KubeCon

CloudNativeCon

THE LINUX FOUNDATION

OPEN SOURCE SUMMIT

AI_dev

Open Source GenAI & ML Summit

China 2024

KubeCon | CloudNativeCon

THE LINUX FOUNDATION OPEN SOURCE SUMMIT

AI_dev Open Source GenAI & ML Summit

China 2024

连接点：走向统一的多集群AI/ML体验 Connecting the Dots: Towards a Unified Multi-Cluster AI/ML Experience

Qing Hao, Red Hat & Chen Yu, Microsoft

K8s多集群：快速历史回顾

K8s multi-cluster: a quick recap

KubeFed V1 → KubeFed V2 →

OCM

Fleet

Clusternet

Karmada

多集群管理平台

Multi-Cluster Management Platform

Kueue

Volcano

Argo

支持多集群的应用

Applications with Multi-Cluster Support

社区：问题与挑战
Community: problems and challenges

我们该如何为开发者提供统一、标准化的多集群管理体验，同时又能最大程度的保留各个多集群解决方案在多集群管理问题上各自的独特视角？

How do we provide developers with a unified and standardized multi-cluster management experience, while preserving the unique perspectives each multi-cluster management platform holds on the subject matter?

社区：问题与挑战

Community: problems and challenges

我们该如何将多集群管理平台同需要多集群能力的应用连接起来，为应用开发者减轻支持多集群的负担，同时帮助用户复用现有的多集群能力？

How do we connect applications that need multi-cluster support with platforms with multi-cluster management capabilities, so that developers no longer need to rebuild the wheel, and end users can leverage the multi-cluster management capabilities they are already familiar with?

社区：问题与挑战
Community: problems and challenges

Good Engineers
Borrow,
Great Engineers
Steal.

良工摹其形
巧匠窃其意

Cluster Inventory API

```yaml
apiVersion: multicluster.x-k8s.io/v1alpha1
kind: ClusterProfile
metadata:
  name: bravelion
  namespace: ...
spec:
  clusterManager:
    name: ...
  displayName: bravelion
status:
  conditions:
  - lastTransitionTime: ...
    message: The control plane is of a healthy state
    observedGeneration: 1
    reason: HealthCheckCompleted
    status: True
    type: ControlPlaneHealthy
  ...
```

KEP 4322

面向多集群的统一接口

Unified interface for multi-cluster management

连点成线
Connecting the dots

状态监控
Status Monitoring

✅
KEP 4322

凭证签发
Credential Signing

🎯
讨论中
In active discussion

多集群应用的自动配置
Auto-configuration for multi-cluster application

✨
社区愿景
Community vision

凭证的自动生成

Auto-preparation of credentials

AuthTokenRequest API

暂定名称

早期讨论中

```
apiVersion: multicluster.x-k8s.io/v1alpha1
kind: AuthTokenRequest
metadata:
  name: work
  namespace: default
spec:
  targetClusterProfile:
    apiGroup: multicluster.x-k8s.io/v1alpha1
    kind: ClusterProfile
    name: bravelion
  serviceAccountName: work
  serviceAccountNamespace: default
  clusterRoles:
  - name: cluster-admin
```

自动配置之一
Auto-configuration: Pt 1

Fleet

由🪟开源的多集群管理平台，为多集群部署提供状态监控、资源调度和分发以及网络能力。

Multi-Cluster management platform open-sourced by Microsoft, with status monitoring, resource scheduling + placement, and networking capabilities.

🔗github.com/azure/fleet

展示：使用Cluster Profile和AuthTokenRequest API，自动为成员集群签发凭证。

Demo: sign a credential for a member cluster using Cluster Profile and AuthTokenRequest API.

🔗https://asciinema.org/a/EAI5mowdlhxygBBKHUrzTKS4A

自动配置之二
Auto-configuration: Pt 2

Kueue

由Kubernetes社区开源的Kubernetes原生作业队列控制器，提供对多集群环境的支持（MultiKueue）。

Kubernetes-native Job queue controller open-sourced by the Kubernetes community, with support for multi-cluster environment.

🔗github.com/kubernetes-sigs/kueue/

自动配置之三
Auto-configuration: Pt 3

查找目标集群
Find target cluster

创建服务账户
Find target cluster

配置权限
Set up RBAC

签发令牌
Issue service account token

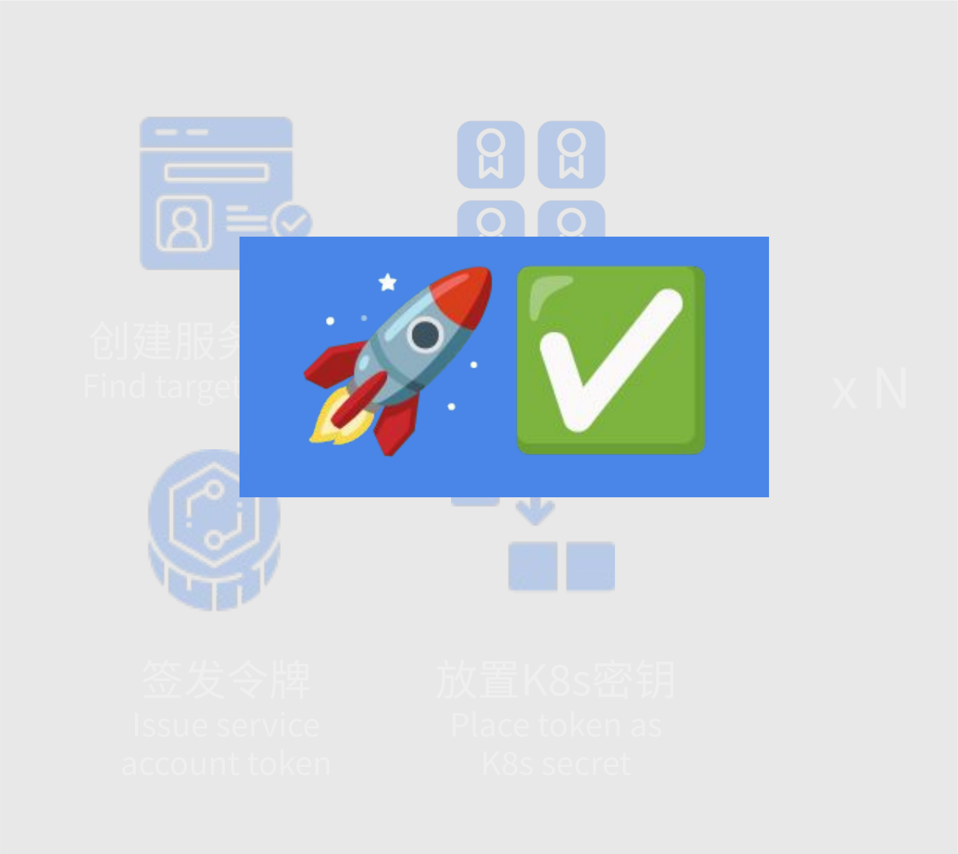
放置K8s密钥
Place token as K8s secret

x N

配置应用
Configure application

自动配置之四
Auto-configuration: Pt 4

查找目标集群
Find target cluster



配置应用
Configure application

自动配置之五

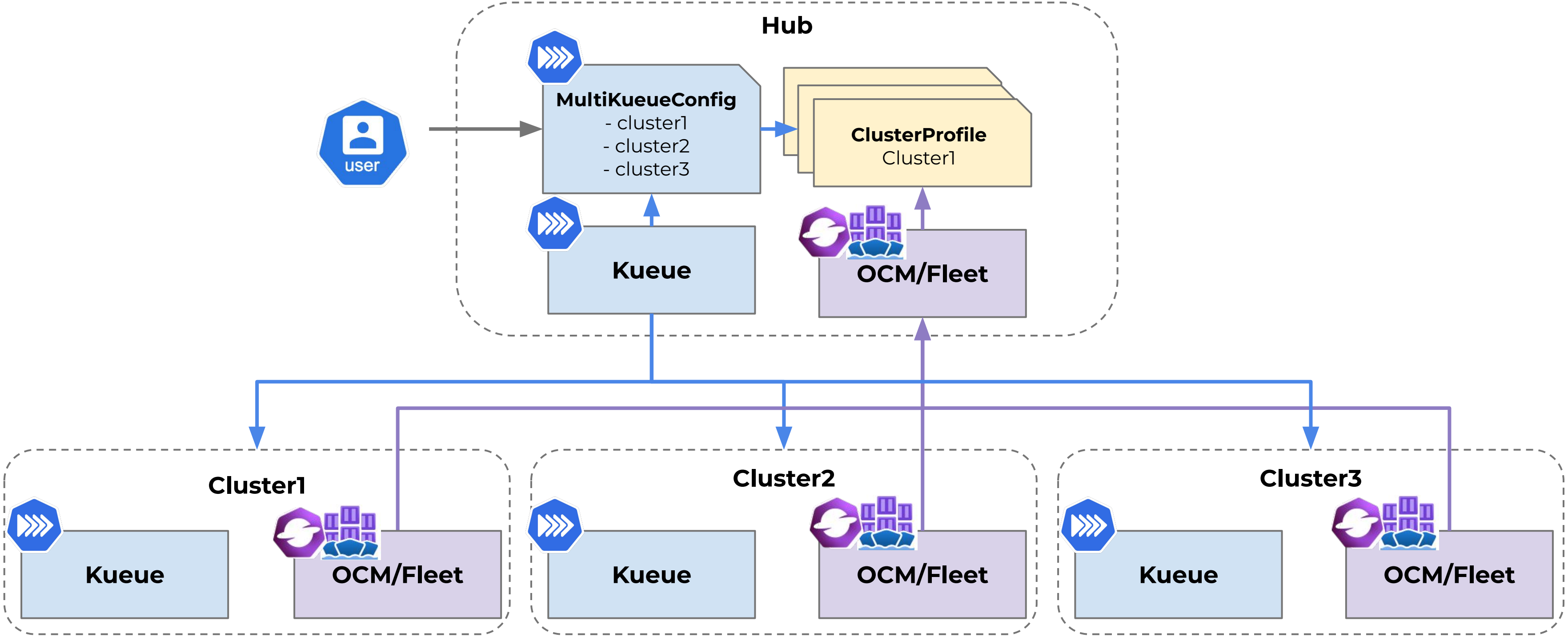
Auto-configuration: Pt 5

展示：使用Cluster Profile和AuthTokenRequest API，Kueue可以自动配置用于分发离线工作负载（数据分析、AI/ML训练等）的MultiKueue多集群环境。

Demo: auto-configuring Kueue to set up a MultiKueue environment for orchestrating offline workloads like data analysis or AI/ML training jobs, with the help of ClusterProfile and AuthTokenRequest API.

🔗https://asciinema.org/a/fJx5yqEajN51wrkc4zBlmOHjl

自动配置之六
Auto-configuration: Pt 6



自动配置之外
Beyond auto-configuration

集群替换
Cluster Replacement

凭证轮转
Credential Rotation

多集群调度能力

Multi-cluster Scheduling Capabilities

Open Cluster Management

OCM旨在解决多集群场景下的集群注册管理，工作负载分发，以及动态的资源配置等功能，目前属于CNCF sandbox阶段。

OCM is focused on multicluster and multicloud scenarios for cluster registration, work distribution, dynamic placement of policies and workloads, and much more. It is a CNCF sandbox project.

🔗open-cluster-management.io

多集群调度能力
Multi-cluster Scheduling Capabilities

预选
Predict

Hard requirements
- Label Selector
- Taint/Toleration

优选
Prioritize

Soft requirements
- Number of clusters
- Score based ranking
- Customized score

生成调度决策
Generate scheduling decision

Decisions can keep steady or update dynamically.

连点成线
Connecting the dots

通过支持ClusterProfile API，更容易复用现有的多集群调度能力到Kueue这样支持多集群的应用，实现多集群环境的动态配置，满足更多的调度场景。

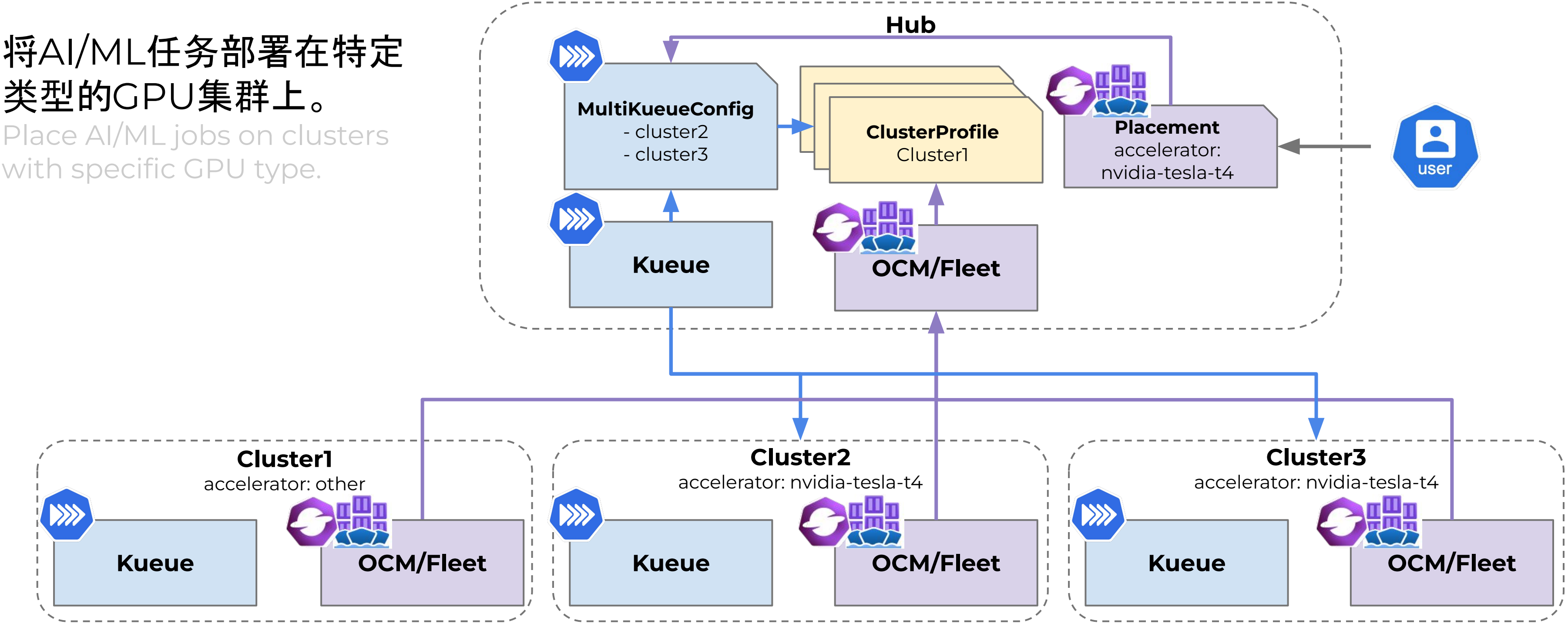
By supporting the ClusterProfile API, it becomes easier to reuse the existing multi-cluster scheduling capability in applications like Kueue that support multi-cluster environments, enabling dynamic configuration and meeting more scheduling scenarios.

复用多集群调度能力之一

Leverage the Multi-cluster Scheduling Capabilities: Pt 1

将AI/ML任务部署在特定类型的GPU集群上。

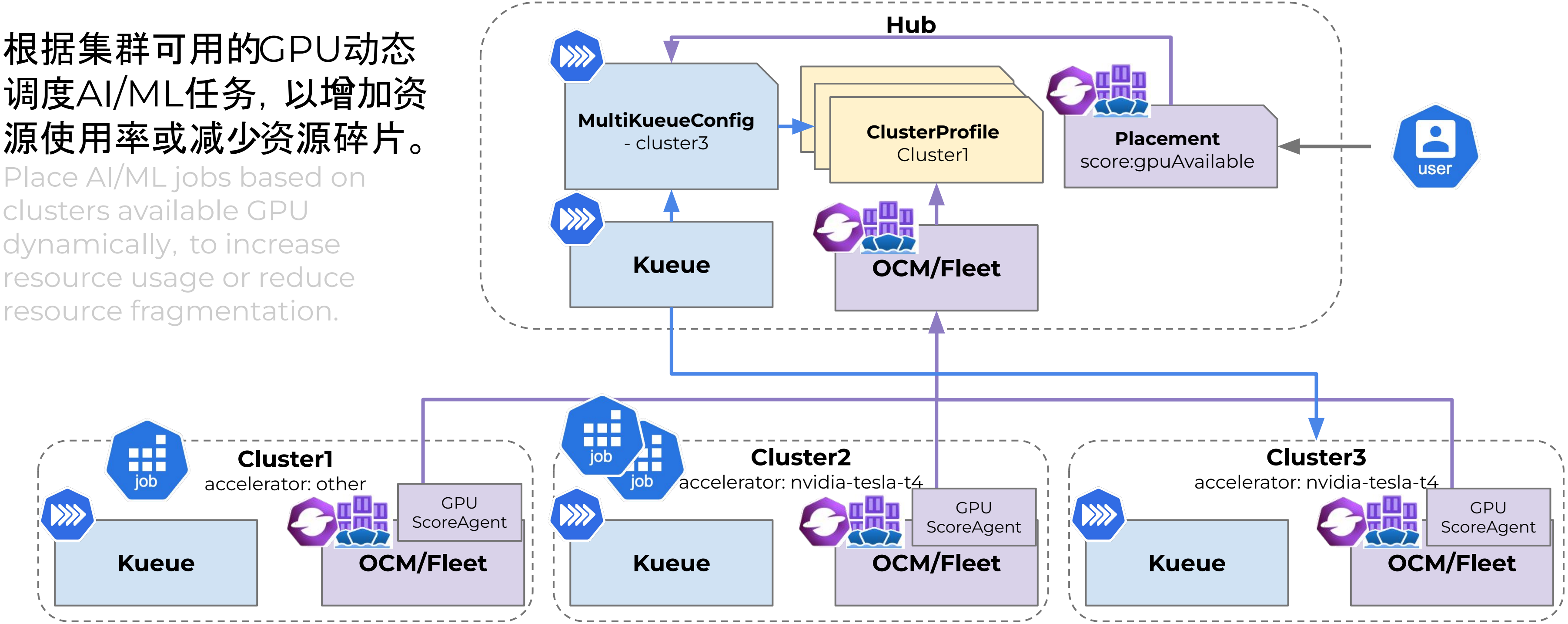
Place AI/ML jobs on clusters with specific GPU type.



复用多集群调度能力之二
Leverage the Multi-cluster Scheduling Capabilities: Pt 2

根据集群可用的GPU动态调度AI/ML任务，以增加资源使用率或减少资源碎片。

Place AI/ML jobs based on clusters available GPU dynamically, to increase resource usage or reduce resource fragmentation.



复用多集群调度能力之三

Leverage the Multi-cluster Scheduling Capabilities: Pt 3

展示一：将Kueue工作负载（数据分析、AI/ML训练等）部署在类型为nvidia-tesla-t4的GPU集群上。

Demo1: Deploy Kueue workloads (data analysis or AI/ML training jobs) on clusters with GPU type nvidia-tesla-t4.

展示二：将Kueue工作负载（数据分析、AI/ML训练等）动态部署至可用GPU多的集群上。

Demo2: Deploy Kueue workloads (data analysis or AI/ML training jobs) on clusters with most available GPU dynamically.

🔗https://asciinema.org/~h222q

下一步
Next

Learn more about Cluster Inventory API
https://github.com/kubernetes-sigs/cluster-inventory-api

Adopting a Standardized ClusterInventory API from SIG Multi-Cluster
https://github.com/open-cluster-management-io/ocm/issues/247

OCM GPU/TPU-resource-usage-collect-addon
https://github.com/open-cluster-management-io/addon-contrib/pull/20

OCM Kueue Admission Check Controller
https://github.com/open-cluster-management-io/ocm/pull/601

加入我们

Get Involved

SIG Multicluster

https://multicluster.sigs.k8s.io/

Fleet

https://github.com/azure/fleet/

Open Cluster Management

https://open-cluster-management.io/

KubeCon | CloudNativeCon

THE LINUX FOUNDATION OPEN SOURCE SUMMIT

AI_dev Open Source GenAI & ML Summit

China 2024

Thanks!