Confluent Platform and Apache Kafka on VMware Cloud Foundation

Table of contents

Confluent Platform and Apache Kafka on VMware Cloud Foundation

Executive Summary

With the rapid expansion of event stream processing, decisions at the eco-system level must be made within a matter of seconds. Event streaming is a new paradigm where data is seen as a continuous stream of events. With the rapid expansion of event stream data, businesses are relying on Apache Kafka to integrate existing systems in real time and build a new class of event streaming applications that unlock business opportunities. Confluent Platform is an enterprise-ready platform that compliments Kafka with advanced capabilities designed to help accelerate application development and connectivity, enable event transformations through stream processing, simplify enterprise operations at scale and meet stringent architectural requirements.

Operating a Kafka environment within traditional IT infrastructure can be challenging, since the demand for resources can fluctuate with business needs, leaving the Kafka cluster either under-powered or over-provisioned. IT needs a more flexible, scalable and secure infrastructure to handle with ever-changing demands of Kafka. With a single architecture that is easy to deploy, VMware Cloud Foundation™ can provision compute, network, and storage on-demand. VMware Cloud Foundation protects network and data with micro-segmentation and satisfies compliance requirements with data-at-rest encryption. Policy-based management delivers business-critical performance. For a true hybrid cloud experience, organizations can combine their on-premises data center with VMware Cloud on AWS (VMC) to address ephemeral workloads. VMware Cloud Foundation delivers flexible, consistent, secure infrastructure and operations across private and public clouds and is ideally suited to meet the demands of Apache Kafka.

Introduction

VMware Cloud Foundation is an integrated software platform that automates the deployment and lifecycle management of a complete software-defined data center (SDDC) on a standardized hyperconverged architecture. It can be deployed on premises on a broad range of supported hardware or consumed as a service in the public cloud (VMware Cloud™ on AWS or a VMware Cloud Provider™). With the integrated cloud management capabilities, the end result is a hybrid cloud platform that can span private and public environments, offering a consistent operational model based on well-known VMware vSphere® tools and processes, and the freedom to run apps anywhere without the complexity of app re-writing.

This document outlines general design and deployment guidelines for Confluent Platform and Apache Kafka on VMware Cloud Foundation 4.0.

Overview

The solution technology components are listed below:

• VMware Cloud Foundation

- VMware vSphere
- VMware vSAN
- VMware NSX Data Center

• Confluent Platform and Apache Kafka

VMware Cloud Foundation

VMware Cloud Foundation is a hybrid cloud platform designed for running both traditional enterprise applications and modern applications. It is built on the proven and comprehensive software-defined VMware® stack, including vSphere with Kubernetes, VMware vSAN™, VMware NSX-T Data Center™, and VMware vRealize® Suite. Cloud Foundation provides a complete set of software-defined services for compute, storage, network security, Kubernetes management, and cloud management. The result is agile, reliable, efficient cloud infrastructure that offers consistent operations across private and public clouds.

VMware vSphere

VMware vSphere is VMware's virtualization platform, which transforms data centers into aggregated computing infrastructures that include CPU, storage, and networking resources. vSphere manages these infrastructures as a unified operating environment and provides operators with the tools to administer the data centers that participate in that environment. The two core components of vSphere are ESXi and vCenter Server. ESXi is the virtualization platform where you create and run virtual machines and virtual appliances. vCenter Server is the service through which is used to manage multiple hosts connected in a network and pool host resources.

VMware vSAN

VMware vSAN is the industry-leading software powering VMware’s software defined storage and HCI solution. vSAN helps customers evolve their data center without risk, control IT costs and scale to tomorrow’s business needs. vSAN, native to the market-leading hypervisor, delivers flash-optimized, secure storage for all of your critical vSphere workloads, and is built on industry-standard x86 servers and components that help lower TCO in comparison to traditional storage. It delivers the agility to easily scale IT and offers the industry’s first native HCI encryption.

vSAN simplifies day-1 and day-2 operations, and customers can quickly deploy and extend cloud infrastructure and minimize maintenance disruptions. Stateful containers orchestrated by Kubernetes can leverage storage exposed by vSphere (vSAN, VMFS, NFS) while using standard Kubernetes volume, persistent volume, and dynamic provisioning primitives.

VMware NSX Data Center

VMware NSX Data Center is the network virtualization and security platform that enables the virtual cloud network, a software-defined approach to networking that extends across data centers, clouds, and application frameworks. With NSX Data Center, networking and security are brought closer to the application wherever it’s running, from virtual machines to containers to bare metal. Like the operational model of VMs, networks can be provisioned and managed independent of underlying hardware. NSX Data Center reproduces the entire network model in software, enabling any network topology—from simple to complex multitier networks—to be created and provisioned in seconds. Users can create multiple virtual networks with diverse requirements, leveraging a combination of the services offered via NSX or from a broad ecosystem of third-party integrations ranging from next-generation firewalls to performance management solutions to build inherently more agile and secure environments. These services can then be extended to a variety of endpoints within and across clouds.

Confluent Platform

Apache Kafka is a community distributed event streaming platform capable of handling trillions of events a day. Initially conceived as a messaging queue, Kafka is based on an abstraction of a distributed commit log. Since being created and open sourced by LinkedIn in 2011, Kafka has quickly evolved from messaging queue to a full-fledged event streaming platform.

Founded by the original developers of Apache Kafka, Confluent delivers the most complete distribution of Kafka with Confluent Platform. Confluent Platform improves Kafka with additional community and commercial features designed to enhance the streaming experience of both operators and developers in production, at massive scale.

Validation

We validate VMware Cloud Foundation with vSAN can support Confluent Platform by deploying an Apache Kafka cluster in a

VMware Cloud Foundation VI workload domain and running sample representative workload against the Kafka cluster. Testing will insure that VMware Cloud Foundation is able to meet Kafka infrastructure requirements and validate design assumptions about the infrastructure.

Design Assumptions

The VMware Cloud Foundation will subject Kafka VM to vMotion events and also provide resilient storage.

vMotion

vMotion is a zero-downtime live migration that allows you to move an entire running virtual machine from one physical server to another, with no downtime. The virtual machine retains its network identity and connections, ensuring a seamless migration process.

• Transfer the virtual machine's active memory and precise execution state over a high-speed network, allowing the virtual machine to switch running on the source vSphere host to the destination vSphere host.

Resilient Storage

vSAN allows the configuration of the number of failures to tolerate (FTT) as a virtual machine policy regulating the number of failures the underlying infrastructure in a cluster can sustain while the VM remains available. When a device failure occurs vSAN automatically rebuilds the components on the failed device to restore storage resiliency. The FTT number, from 0 to 3, represents the number of simultaneous device failures vSAN can withstand from the time of failure until the rebuild completes.

In vSAN a device is any one of the following:

• Capacity disk

• Cache disk

• ESXi Host

Test Tools

We leverage the following monitoring and benchmark tools in the scope of our functional validation of Kafka on VMware Cloud Foundation.

Monitoring tools

vSAN Performance Service

vSAN Performance Service is used to monitor the performance of the vSAN environment, using the vSphere web client. The performance service collects and analyzes performance statistics and displays the data in a graphical format. You can use the performance charts to manage your workload and determine the root cause of problems.

vSAN Health Check

vSAN Health Check delivers a simplified troubleshooting and monitoring experience of all things related to vSAN. Through the vSphere web client, it offers multiple health checks specifically for vSAN including cluster, hardware compatibility, data, limits, physical disks. It is used to check the vSAN health before the mixed-workload environment deployment.

Grafana

Grafana is a multi-platform open source solution for running data analytics, pulling up metrics that make sense of the massive amount of data, and monitoring apps through customizable dashboards. Available since 2014, the interactive visualization software provides charts, graphs, and alerts when the service is connected to supported data sources.

Workload generation and testing tools

HCIBench

HCIBench is an automation wrapper around the popular and proven open source benchmark tools: Vdbench and Fio that make it easier to automate testing across an HCI cluster.

Kafka CLI testing tools

The standard Apache Kafka application deployment provides several command line utilities that provide mechanism to simulation message production and consumption. We use the kafka-topic, kafka-producer-perf-test, and kafka-consumer-perf-test command line utilities to generate our test workloads.

Validation Environment Configuration

This section introduces the resources and configurations:

- Environment diagram
- Hardware resources
- Software resources
- Network configuration
- vSAN Configuration

Environment Diagram

Our VMware Cloud Foundation test environment is composed of a Management workload domain and a Virtual Infrastructure (VI) workload domain. We deploy all the VM required for the Kafka test cluster in the VI Workload Domain and all other infrastructure VM are located in the separate management workload domain (figure 1).

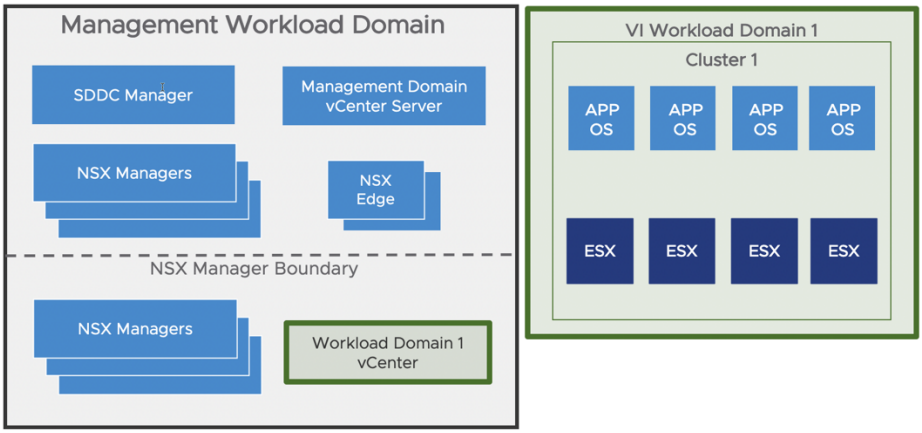

Figure 1. Environment Overview

In our deployment we use a 4-node Dell PowerEdge R640 cluster for the VMware Cloud Foundation management cluster, running multiple virtual machines. All VMware Cloud Foundation infrastructure virtual machines are deployed in the management workload domain with their default hardware configuration with the exception of vRealize Operations. Because of the small size of our environment vRealize Operations was deployed as a single medium sized VM rather than the typical multi instance cluster prescribed for high availability in larger environment (table 1).

VM Role	vCPU	memory (GB)	vm Count
Management Domain vCenter Server	8	24	1
Platform Server Controller	2	4	2
SDDC Manager	4	16	1
Management Domain NSX-V Manager	12	48	3
Management Domain NSX-V Controller	4	4	3
VI Workload Domain NSX-T Manager	12	48	3
vRealize Log Insight	8	16	3
vRealize Lifecycle Manager	2	16	1
vRealize Operations	8	32	1

Table 1. Management Domain VM

We deploy another 4-node Dell PowerEdge R640 cluster for the VI workload domain cluster running our Kafka cluster. The deployment of the Kafka virtual machines is in the quantities and configuration outlined in table 2.

VM Role	vCPU	memory (GB)	vm Count	vmdk	vmdks Size
Kafka Connect	4	8	2	1	100GB (OS)
Kafka Control Center	8	64	1	1	100GB (OS)
Kafka Broker	4	64	4	3	100GB (OS) 500GB (Data)
Kafka Zookeeper	2	16	3	2	100GB (OS) 128GB (Data)

Table 2. VI Workload Domain VM

All Kafka VM are deployed in a single resources pool in the VI workload domain cluster. This resource pool has no limits or reservations, and it is used exclusively to group related VM together (figure 2).

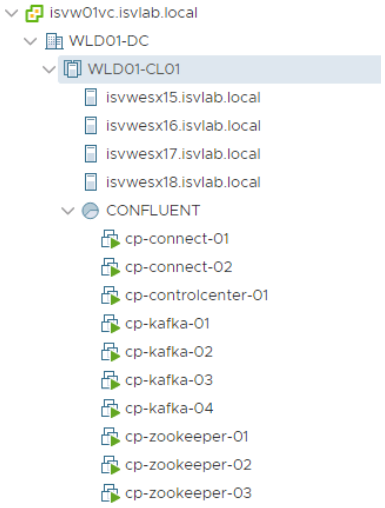

Figure 2. Kafka VM in the VI Workload Domain

Anti-affinity rules are created to prevent more than one zookeeper from running on an ESXi hosts as well as another rule to prevent more than one broker from running on an ESXi host. These rules ensure that a host failure will impact at most one broker and zookeeper. Separating the brokers provides optimal performance by balancing the workload evenly across all the physical hosts in the VI workload domain cluster (figure 3).

Figure 3. Kafka VM in the VI Workload Domain

In the VI workload domain all the Kafka VM are connected to a routed L2 segment (ls-vlan2010-vm-network-static). The Kafka

broker and zookeeper are additionally connected to a second non-routed L2 segment (ls-vlan2040-vm-network-static) for internal communication (figure 4).

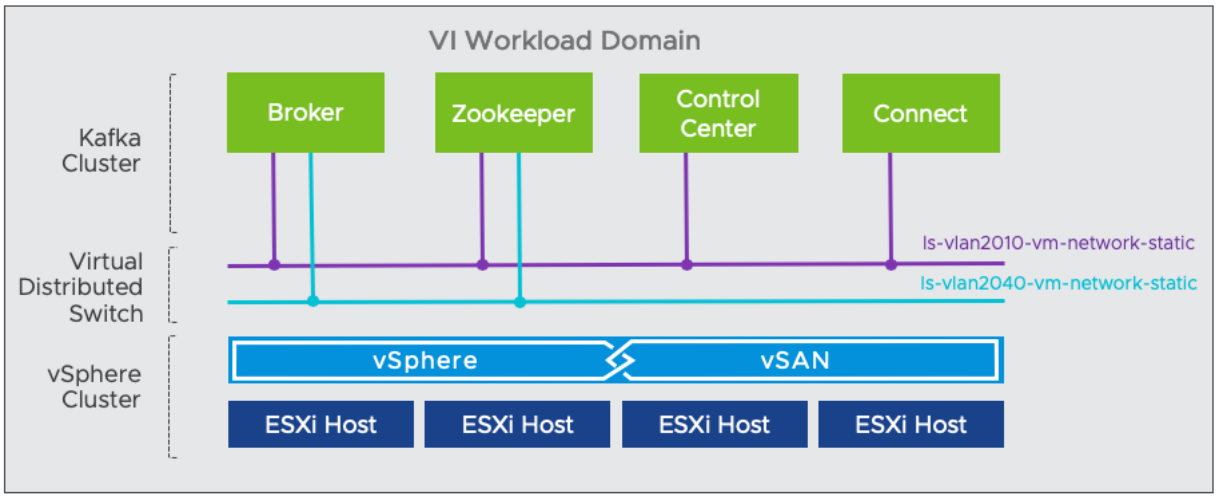

Figure 4. Kafka VM in the VI Workload Domain

Hardware Resources

In our environment a total of eight Dell PowerEdge R640 with each disk group consisting of one cache-tier NVMe and capacity-tier read-intensive SATA SSDs (table 3).

Note: Although our servers are configured with NIC capable of supporting up to 100GbE, our top of rack (TOR) switches only support a maximum speed of 40GbE per port. All server NICs are running at 40GbE, the maximum rate supported by the TOR switches.

Each server node in the cluster had the following configuration:

PROPERTY	SPECIFICATION
Server model name	Dell PowerEdge R640
CPU	2 x Intel(R) Xeon(R) Platinum 8260 CPU @ 2.40GHz, 48 core each
RAM	768GB
Network adapter	2 x Mellanox MT28800 ConnectX-5 1/10/25/40/50/100Gbps Ethernet Controller
Storage adapter	1 x Dell HBA330 Mini Adapter
Disks	Cache - Dell Express Flash NVMe P4610 Capacity – Intel D3-S4510 RI SATA SSD

Table 3. Server Hardware Configuration

The TOR networking in our environment is provided by a pair of Dell S6000-ON switches. The TOR switches provide both L2 switching and L3 routing between the subnets in our environment.

PROPERTY	SPECIFICATION
Switch model name	Dell S6000-ON
Number of ports	32 x 40GbE QSFP+
Switching bandwidth	Up to 2.56Tbps non-blocking (full-duplex)

Table 4. Dell S6000-ON Switch Characteristics

Software Resources

Testing was based on the following software resources (table 5).

Software	Version	Purpose
VMware Cloud Foundation	4.0	A unified SDDC platform that brings together VMware ESXi, vSAN, NSX and optionally, vRealize Suite components, into a natively integrated stack to deliver enterprise-ready cloud infrastructure for the private and public cloud. See BOM of VMware Cloud Foundation on VxRail for details.
Centos	7.7.1908	Operating System
Confluent Platform	5.4.1-CE	Confluent Community Edition
HCIBench	2.3.1	General purpose storage workload testing appliance for VMware vSphere environments.

Table 5. Software Resources

Network Configuration

Figure 5 shows the VMware vSphere Distributed Switches configuration for the workload domain of the VMware Cloud. Two 40 GbE vmnics were used and configured with teaming policies.

In the management workload domain NSX-T reside on the management port group of the vDS. In the Kafka VI workload domain, we use NSX-T to add segments to the VDS. For NSX-T design guidance see NSX-T Reference Design.

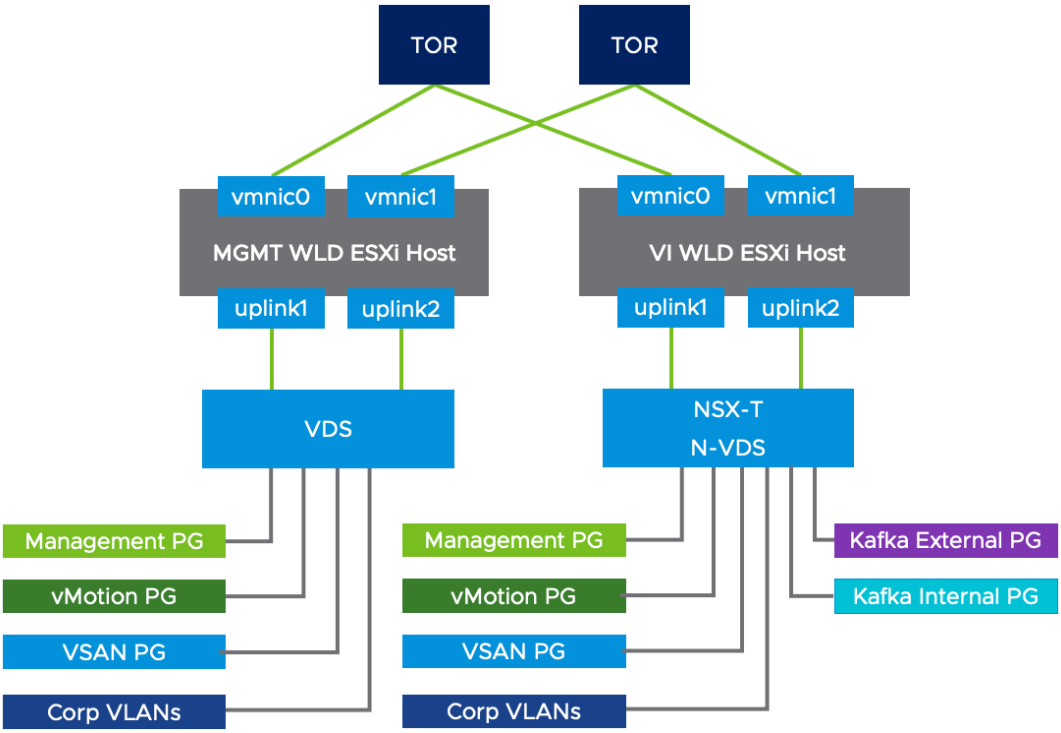

Figure 5. Distributed Switches Overview

Beyond the basic portgroups created by VMware Cloud Foundation we create two additional portgroups to provide Kafka environment with one external network (ls-vlan2010-vm-network-static) and one internal network (ls-vlan2040-vm-network-static) network.

Port Group	Teaming Policy	VMNIC0	VMNIC1
Management network	Route based on Physical NIC load	Active	Active
VM network	Route based on Physical NIC load	Active	Active
vSphere vMotion	Route based on Physical NIC load	Active	Active
vSAN	Route based on Physical NIC load	Active	Active
VXLAN VTEP	Route based on the originating virtual port	Active	Active
ls-vlan2010-vm-network-static	Load Balance Source	Active	Active
ls-vlan2040-vm-network-static	Load Balance Source	Active	Active

Table 6. Virtual Distributed Switch Teaming Policy for 2x40 GbE Profile

vSAN Configuration

Validation was conducted using the default vSAN datastore storage policy of RAID 1 FTT=1, checksums enabled, dedupe and compression disabled, and no encryption. This storage policy offers the best performance with the ability to tolerate up to one device failure (figure 6,7).

Figure 6. vSAN Storage Policy Availability Settings

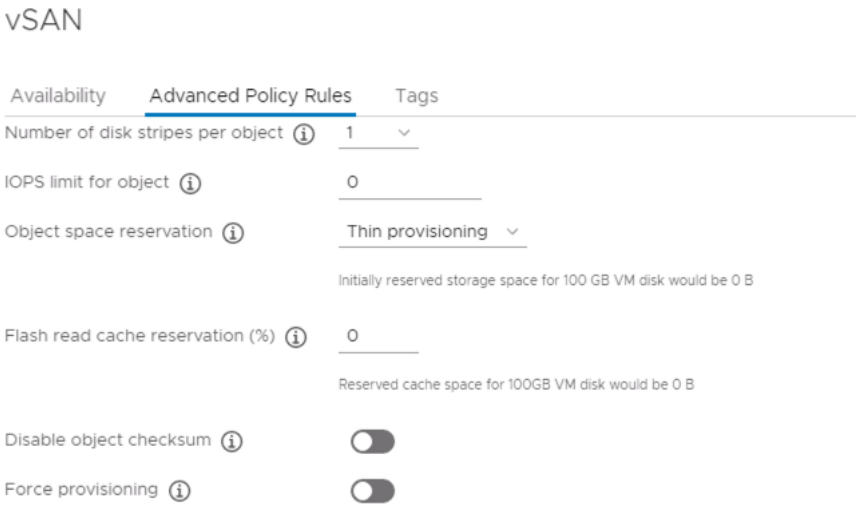

Figure 7. vSAN Storage Policy Advanced Policy Rules

Platform Validation

Prior to deployment it is highly recommended to validate the performance capabilities of the intended platform. HCIBench is the preferred tool to validate both overall and I/O specific profile performance using synthetic I/O. HCIBench provides the ability to run user-defined workloads as well as a series of pre-defined tests, known as the EasyRun suite. When leveraging EasyRun the HCIBench appliance executes four different standard test profiles that sample system performance and report key metrics.

Beyond synthetic testing it is advised to leverage the I/O tool designated by the software vendor. Given Kafka's open source status there are many published sample tests capable of simulating different representative workloads. Users should explore which tests and parameters are best suited to replicate the I/O profiles matching their actual workload. Once tests and optimal parameters are identified, a baseline test should be conducted using a subset of the selected test cases. Running a limited subset allows the user to rapidly expose potential performance anomalies present in the system or configuration while reducing in between test iterations.

Production Criteria Recommendations

Kafka provides an application level fault tolerance through application clustering. When deploying on VMware Cloud Foundation, it is best to consider the following settings within Storage Policy Based Management (SPBM) and the vCenter vSAN Cluster level settings:

SPBM

Availability:

FTT

The Number of Failures to Tolerate capability addresses the key customer and design requirement of availability. With FTT, availability is provided by maintaining replica copies of data, to mitigate the risk of a host failure resulting in lost connectivity to data or potential data loss. The FTT policy works in conjunction with VMware vSphere High Availability to maintain availability and provide consistent and near continuous uptime to workloads.

Recommendation: FTT=1

RAID

vSAN has the ability use RAID1 for mirroring or RAID5/6 for Erasure Coding. Erasure coding can provide the same level of data protection as mirroring (RAID 1), while using less storage capacity.

Recommendation: RAID1

We recommend both FTT=1 and RAID1 from a performance and cost perspective. Using RAID1 will provide the best level of performance in conjunction with FTT=1 provides operational efficiency and availability coupled with Kafka clustering.

vCenter vSAN Cluster

Dedupe/Compression

Dedupe and Compression can greatly enhance space savings capabilities, however, for optimal performance with Confluent Platform and Apache Kafka we do not recommend enabling Dedupe and Compression.

Recommendation: Disable Dedupe/Compression

Encryption

vSAN can perform data at rest encryption. Data is encrypted after all other processing, such as deduplication, is performed. Data at rest encryption protects data on storage devices, in case a device is removed from the cluster. Use encryption as per your company's Information Security requirements.

Recommendation: Enable Encryption as required per your InfoSec.

High Availability

vSphere HA provides high availability for virtual machines by pooling the virtual machines and the hosts they reside on into a cluster. Hosts in the cluster are monitored and in the event of a failure, the virtual machines on a failed host are restarted on alternate hosts.

Recommendation: HA Enabled

DRS

vSphere® Distributed Resource Scheduler™ (DRS) is the resource scheduling and load balancing solution for vSphere. DRS works on a cluster of ESXi hosts and provides resource management capabilities like load balancing and virtual machine (VM) placement. DRS also enforces user-defined resource allocation policies at the cluster level, while working with system-level constraints.

Recommendation: DRS – partially automated

Kafka Compression

We recommend using a Kafka supported type of compression (.gzip, lz4, snappy, zstd). Using compression will greatly improve performance and minimize the impact to other workloads residing on the same vSAN cluster.

Recommendation: Use compression

Kafka Partitions and Replication Factors

Environment and use case will dictate the optimal number of Kafka partitions and/or replication factors. We recommend consulting Confluent Best Practices determining the number of partitions and/or replication factor that will meet your use case needs.

Recommendation: Refer to Confluent Best Practices

Hardware Recommendations

Although VMware vSAN supports a wide range of hardware for optimal performance with applications, we recommend:

- Minimum Dual Intel Gold processors with minimum of 18 cores per socket and 2.6GHz frequency base[1]
- Minimum of 576GB RAM per vSAN node
- Minimum (4) 10GbE, preferred 25GbE Network Interface Cards
- Two disk groups minimum per vSAN node, with a minimum of (5) disks per disk group
- Mixed Use or Write Intensive NVMe SSDs for the vSAN cache tier disks
- Mixed Use NVMe or 6Gb SATA SSDs for the vSAN capacity tier disks
- Network Switches must be a non-blocking architecture and with high-buffers
- Ensure all components are on the VMware vSAN Hardware Compatibility Guide

Conclusion

VMware Cloud Foundation delivers flexible, consistent, secure infrastructure and operations across private and public clouds and is ideally suited to meet the demands of Confluent Platform and Apache Kafka. Using micro-segmentation, administrators can isolate traffic to a given set of consumers for workload and regulatory purposes. With SPBM, VMware Cloud Foundation can scale performance for both department and enterprise level clouds. Data-at-rest encryption meets both operational and regulatory compliance. CTO's and CFO's budget objectives can be achieved with dynamic provisioning, allowing enterprises to scale-up and scale-down as needed. VMware Cloud Foundation with VMware vSAN enables Apache Kafka developers to effortlessly stream data in real-time, allowing enterprises to make business critical decisions instantaneously using a single hybrid cloud platform.

Reference

- VMware Cloud Foundation
- VMware vSphere
- VMware vSAN
- VMware NSX Data Center
- VMware vRealize Suite
- Confluent Best Practices for Apache Kafka in Production

About the Authors

Charles and Christian wrote the original contents of the reference architecture:

- *Charles Lee*, *Solutions Architect* in the Solutions Architecture team of the Cloud Platform Business Unit
- *Christian Rauber*, *Staff Solutions Architect* in the Solutions Architecture team of the Cloud Platform Business Unit

Tom and Jeff reviewed and contributed to the paper:

- *Tom Nagelmeyer, Product Line Marketing Manager* in the Storage Product Marketing of the Cloud Platform Business Unit
- *Jeff Bean, Senior Technical Marketing Manager* from Confluent
- *Chris Matta, Principal Solutions Engineer* from Confluent