KubeCon

CloudNativeCon

THE LINUX FOUNDATION

OPEN SOURCE SUMMIT

AI_dev

Open Source GenAI & ML Summit

China 2024

Multi-Cluster Networking and Service Discovery Leveraging NRI

基于NRI的多集群网络和服务发现

夏令明 紫金山实验室

Di Xu (徐迪, Github: @dixudx, CNCF Ambassador) Xiaohongshu(小红书)

Agenda

- Why Multi-Cluster
- Challenges in Multi-Cluster Networking
- Introducing FleetBoard
- Demo & Some Use Cases
- Q&A

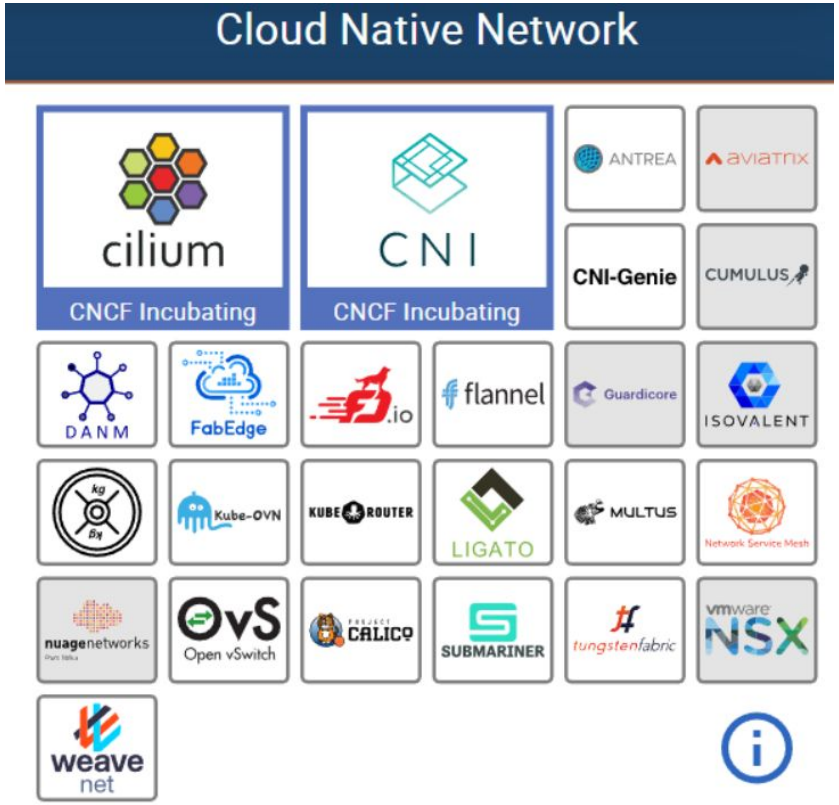
Why Multi-Cluster

- Business Needs
  - Compliance / Isolation / Availability / Security / Latency / etc.
  - Multi-Cloud Strategy
    - avoid vendor lock-in
- HARD to maintain a very large cluster
  - single cluster has its own limitations too
    - 5k nodes / 15k pods / 30w containers
  - Upgrading / Performance / Latency / Data Backup & Recovery / Failover
- Don’t put all your eggs in one basket!

Where We Are Now

- Multi-Cluster Management & Scheduling
  - Stitch Multiple Clusters into One
  - ~~KubeFed~~ / Clusternet / Karmada / OCM /…
- Multi-cluster networking
  - Submariner (https://github.com/submariner-io/submariner)
  - Still outstanding
- Multi-Cluster Service Discovery

Challenges in Multi-Cluster Networking



- Deployment Complexity
  - Requires Broker Deployment
  - Copy Broker Token
  - Manually select a node as gateway
  - Set up gateway node IP

- Numerous Limitations
  - Requires Public IPs
  - Pre-planning of non-conflicting CIDR ranges is essential
  - Poor CNI (Container Network Interface) compatibility

Challenges in Multi-Cluster Networking (cont.)

- Highly Invasive
  - Mixing network policies with node policies
  - Tunnels are visible on the node

- Architectural Complexity
  - High number of tunnels
  - Limited support for large number of clusters
  - Increased Complexity
  - Higher instability and management challenges

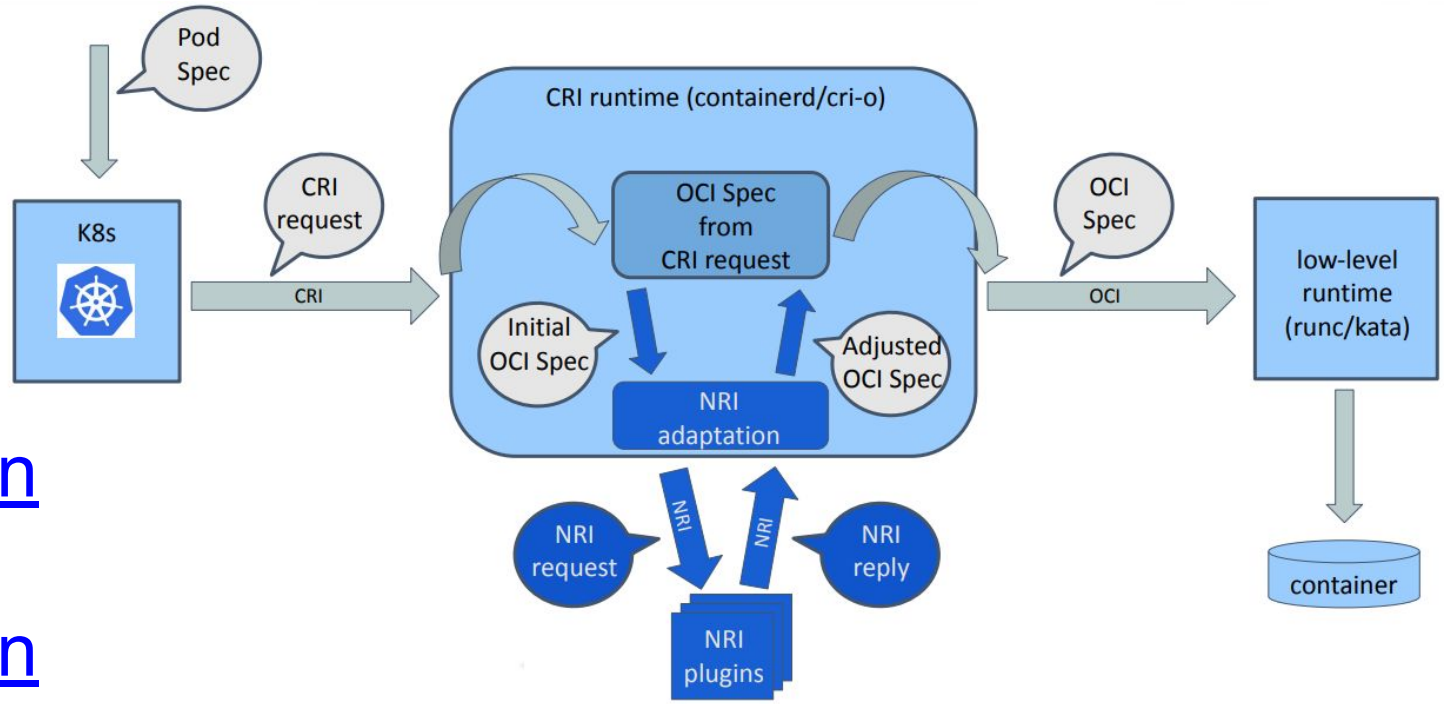
About FleetBoard

- Multi-cluster Interconnection and Service Discovery Management
- Consistent Service Discovery Mechanism
- Accessing Cross-Cluster Services: simple as visiting local clusters
- Configurable Network Tunnel Settings
  - Hub or Dedicate Gateway
- Github: https://github.com/fleetboard-io/fleetboard
  - First Commit (Dec. 2023)

Leveraging NRI

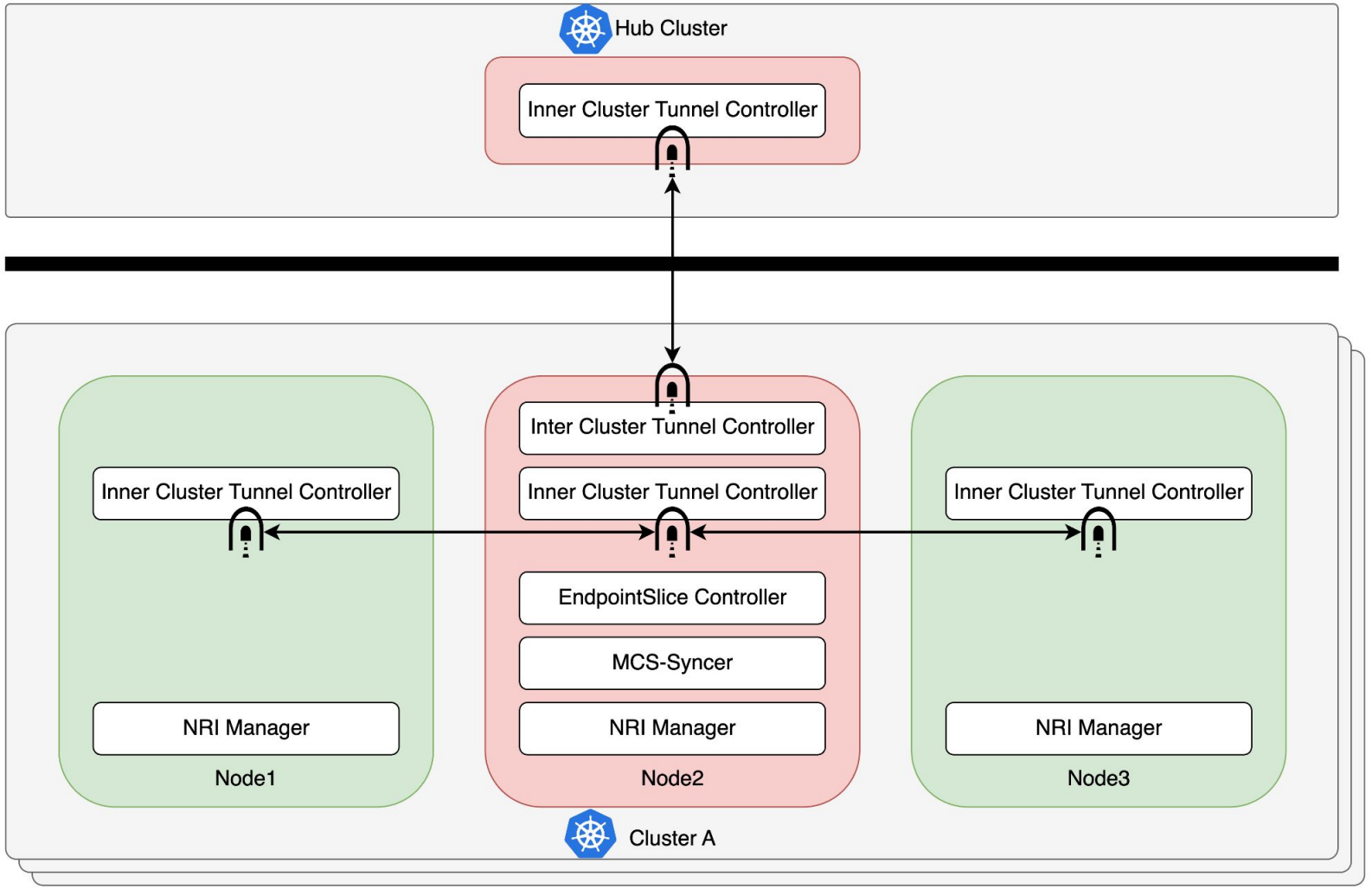
NRI (Node Resource Interface)

- containerd 1.7+
- CRI-O 1.26+
- Mutating / Non-blocking
- https://github.com/containerd/nri/
- https://github.com/containers/nri-plugins
  - Resource policy / Memory Management



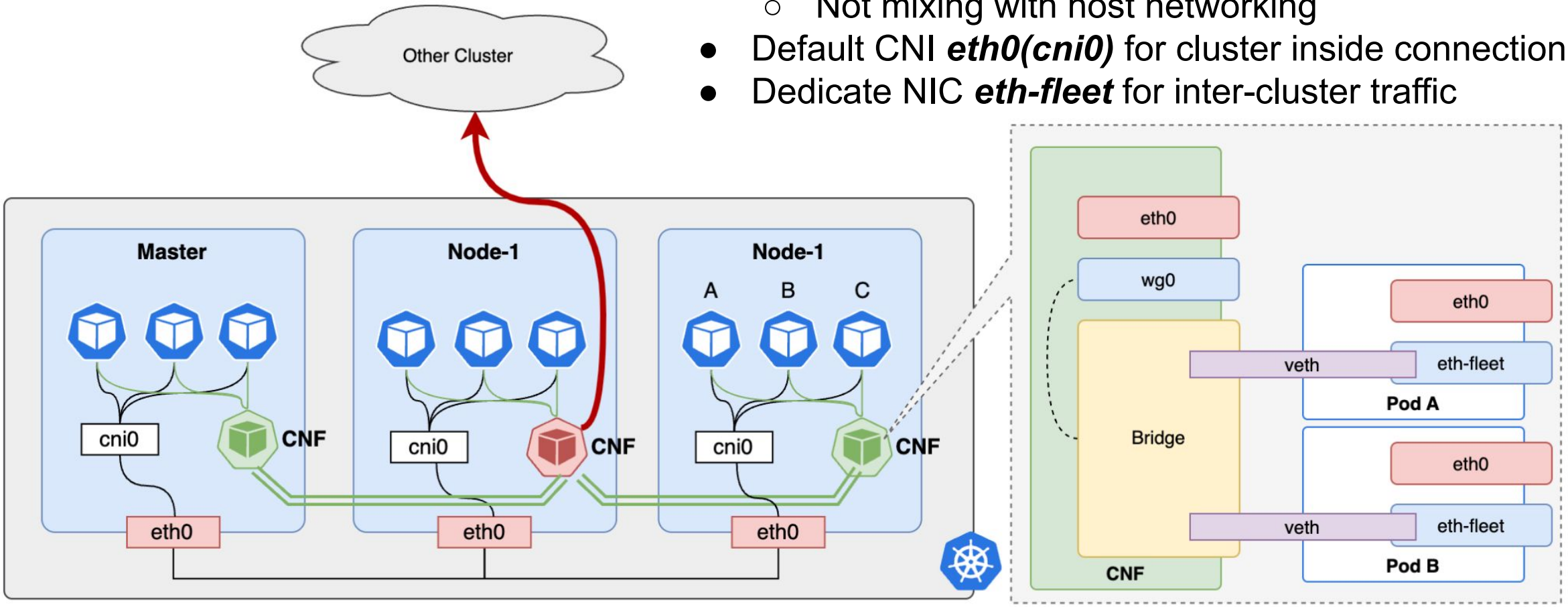
The Architecture

The Architecture of FleetBoard



Node Network

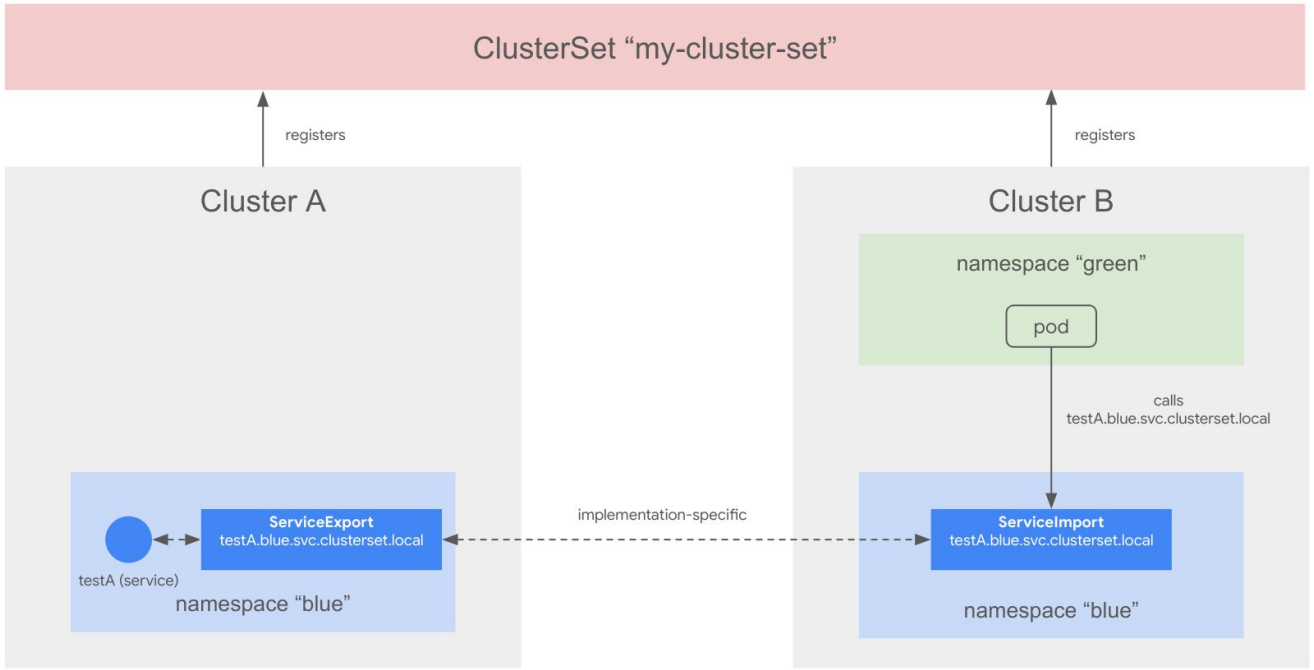
- Cloud-Native Network Function (CNF)
  - Not mixing with host networking
- Default CNI ***eth0(cni0)*** for cluster inside connection
- Dedicate NIC ***eth-fleet*** for inter-cluster traffic



Multi-Cluster Service Discovery

MCS: KEP-1645: Multi-Cluster Services API

Multi-Cluster EndpointSlices Syncing



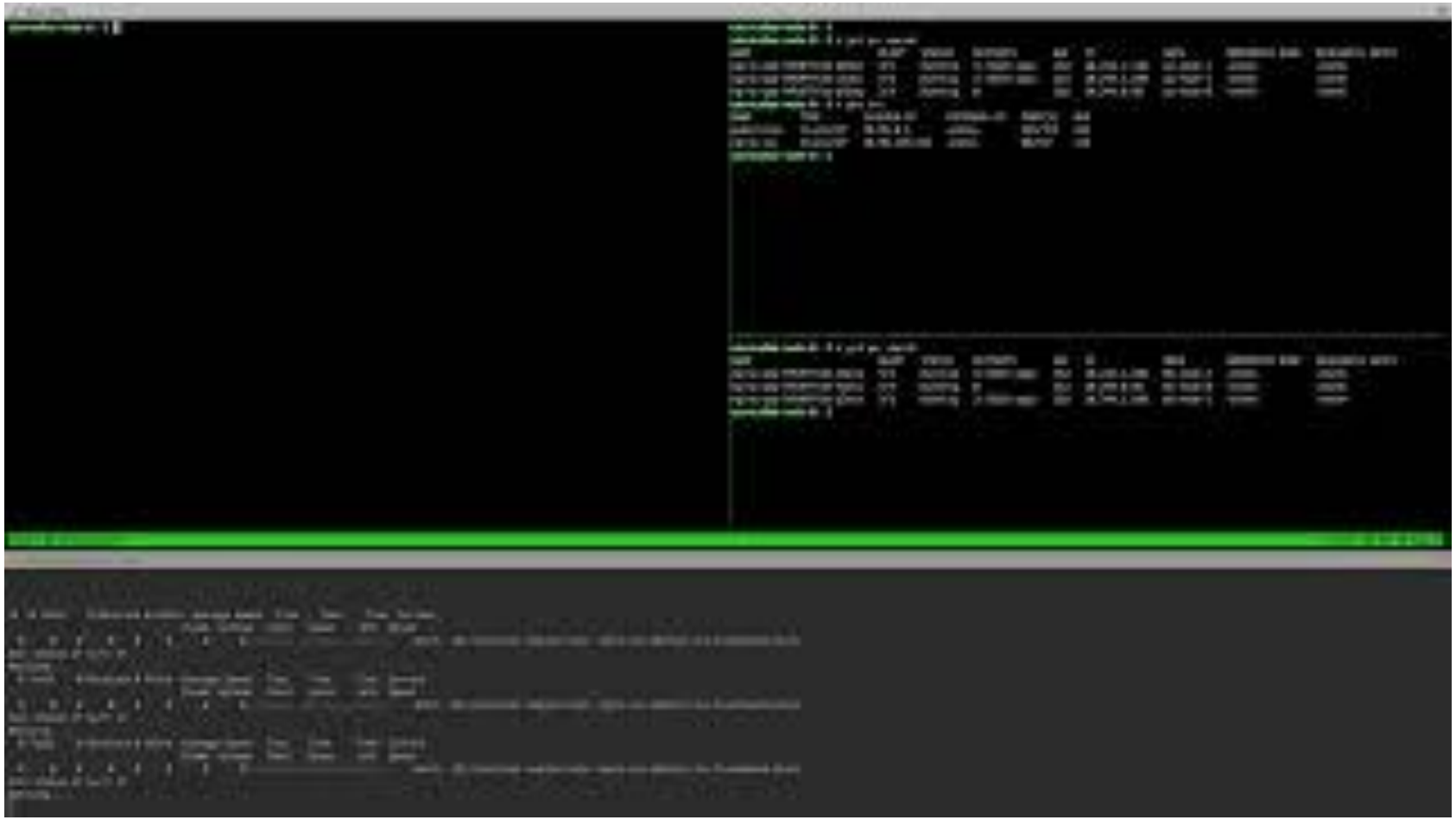
- Explicit Service Exposing
  - ServiceExport
  - ServiceImport
  - Multi-Cluster Network Policy (TBD)
- CoreDNS plugin
  - https://github.com/fleetboard-io/fleetboard/tree/main/cmd/crossdns
- EndpointSlices (eth-fleet IPs)

Advantages of FleetBoard

- Compatible with all CNI plugins
- Support Pod CIDR Overlap
- No cluster is required to provide an accessible IP
- Cross-Cluster Service Discovery
- Network Tunnel and Secure Access
- Non-invasive Network Configuration

Demo

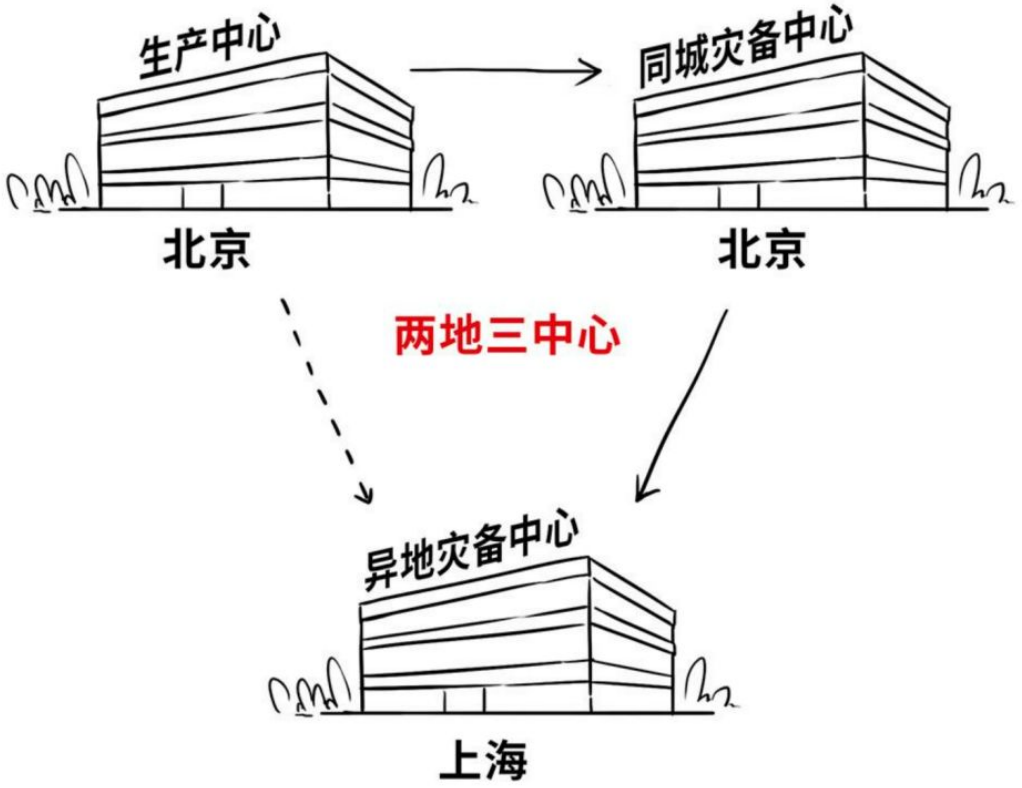
- Cluster “cc” as Hub cluster
- Cluster “bb” calls a service request to Cluster “aa”.



Some Use Cases

East-to-West Computing Resource Transfer Project

On-prem Cluster interconnection



Roadmap

- Performance
- Control Plane Load Balancing
- Integration with multi-cluster projects, such as Clusternet ✅ / OCM / Fleet / Karmada / etc.
- Multi-Cluster Network Policy
- Scalability
  - Upgrade without interruption

Thanks

- https://github.com/fleetboard-io/fleetboard

fleetboard demo