Arktos: Current Work and Next

Xiaoning Ding

Agenda

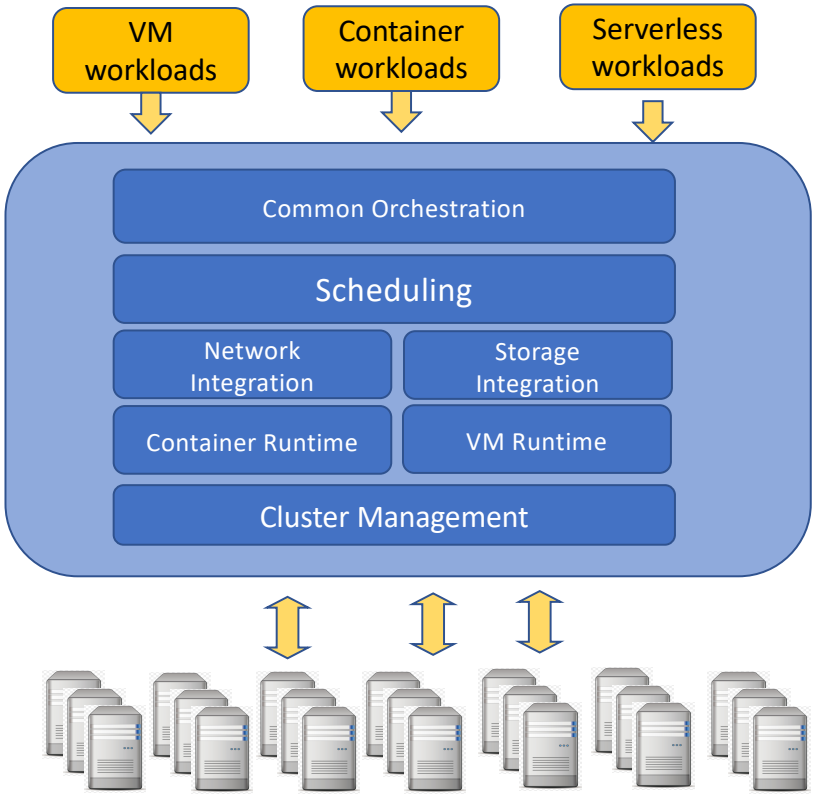
- Background
- Arktos
  - Features Overview
  - Current work
  - Next

CloudFabric Overview

1. Large-scale Cloud Platform: challenges and opportunities
   - Challenges : scalability; resource utilization, new workload types (containers & functions); new heterogenous hardware & high-speed network, etc.
   - Opportunities: new technologies in hardware and software, including hardware offloading, CPU QoS control, new networking capabilities in Linux kernel, etc.

2. CloudFabric: an open-source project group for large scale cloud platform
   - Arktos, computing platform, https://github.com/futurewei-cloud/arktos
   - Mizar, network data plane, https://github.com/futurewei-cloud/mizar
   - Alcor, network control plane, https://github.com/futurewei-cloud/alcor

Arktos: Overview



Unified Stack

- VM and containers share a same control plane.
- VM and containers share a shared resource pool.
- Built-in optimization for serverless functions.

Large Scalability

- Targeting at 100K nodes per cluster and 300K nodes per region
- Make all components partitioned and replicated

Multi-tenancy

- Enable multi organizations and teams to share a same resource pool.
- Provide strong isolation and quota mechanism for resources.

Resource Utilization

- Vertical scaling and SLO-driven scaling.
- intelligent distributed scheduling.
- GPU scheduling.

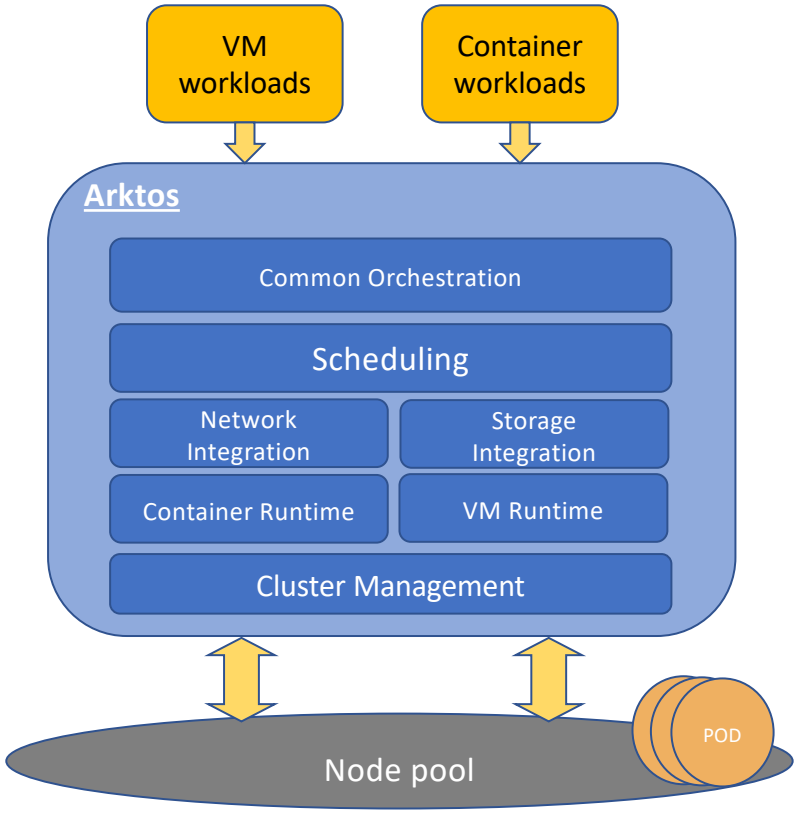
Virtualization

- ARM virtualization & optimization.
- lightweight virtualization.

Edge Cloud

- Cloud/Edge Scheduling
- Next version of KubeEdge

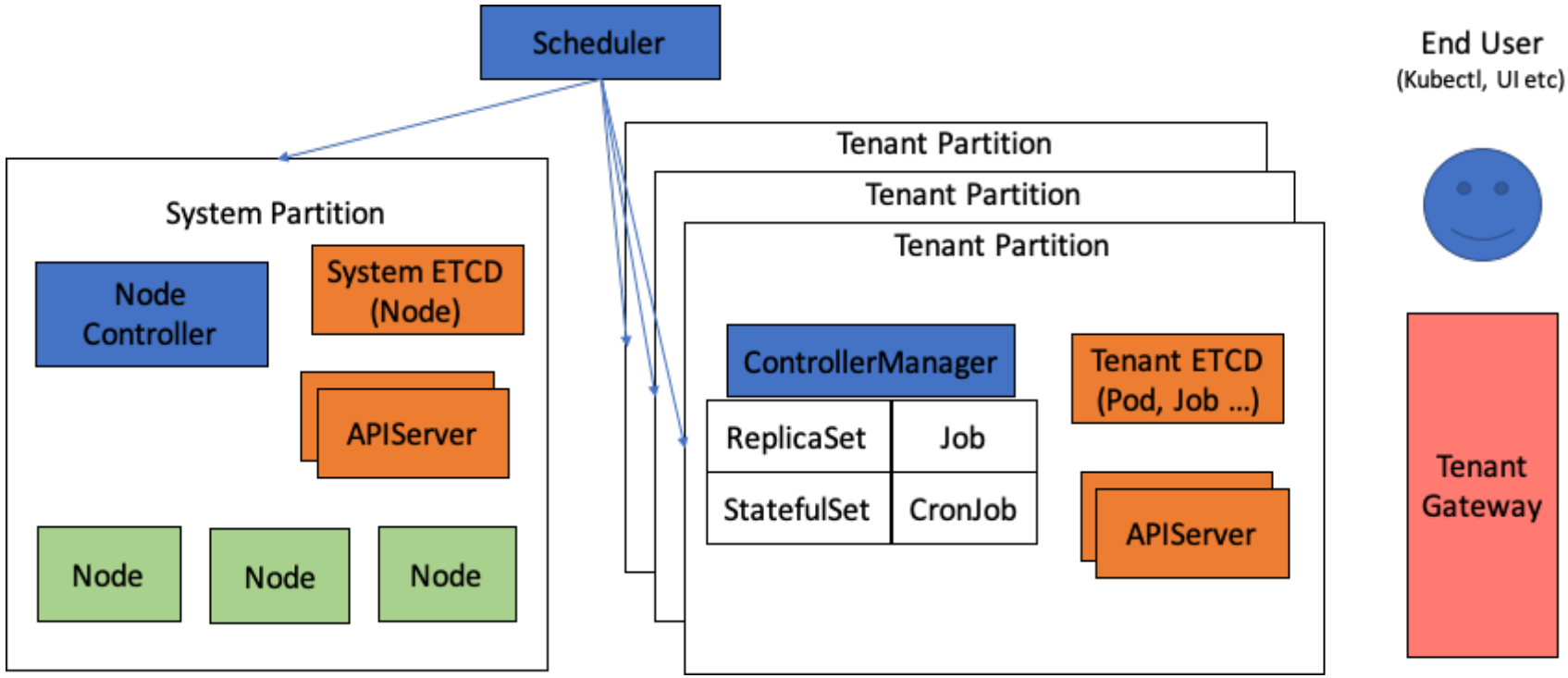
Arktos: Unified Stack




```yaml
apiVersion: v1
kind: Pod
metadata:
  name: vm1
spec:
  virtualMachine:
    name: vm
    image: download.cirros-cloud.net/0.3.5/cirros-0.3.5-x86_64-disk.img
    resources:
        requests:
          cpu: "1"
          memory: "1Gi"
```

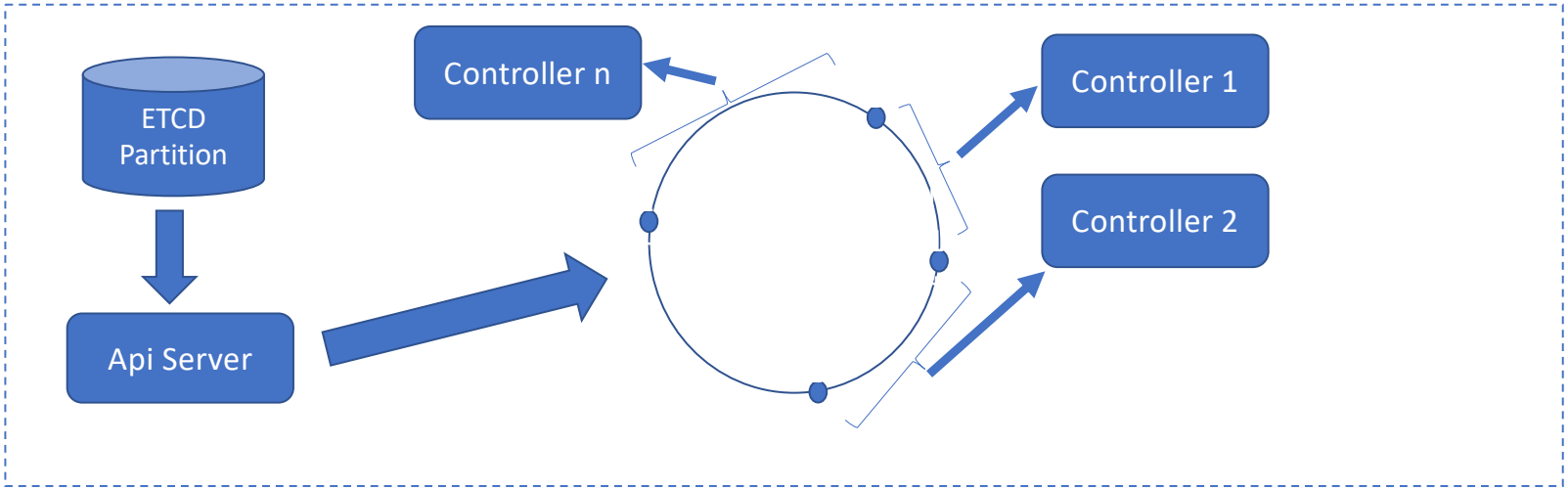
```yaml
apiVersion: v1
kind: Pod
metadata:
  name: container1
spec:
  containers:
    - name: container1
      image: ubuntu
      command: ["/bin/bash", "-ec", "while :; do echo '.'; sleep 5 ; done"]
      resources:
        requests:
          cpu: "1"
          memory: "1Gi"
```

Arktos: Scalability



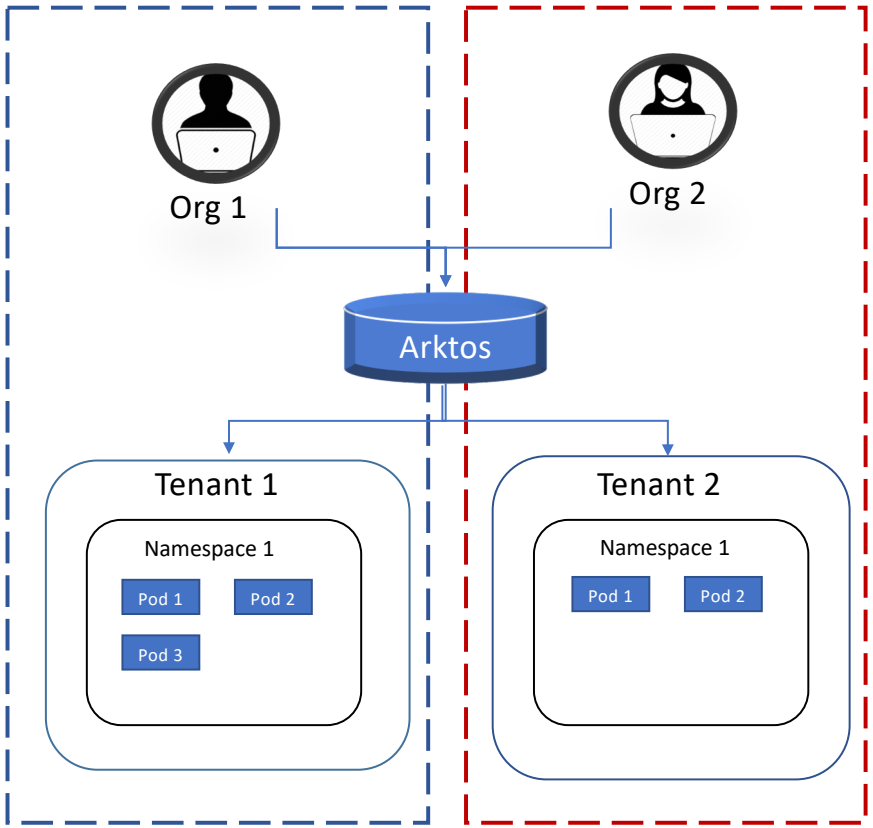
- Shard tenant data
- Scheduler has global view of all nodes in cluster

Arktos: Scalability



- List/watch by range of field value
- Multiple controller instances
  - Multiple controller managers works in active-active mode

Arktos: Multi-Tenancy



Hard Multi-Tenancy

- Enable organizations/departments to safely share one infrastructure, without deploying/operating multiple clusters.
- Support per-tenant resource view, access control, quota, etc.
- Assume no trust among tenants; ready for strict scenarios like public cloud.

Key Changes:

- A new API object: ***tenant***
- All API objects have a new field ***Tenant*** in its ***ObjectMeta*** section
- A new resource URL scheme: ***tenants/{tenant}/namespaces/{namespace}/{objectTypes}/{objectName}***
- Tenant-aware Client-Go library, scheduler, controllers, agent and CLI tools.

Next Step: Scheduling, Resource Utilization & SLO

- Cross-site scheduling
  - A global entry to schedule workloads across cloud data centers and Edge sites
  - Geo locality aware and application traffic-pattern aware

- Resource Utilization Improvements
  - In-place vertical scaling. Applications don’t need to reserve the peak resource requirement up front.
  - SLO-driven resource management, including SLO specification, monitoring and intelligent resource adjustment to enforce the SLO requirements.

Next Step: ARM Optimization

- The eco system of ARM in data center is emerging. AWS released its Graviton cpu and Huawei released Kunpeng cpu.
- ARM CPU is usually more energy efficient and cost efficient than x86, but per-core performance is less than x86.
- Besides simply recompiling software for ARM platform, what software optimizations can we do? What new features can we build on ARM?
- Some initial efforts
  - Work with university researchers to implement enclave in normal world of ARM TrustZone. This potentially can provide a large enclave for sensitive cloud applications.

Thank you.