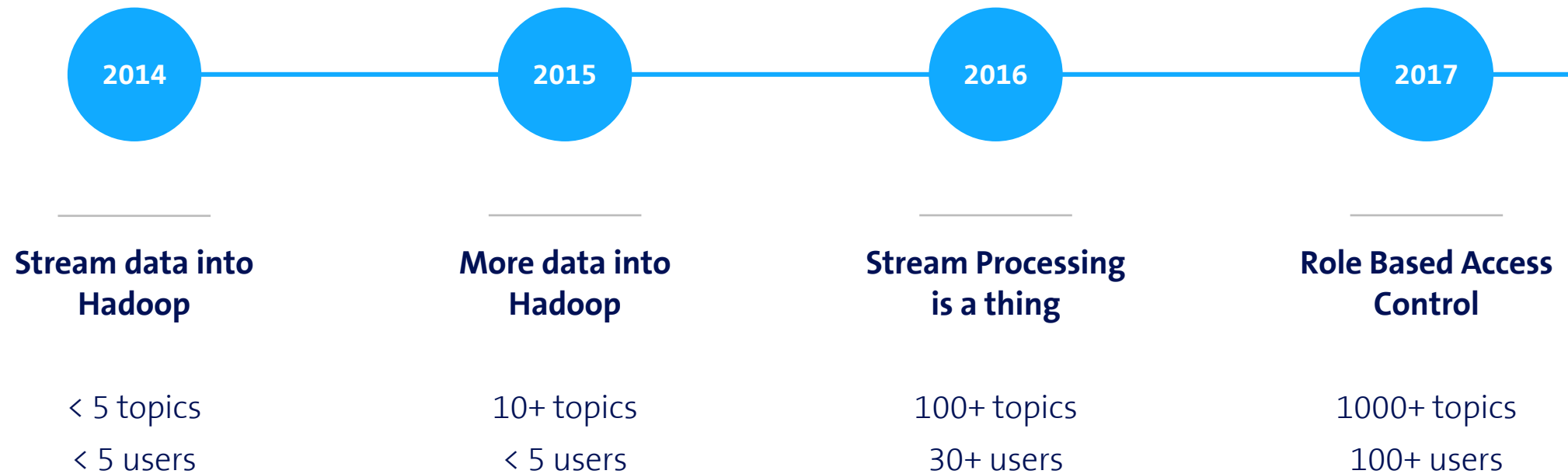
Store and present the insights in a useful manner for end-users

Batch vs Real time / Streaming

Swisscom's Streaming Platform



History of Kafka @ Swisscom



I want moore!!!



How to scale



Challenges

Lots of requests for topics
creation, modification,
access, ...

Who does approve the access
to a topic?

Policies are hard to maintain
over time using existing tools




Solutions

Self-Service API

Decentralized Approval
Process

Define a higher-level
Abstraction for Governance

A background image of the Golden Gate Bridge in San Francisco, viewed from a low angle looking up at the tower. The bridge is illuminated with warm lights, and the sky is a mix of orange and blue from the setting or rising sun. The water of the bay is visible in the foreground.

Firehose 2.0

Data Analytics Streaming Platform

Simple yet powerful concepts



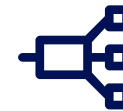
Organization

Group of **individuals** working on something together (e.g. project), characterized by a **name**.



Space

Group of **isolated resources** (e.g. Topics) belonging to an organization, characterized by a **name** and a **classification**.



Topic

Kafka Topic characterized by a **name**, a **description**, a **schema**, a **classification** (C1-C4), and a **shareable flag**.



Identity

Identity delivered to a Kafka Client, characterized by a **name**, a **network zone**, a **set of IPs**, and a **certificate**. Identity has RW access to all resources within its space.



Constraints on
Classification <-> Network Zone

Clear roles and responsibilities



Org. Owner

Responsible of the resources (usage, legal, approvals, etc.) **and the members** within his organization.



Org. Member

Use the resources within his organization.



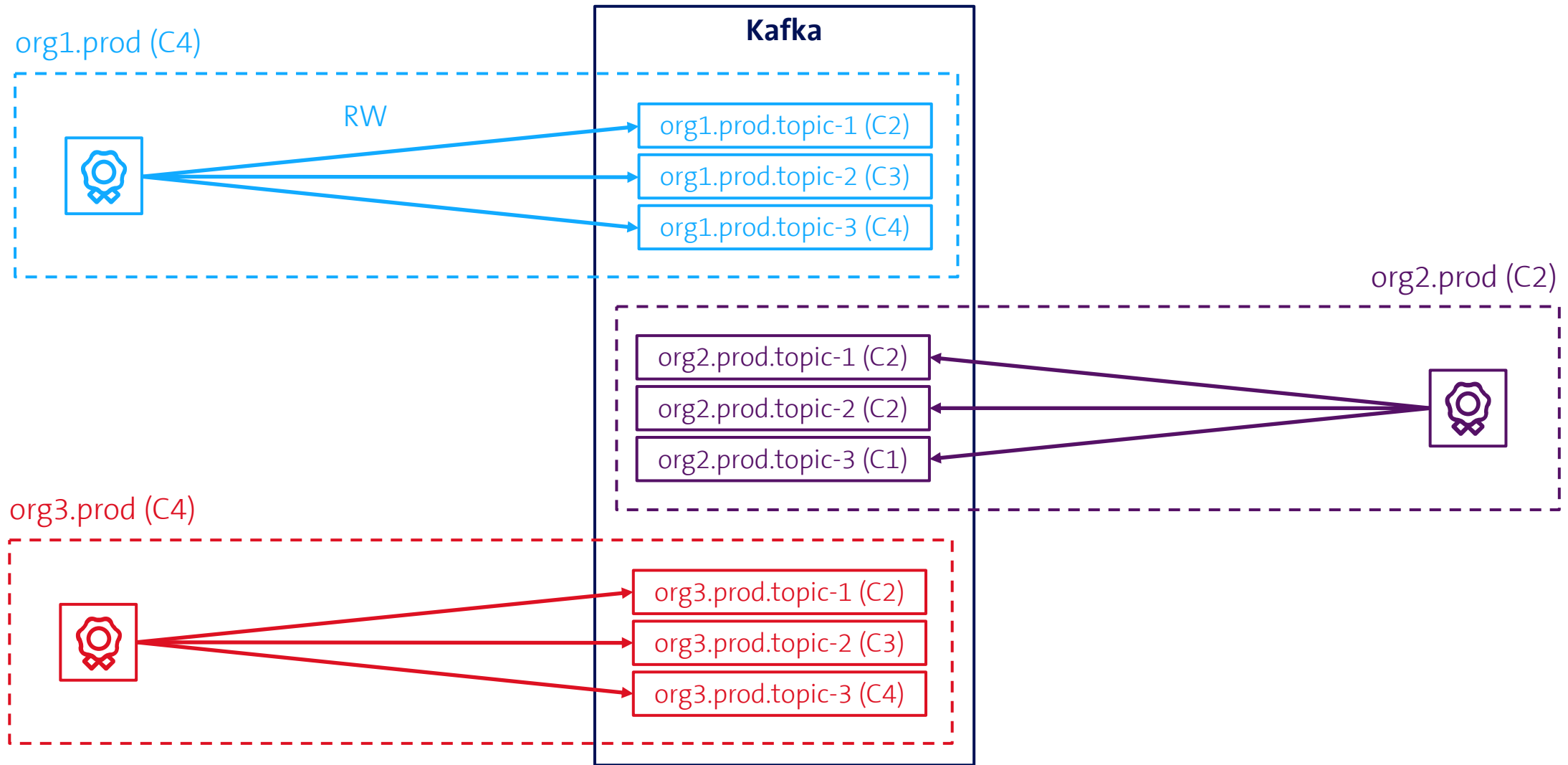
Admin

Enforce the security and governance rules by technical means.



Auditor

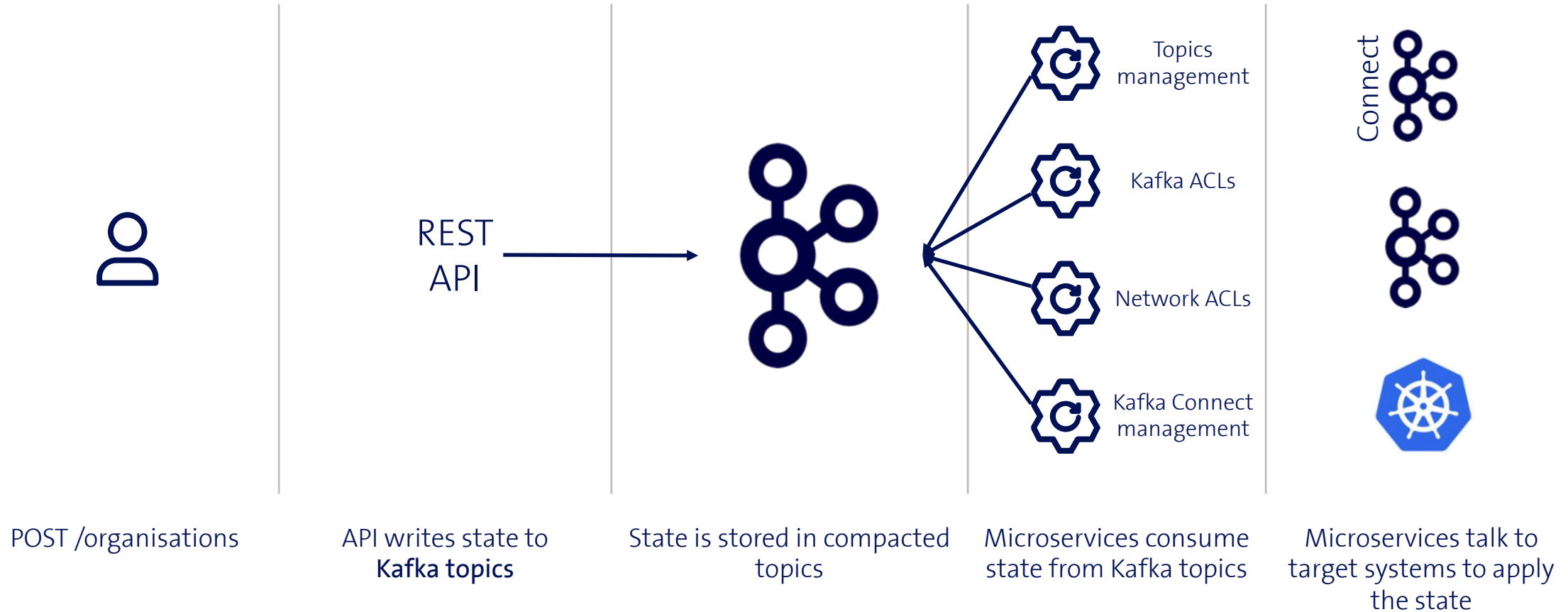
Audit all the organizations to ensure rules are enforced and respected.



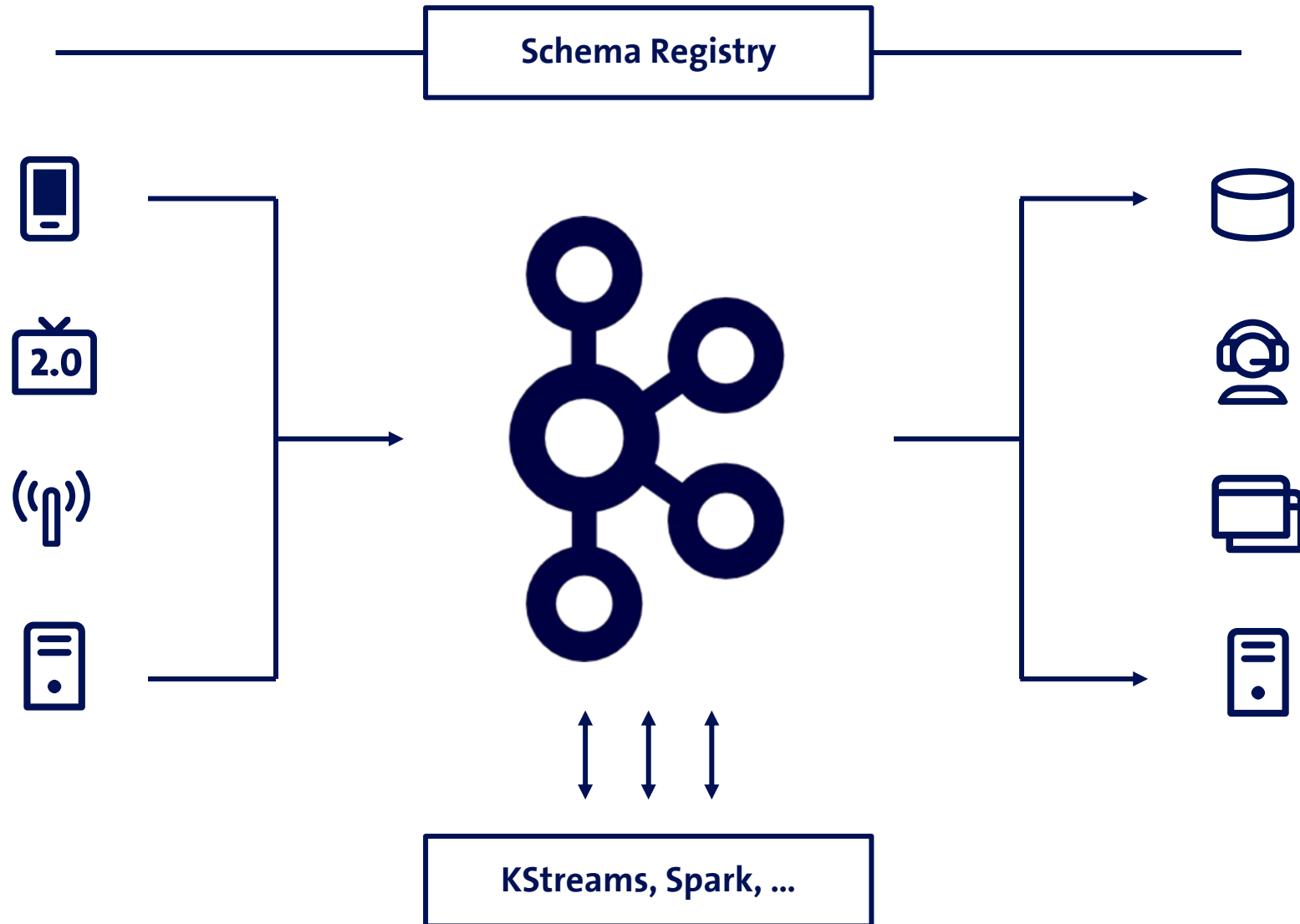
How do we share topics between organizations/spaces?



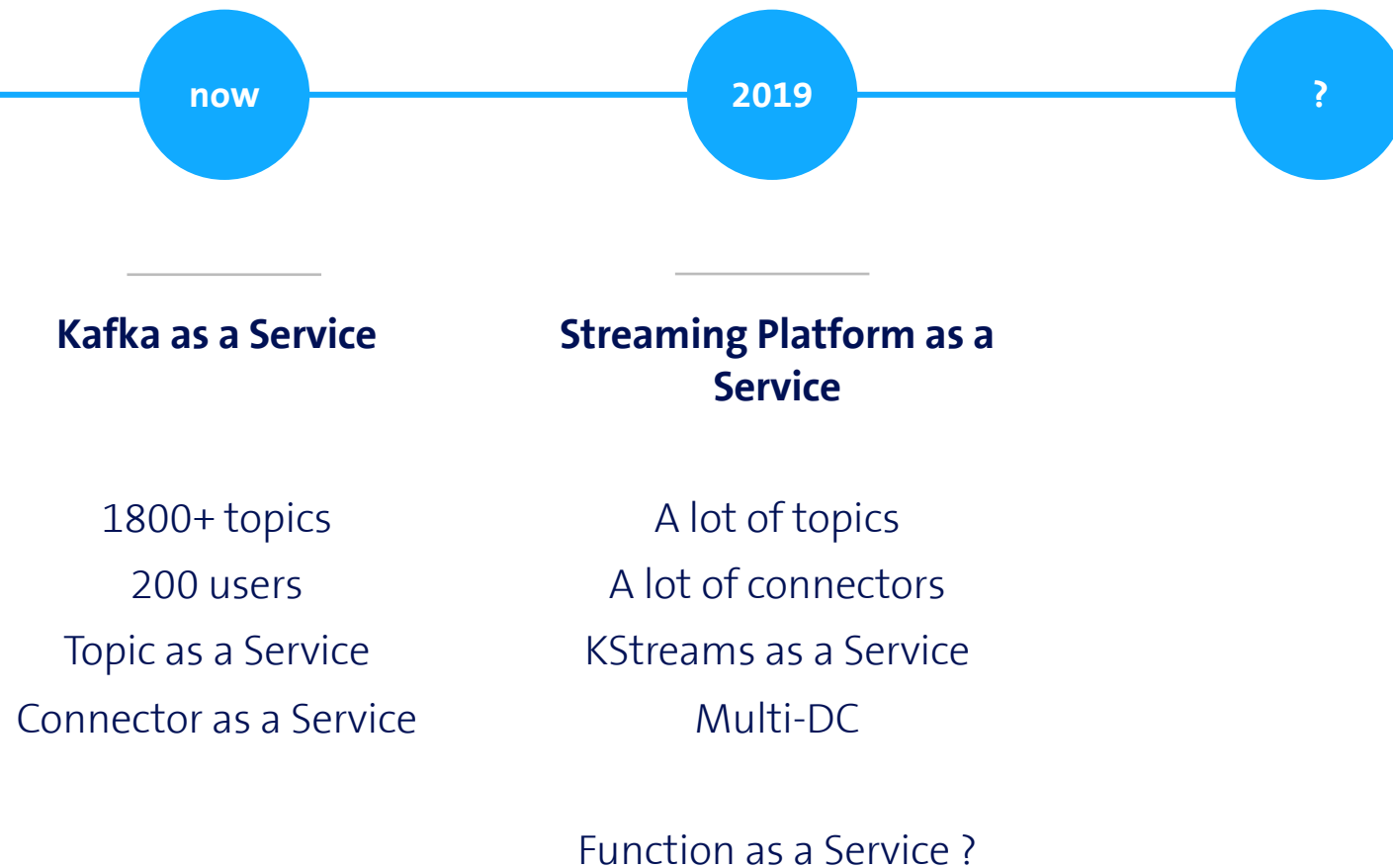
API Design



Swisscom's Streaming Platform



History of Kafka @ Swisscom





Kubernetes

as Enabler for a Data Analytics Streaming Platform

Kubernetes

Production-Grade Container Orchestration



Service discovery and load
balancing



Automated rollouts and
rollbacks



Self-healing



Horizontal scaling
and
Automatic binpacking



Secret and Configuration
Management



Why Kubernetes?



Scheduler



Network Policies



Containers

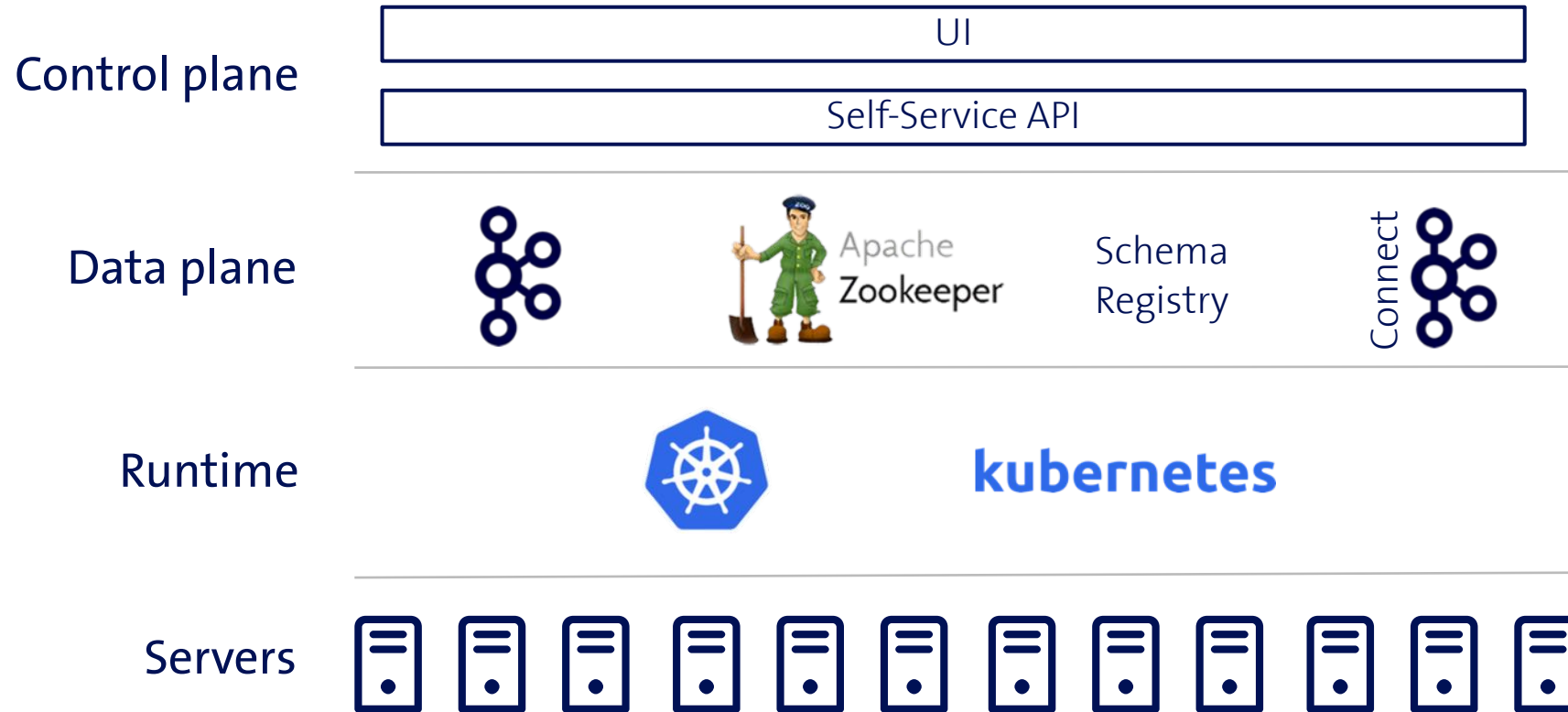


Increase hardware
utilization

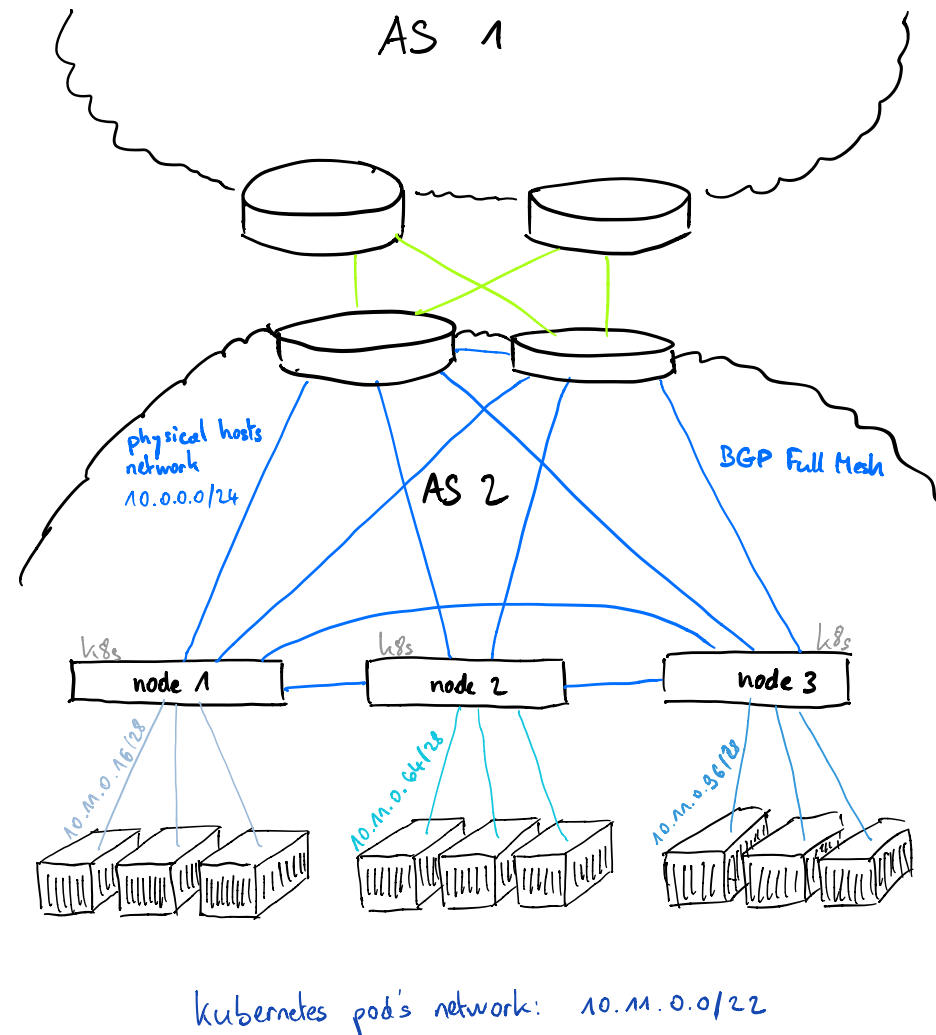


It rocks!

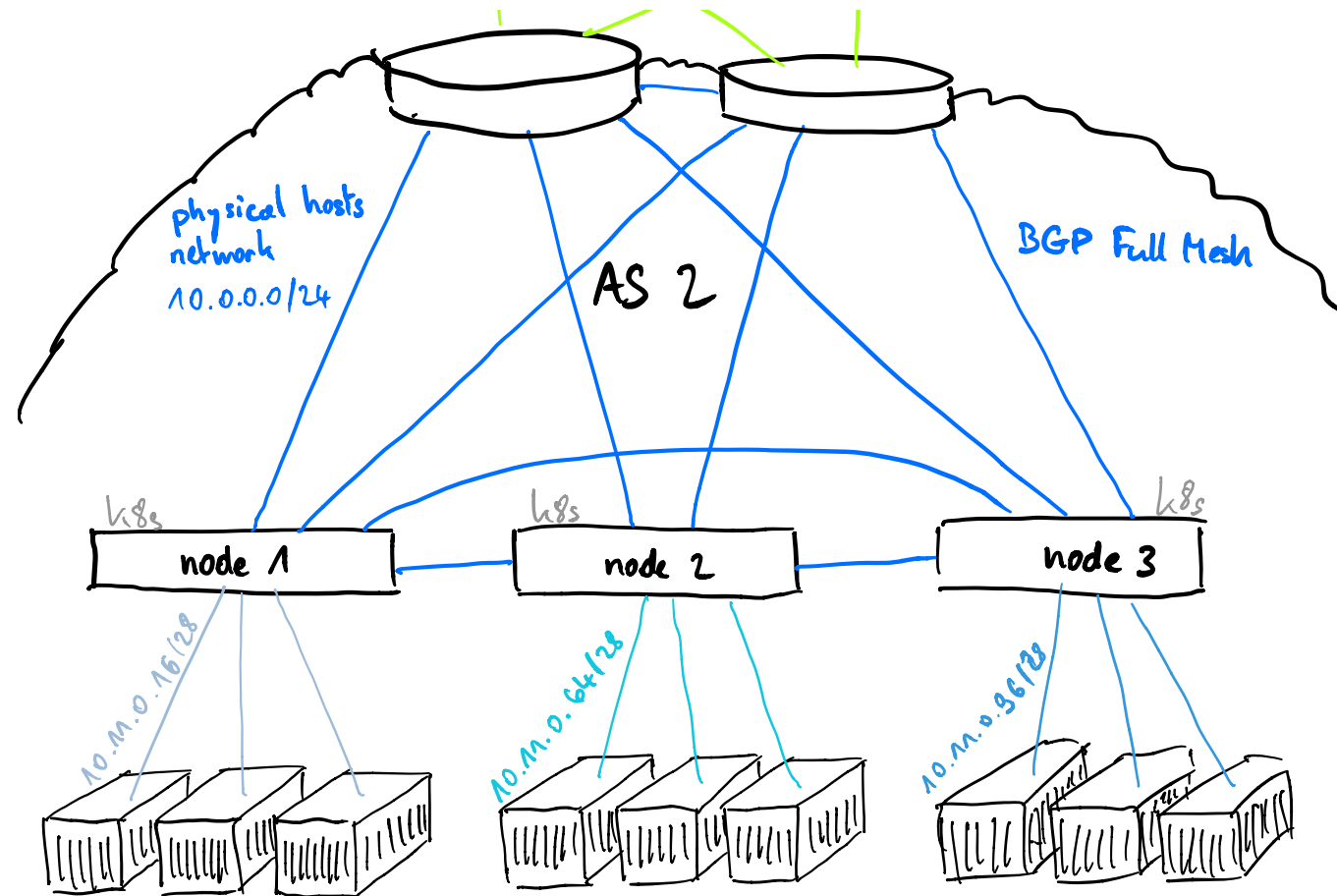
Architecture



Network Topology

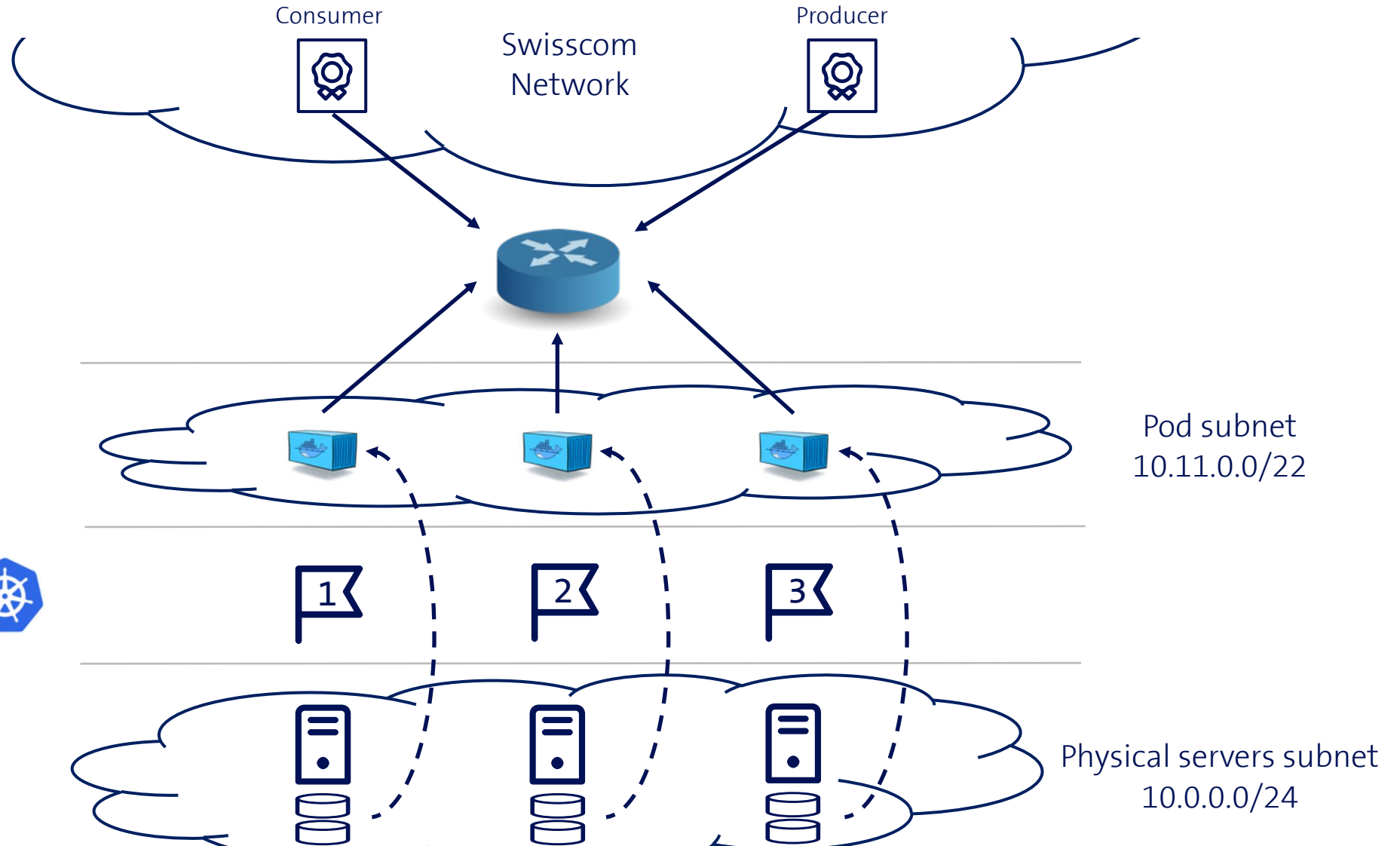


Network Topology



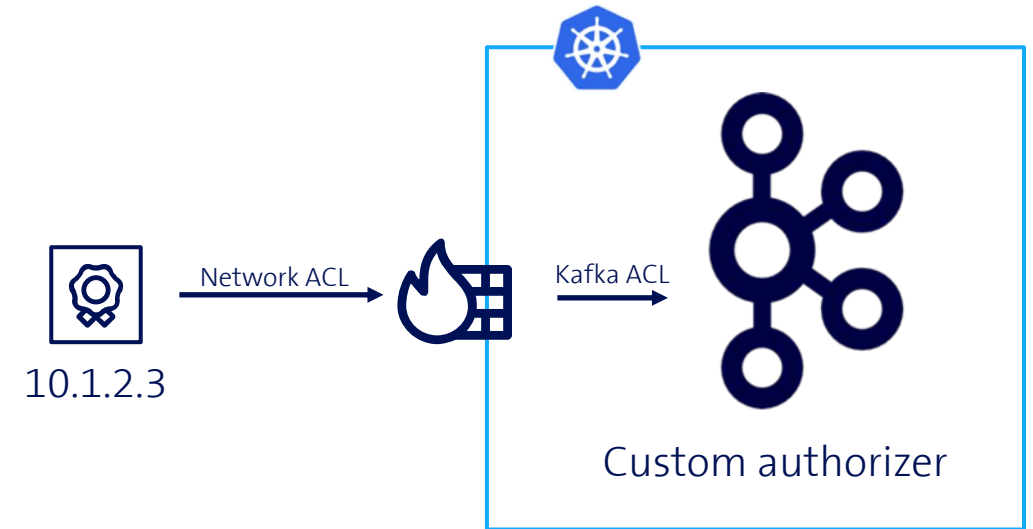
kubernetes pod's network: 10.11.0.0/22

Kafka on Kubernetes



Kafka and Network ACLs

1. Identity created
2. Identity approved
3. Network ACL microservice:
 - creates Kubernetes network policy
4. Kafka ACL microservice:
 - adds Read/Write ACL for <org>.<space>.*
 - adds Read Only ACL for shared topics
5. User generates a certificate and starts to consume/produce to topics belonging to the space



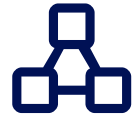
What's next?



Improved management of
physical volumes and better
integration with Kafka



Quotas for Kafka clients



Firehose in second data
centre for geo-redundancy



Transformation as a Service
Function as a Service
Container as a Service



Include other platforms in our
new governance framework



Thank you!

Neuchatel, 27th of September 2018



By the way....

We are hiring 🤗