Leveraging and Evaluating Kubernetes support for HPCC Systems on Azure

Yash Mishra | Advisor: Dr. Amy Apon | Mentor: Dan Camper

Introduction

Deployment of HPCC Systems to commercial clouds can be done in multiple ways, such as Lift-and-Shift or Containerization, depending on various business needs. This project utilizes the containerized version of HPCC Systems and orchestrates the new environment via Kubernetes, targeting Microsoft Azure. In the new Kubernetes orchestration of HPCC Systems, several things appear to be different than in the legacy version. HPCC Systems components are converted to pods, completely decoupling it from the node-level dependencies. Pods run the system processes that communicate with other pods in the cluster. Moreover, the storage handling and scaling also changes. The project explores these options to understand the operation of HPCC Systems in cloud-native environment.

Kubernetes considerations on Azure

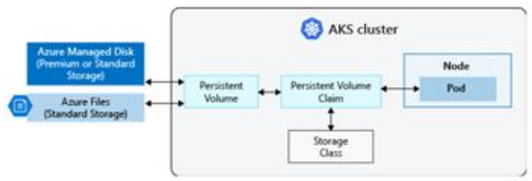
Subscription
Resource Group| Deployment Region
Primary Node Pool - Number and size of nodes in the cluster along with node type
Authentication - Service Principal or System-assigned Managed Identity
Helm Manifest Configuration

Storage Option Considerations

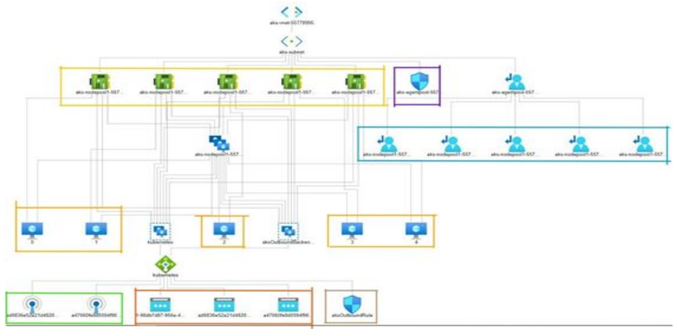
- **Azure File** - Offers SMB access to file shares. This meets the shared data requirement for dllStorage and dataStorage classes.
- **Azure Disk:** Mounted as *ReadWriteOnce,* so it is only available to a single pod. This does not meet the shared storage requirement for dataStorage and dllStorage classes in a multi-node cluster
- Orchestrated via Persistent Volumes and referenced by Persistent Volume Claims

Persistent Volumes

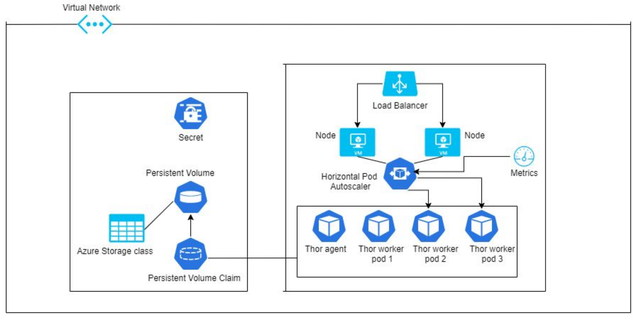
The following diagram illustrates storage architecture in Kubernetes [1]



Network Topology

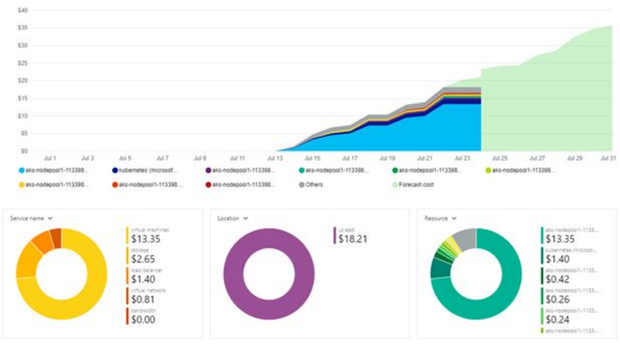


Pod Scaling and Shared Storage



Cloud Costs

Cloud costs vary by region. For example, Standard_D2s_v2 instance type costs $0.146/hour in the US East region, and costs $0.14/hr in West US region. Choosing a different region may be cheaper, but it might impact the latency.



Challenges and Future Work

- Getting data in and out of the cluster
- Persisting data longer than helm charts
- Exploring alternate storage options - Azure Blob | Azure Data Lake

References:

[1] *Concepts - Storage in Azure Kubernetes Services (AKS).* https://docs.microsoft.com/en-us/azure/aks/concept-storage
[2] *Setting up a Default HPCC Systems Cluster on Microsoft Azure Cloud Using HPCC Systems 7.8.x and Kubernetes*, Jake Smith | HPCC Systems. https://hpccsystems.com/blog/default-azure-setup
[3] *Persisting Data in an HPCC Systems Cloud Native Environment*, Gavin Halliday | HPCC Systems. https://hpccsystems.com/blog/persisting-data-cloud.