Optimizing Resource Provisioning Across Diverse Computing Facilities with Virtual Kubelet Integration

Authors: Jeng-Yuan Tsai, Christopher Larrieu, David Lawrence, Graham Heyes, Vardan Gyurjyan

Motivation

In the dynamic landscape of High-Performance Computing (HPC), managing distributed compute resources across disparate sites presents a formidable challenge. JIRIAF aims to address this specific issue by proposing innovative solutions that operate within user space, ensuring seamless resource management without compromising system integrity.

Key Solution

JIRIAF: An Elastic Kubernetes (K8s) Cluster
* JIRIAF employs Virtual-Kubelet as a resource manager, strategically deployed in user space across different compute sites.
* The K8s framework ensures efficient control and monitoring of workloads.
* JIRIAF optimizes its operations by utilizing readily available on-the-market resources for Kubernetes.

Proof of Concept

* CLAS12 experiment event streamed to the NERSC computing facility via the EJFAT transport system.
* JRMs were deployed on 8 nodes within the NERSC cluster, awaiting the deployment of stream processing workflows.
* Deployed the stream processing workflow ERSAP for CLAS12 reconstruction.
* Monitored the resources (CPU and memory utilization) of nodes and workflow by k8s metrics server and Prometheus server.

Future Work

* Implement Horizontal Pod Autoscaling (HPA): Optimize resource utilization by enabling HPA for flow-based workflows.
* Leverage AI-Based Tools: Detect and utilize opportunistic resources using AI-driven approaches.
* Develop a digital twin framework. Construct a model that forecasts user workflow within a specific High-Performance Computing (HPC) setting.

JLAB Integrated Research Infrastructure Across Facilities
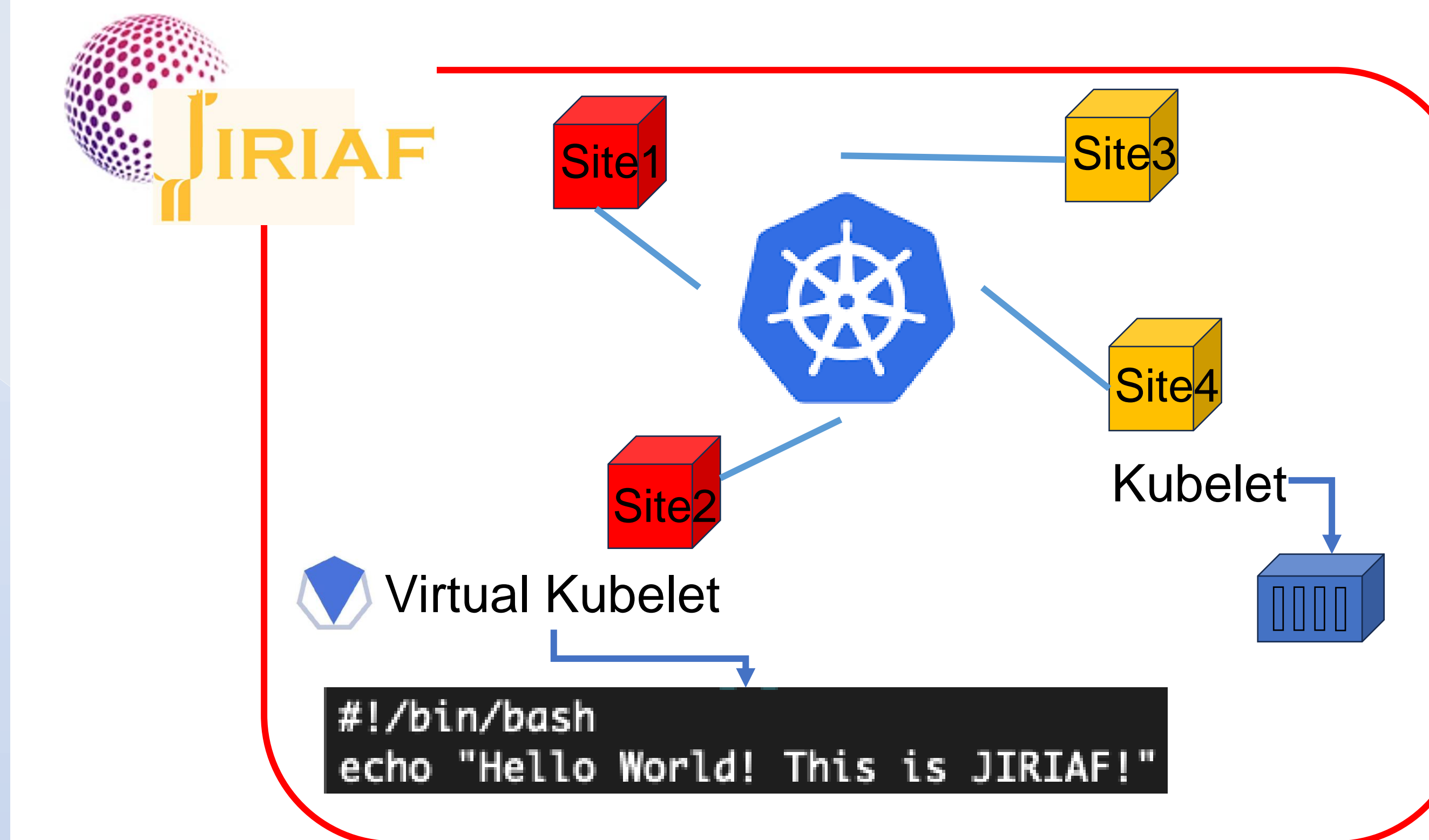

Challenge

* Coordinating and overseeing a large-scale, distributed infrastructure is challenging.
* Efficiently migrating and scaling workloads across distributed compute sites is crucial but complex for optimal resource utilization.
* Intelligently harnessing opportunistic resources in HPC environments is essential for enhancing efficiency.

Core Component

JIRIAF resource manager (JRM)
* JRM translates regular Kubernetes (K8s) objects into BASH commands specifically tailored for JIRIAF.
* JRM leverages the metrics server within K8s to monitor pods and nodes effectively.
* JIRIAF benefits from JRM's ability to manage the lifecycle of pods and containers.
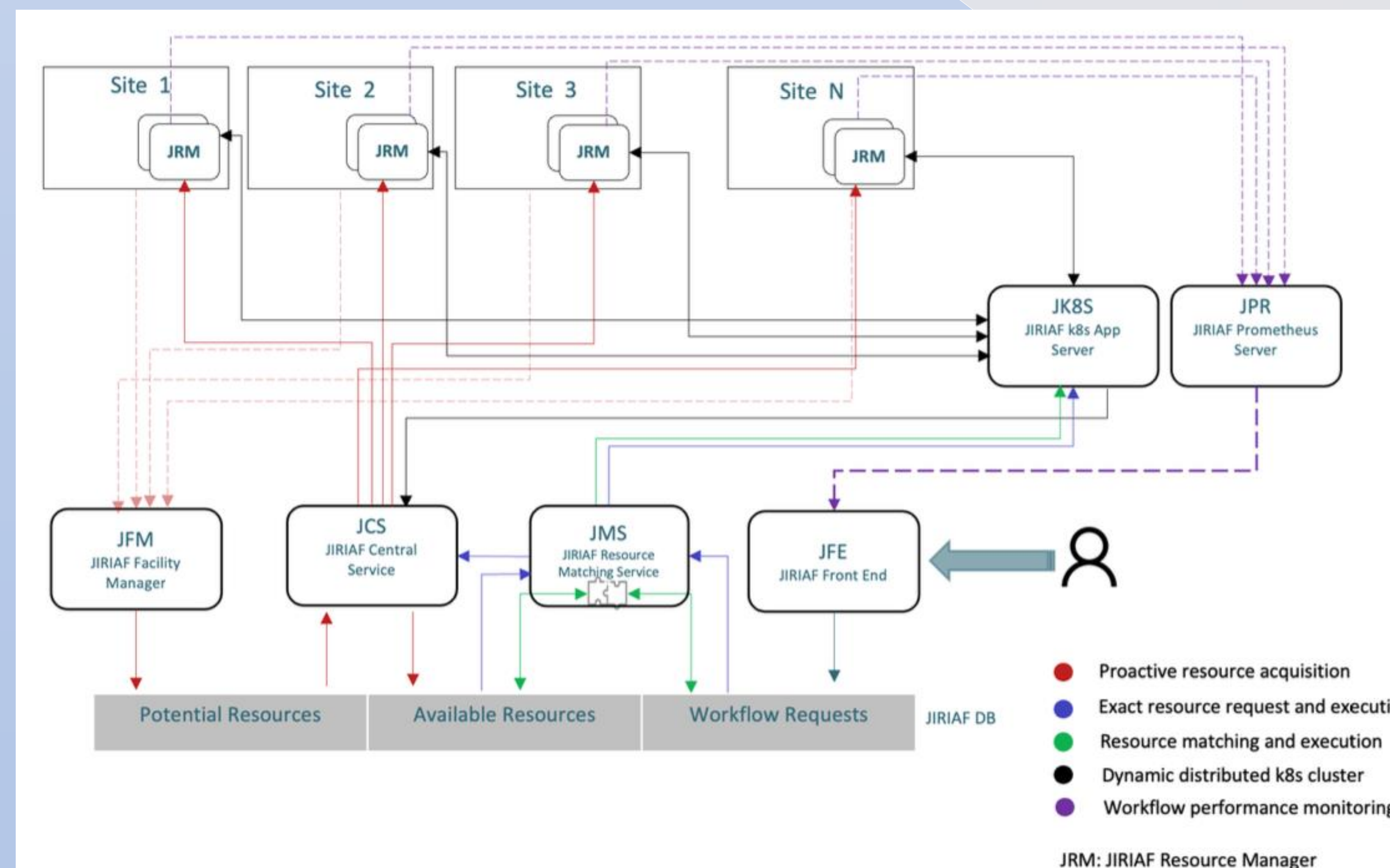
Architecture

* JFM (JIRIAF facility manager) updates the resource pool by scraping resource data from each computing facility periodically.
* JCS (JIRIAF central service) initiates pilot jobs via JRM (JIRIAF resource manager) by leasing resources reported by JFM.
* After JRMs execute, JCS updates the available resource DB table to match user request table using JMS (JIRIAF matching service algorithm).
* User requests are managed through the JFE (JIRIAF front end) component, populating the user workflow request table.

Acknowledgement