
KubeCon
CloudNativeCon
North America 2019

Alcor

Hyperscale Cloud Network Management

Futurewei Technologies

Introduction

Alcor is a cloud native platform that aims to provides high availability, high performance, and large scale virtual networking control plane and management plane at a high resource provisioning rate.

A	P	S	E
Availability	Performance	Scalability	Extensibility
• Always-on control plane without a single point of failure • Cross AZ resilience for services and data • Fault-tolerant design with multiple resource provisioning paths	• Throughput-optimal design to allow batched processing of network resources • Fast provisioning path to support time-critical applications such as serviceless	• Management of large numbers of network resources • Scale to half a million hosts and tens of millions network ports	• Unified network management of both VMs and containers • Plug-able model to support various implementations of data plane

Architecture Overview

Management plane

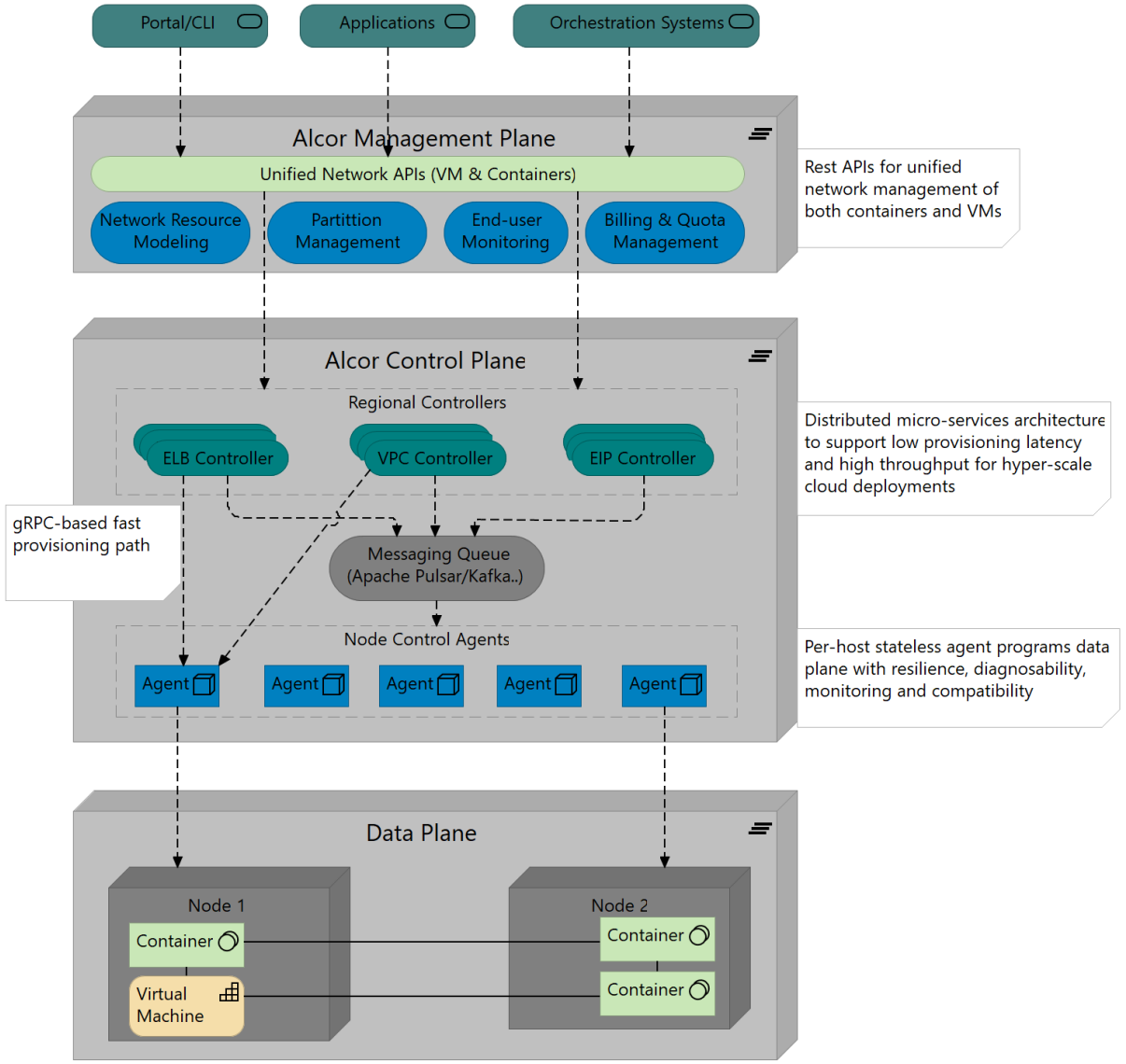
- Expose REST APIs to clients
- Provide partition management, end-user monitoring, and billing & quota management

Control Plane

- Offer multiple network management services including VPC, ELB and EIP
- Drive network configurations in scale to on-host network agents
- Support multiple provisioning path including fast path, normal path and rescue path

Data Plane

- Provide network connection between containers and VMs in the same VPC



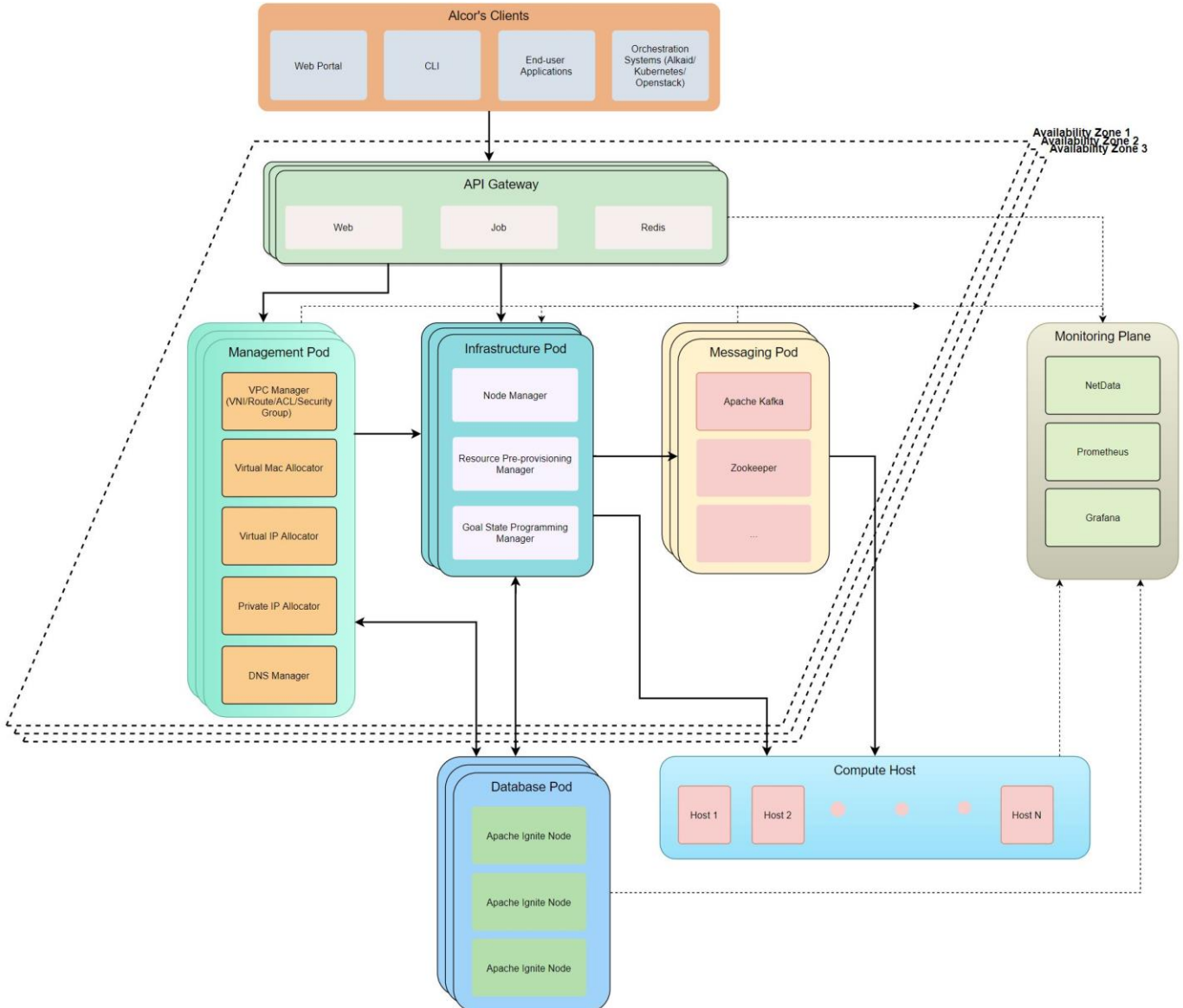
Network Controller

Design Principals

- Regional Scope
- Simple network resource abstraction
- Loosely coupled components for flexible partitioning and easy scale out
- Top-down configuration driving from controller to control agents

Distributed Micro-Services Architecture

- A uniform way to secure, connect, and monitor control plane micro-services powered by Istio
- Offer fine-grained control of service-to-service communication including load balancing, retries, failovers, and rate limits.



AZ-Resilience Design

AZ resilience

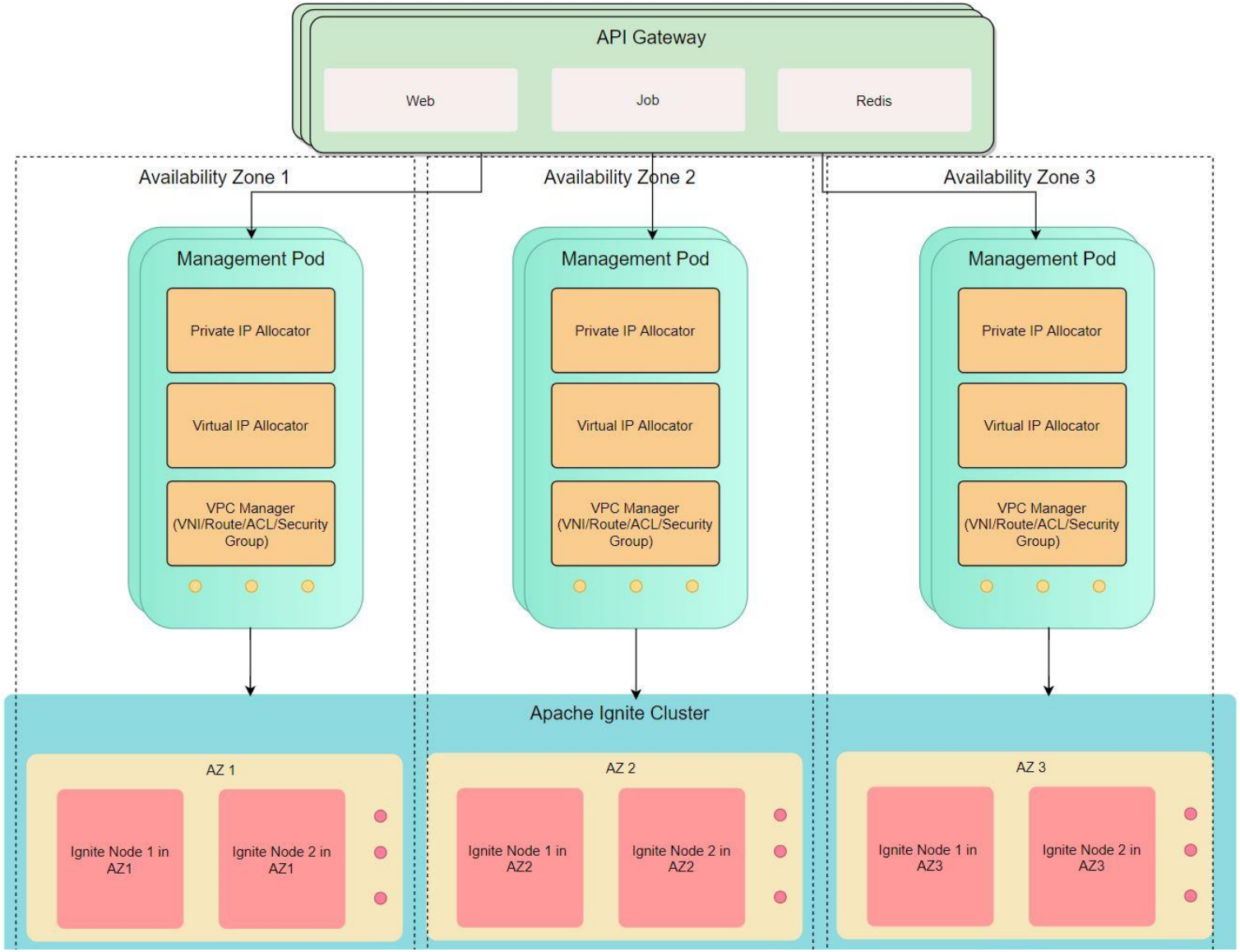
- Controller services and database/cache clusters are deployed across Availability zones
- If any AZ goes down, control plane remains up and fully capable of assuming its load and data

Geo-replication

- Enable Ignite's data center replication
- Supports both active-active and active-passive modes for replication

Fast read operations

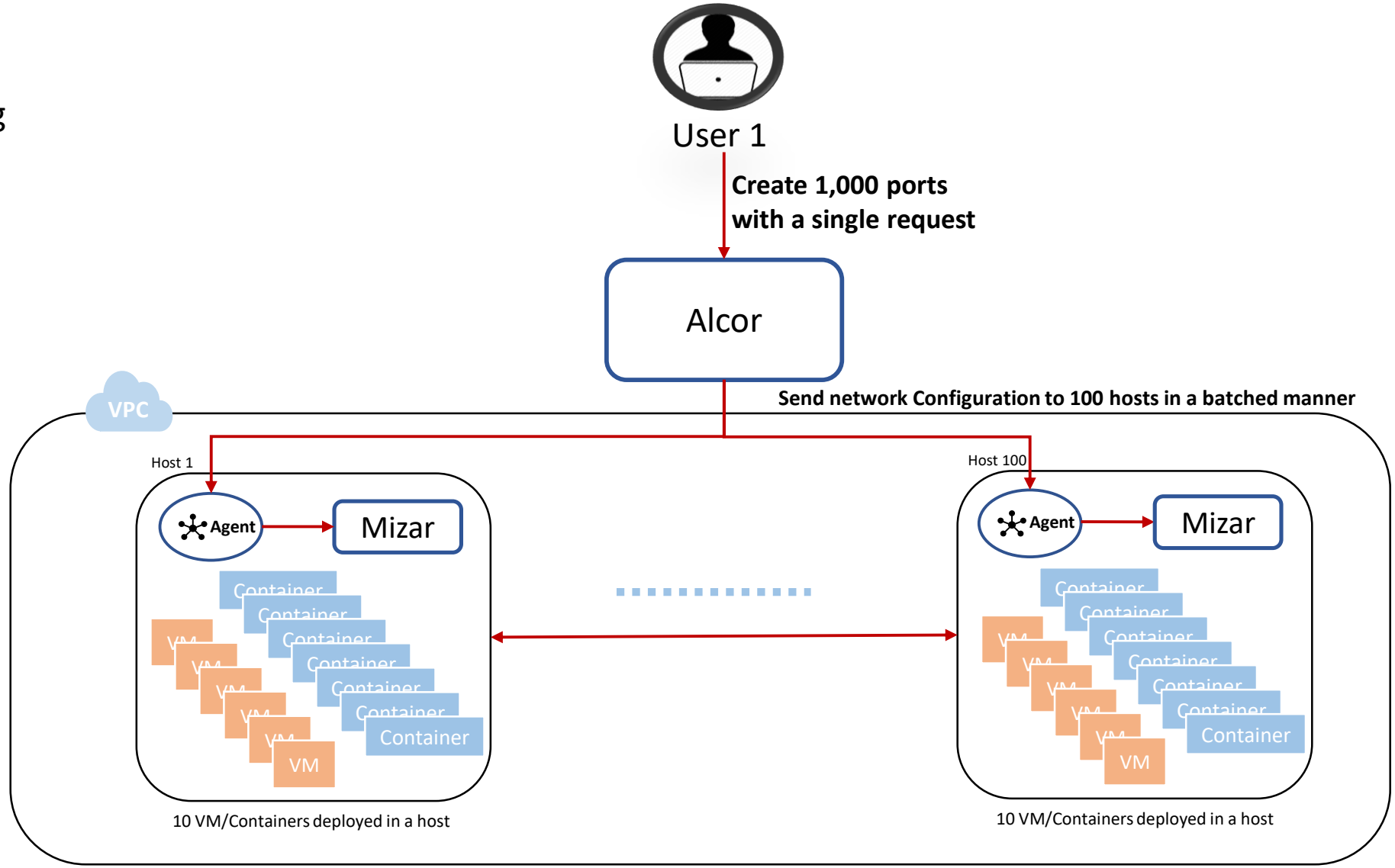
- Ultra-low latency due to data locality and in-memory cache capability.



Demo:Large-Scale VPC Provisioning

Throughput-optimal design

a. Enable batched processing of network ports
   i. Boost overall system throughput
   ii. Support large VPC deployment
b. Optimized on every system layer
   i. API
   ii. Controller
   iii. Controller-to-agent communication
   iv. Host
c. Provision 1,000 ports within XX seconds



Video

Content

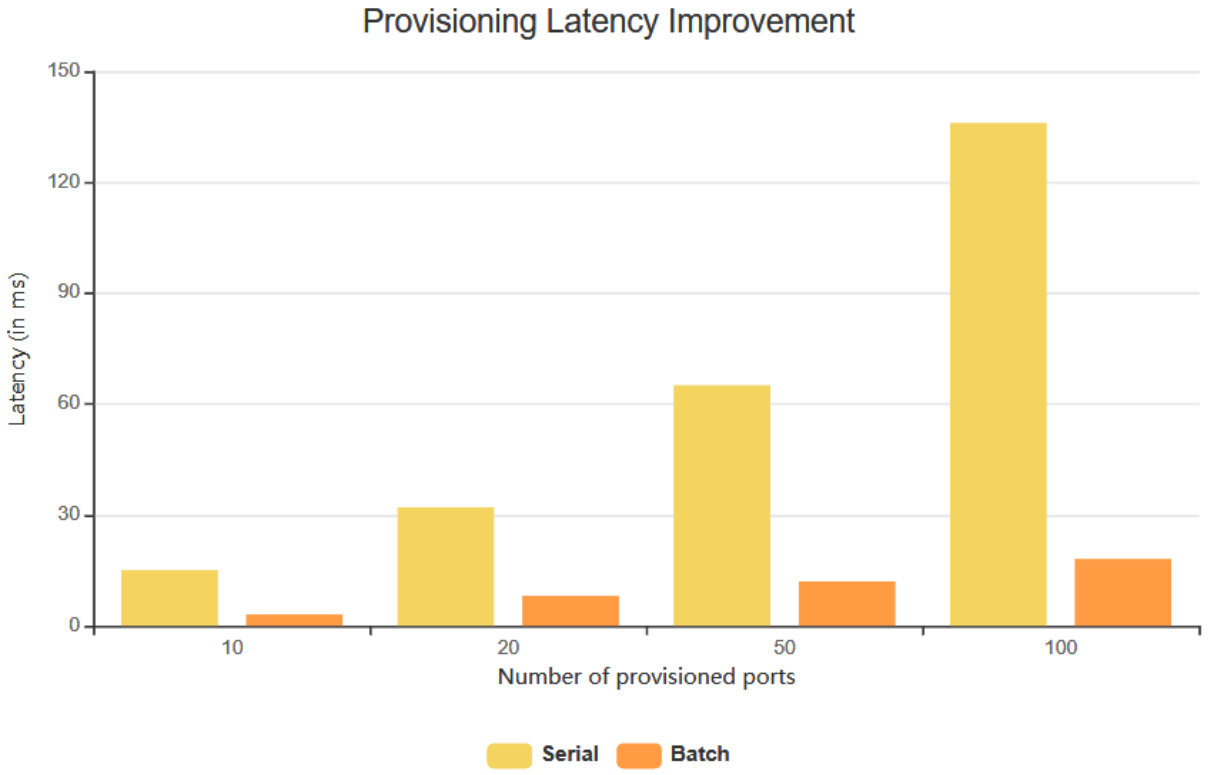
Provisioning Latency

Serial Provisioning:

- Ports are created and distributed one by one.
- Unable to scale to a large number of port creation.

Batched Provisioning:

- Created with a single API call (portgroup creation)
- Distributed to hosts in a batched manner.



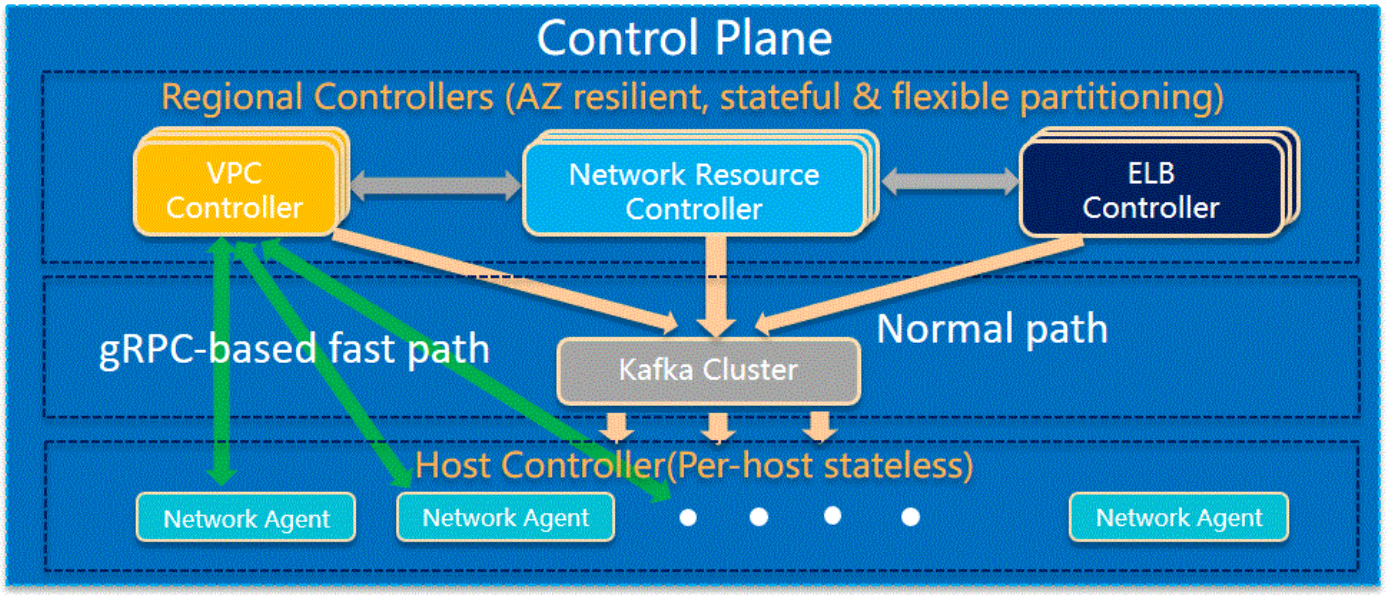
Fast Port Provisioning

Fast path design

- Direct communication channel from controller to agent
- Alternative provisioning path for control plane reliability

Background

- Customer scenarios requires ultra-low latency for E2E network configuration provisioning (in a few 10 *ms* or 100 *ms*)
- Message queue subsystem, usually adopted as a high-throughput and scalable solution for network configuration updates, may not fit into time-critical customer scenarios



Video

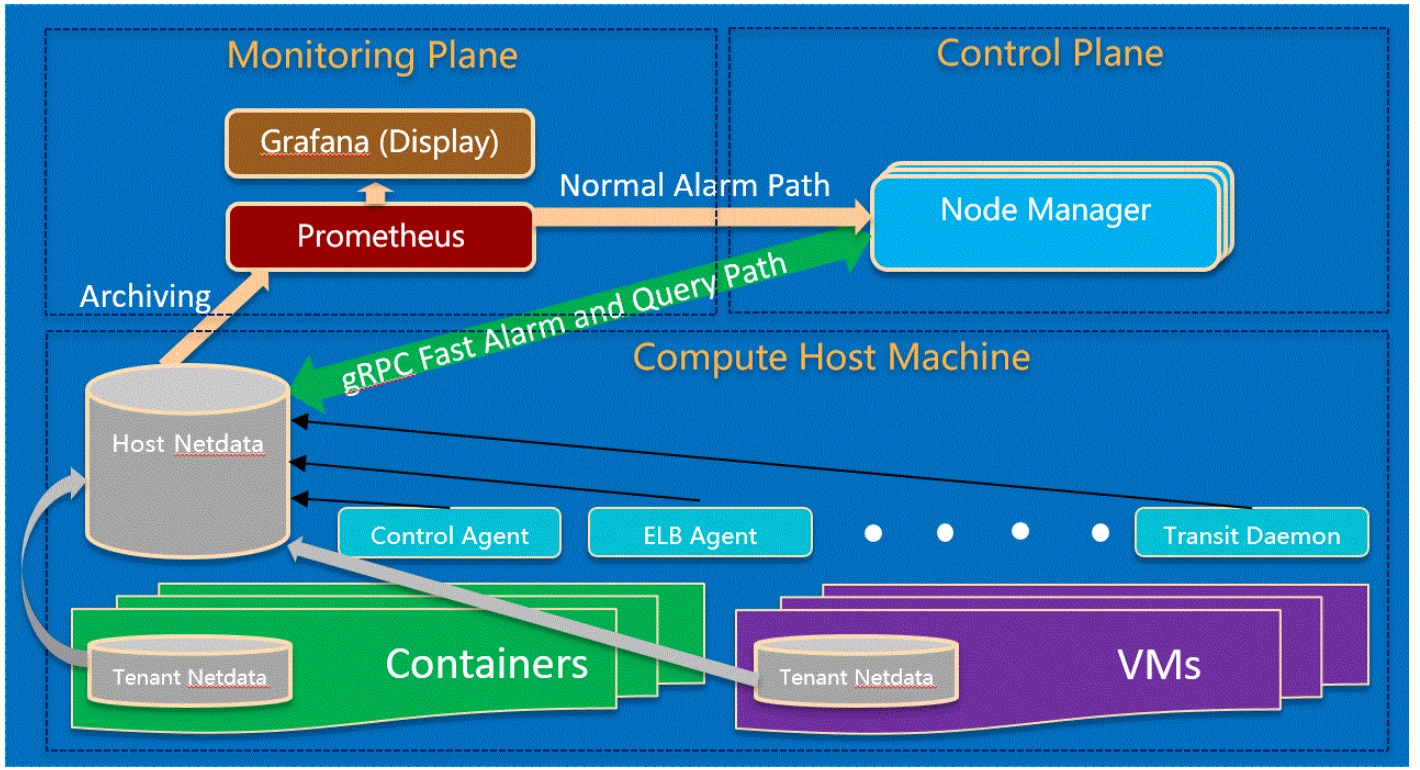
Content

Monitoring Design

Node Monitoring

- Leverage Host Netdata, a decentralized data collection and processing engine running on each compute host.

- Responsibility
  - Monitor host-level agents including ACA and ELB agent
  - Collect data-plane health metrics
  - Monitor host node networking health (e.g. TCP/UDP/ICMP packets drop rate, connection counter, throughput etc.)
  - VM/Container network monitoring (requiring customer opt-in)



Monitoring Demo

Content

Rescue Path

Agent Outage Handling

During Alcor Network Agent (ACA) outage, a neighbor ACA can still setup Endpoint and program dataplane.

ACA were designed to have high throughput, low latency when running in local host, it can also invoke tasks on remote host with minimal added cost.

When rescue path is used, all existing deployment should continue to work without impact. The compute node may be marked as unhealthy. New VMs/Containers or endpoints can still be deployment to this node without customer visible degradation.