VxFlex Integrated Rack for Google Cloud’s Anthos

Building a hybrid cloud with VxFlex integrated rack and Google Cloud’s Anthos

Abstract

This white paper highlights a hybrid cloud solution that is powered by VxFlex integrated rack for Google Cloud’s Anthos delivering a managed Kubernetes service across public cloud and on premises data centers.

May 2019
**Revisions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 2019</td>
<td>Initial release</td>
</tr>
<tr>
<td>May 2019</td>
<td>Branding updates</td>
</tr>
</tbody>
</table>

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2019 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners. [5/6/2019] [Solution White Paper] [000050]
# Table of contents

Revisions..............................................................................................................................................2

Table of contents ......................................................................................................................................3

Executive summary .................................................................................................................................4

1 Introduction...........................................................................................................................................5
   1.1 Audience..........................................................................................................................................5
   1.2 Terminology .................................................................................................................................5
   1.3 VxFlex integrated rack overview ...............................................................................................6
   1.3.1 VxFlex OS components .........................................................................................................6
   1.4 Anthos overview .........................................................................................................................7

2 Solution overview ................................................................................................................................8
   2.1 Prerequisites..................................................................................................................................8
   2.2 Physical design ............................................................................................................................8
   2.3 Logical design ..............................................................................................................................9

3 Deploying Anthos on VxFlex integrated rack .....................................................................................12
   3.1 Deployment tasks ......................................................................................................................12
   3.2 Register GKE On-Prem cluster with GCP Console .................................................................12

4 Deploying Anthos applications on VxFlex integrated rack .............................................................18
   4.1 Sample workload: NGINX web server deployment ..............................................................18
      4.1.1 Deploying NGINX ...........................................................................................................18
      4.1.2 Scaling NGINX deployment .............................................................................................20

5 Conclusion ..........................................................................................................................................22
Containers are an abstraction at the application layer (sometimes called “operating system virtualization”). These containers package software into a standard lightweight format (for example: Docker). These containers package everything an application requires to run successfully (source code, runtime, system tools, system libraries, environment variables, and so on) making a container portable. Containers help lower the barrier of entry to developing microservice based applications. Containers enable developers to spend less time worrying about runtimes, dependencies, and differences between test/dev and production environments and more time innovating to meet the demands of today’s dynamic business environment. Container Orchestrators (COs), like Kubernetes, make running containers at production scale possible by handling the complexity of managing hundreds or thousands of containers at any one time.

Anthos from Google Cloud is an integrated platform that lets you modernize how you develop, secure, and operate hybrid cloud and cloud-native environments across the Google Cloud Platform (GCP) and your on premises data center. Anthos is built on open-source technologies pioneered by Google, including Kubernetes, Istio, and Knative, enabling consistency between cloud and on-premises environments like VxFlex integrated rack. Google Kubernetes Engine (GKE), GKE On-Prem, and Anthos Config Management are the core building blocks of Anthos. In addition, Anthos comes included with GCP Marketplace and integration with platform-level services such as Stackdriver, Cloud Build, and Binary Authorization. Customers get the benefit of managed services provided by GKE, but the choice to run their own infrastructure on-premises, in the GCP public cloud, or hybrid. For many organizations, a hybrid solution provides the best of both worlds.

Working closely with Google Cloud, Dell EMC brings to market a VxFlex integrated rack engineered system configuration that is optimized to power the full Anthos stack including GKE On-Prem operating environment enabled by Anthos. The VxFlex integrated rack has been jointly validated for Anthos¹. Customers can purchase Anthos knowing the VxFlex integrated rack provides an operationally efficient, lifecycle managed and scalable option for on premises deployments.

¹Validation on VxFlex integrated rack performed with GKE On-Prem beta release
1 Introduction

1.1 Audience
The audience for this paper includes sales engineers, field consultants, IT administrators, customers, and anyone else interested in configuring and deploying Anthos on Dell EMC VxFlex integrated rack with VxFlex OS as the underlying software defined storage layer.

Readers are expected to have an understanding and working knowledge of containers, Kubernetes, GCP, GKE, vSphere, and VxFlex OS.

1.2 Terminology
The following table lists terminology and acronyms that are used throughout this document:

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>Storage Data Client</td>
</tr>
<tr>
<td>SDS</td>
<td>Storage Data Server</td>
</tr>
<tr>
<td>MDM</td>
<td>Metadata Manager</td>
</tr>
<tr>
<td>VM</td>
<td>Virtual Machine</td>
</tr>
<tr>
<td>OS</td>
<td>Operating System</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>SAN</td>
<td>Storage Area Network</td>
</tr>
<tr>
<td>HDD</td>
<td>Hard Disk Drive</td>
</tr>
<tr>
<td>SSD</td>
<td>Solid State Drive</td>
</tr>
<tr>
<td>PCIe</td>
<td>Peripheral Component Interconnect Express</td>
</tr>
<tr>
<td>DNS</td>
<td>Domain Name System</td>
</tr>
<tr>
<td>DHCP</td>
<td>Dynamic Host Configuration Protocol</td>
</tr>
<tr>
<td>RCM</td>
<td>Release Certification Matrix</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCP</td>
<td>Google Cloud Platform</td>
</tr>
<tr>
<td>GKE</td>
<td>Google Kubernetes Engine running on GCP</td>
</tr>
<tr>
<td>GKE On-Prem</td>
<td>Google Kubernetes Engine running on premises</td>
</tr>
<tr>
<td>GCE</td>
<td>Google Compute Engine</td>
</tr>
</tbody>
</table>
1.3 **VxFlex integrated rack overview**

VxFlex integrated rack is a rack-scale engineered system built by combining VxFlex OS storage virtualization software with Dell EMC PowerEdge servers and integrated networking to deliver flexibility, scalability and capacity delivered in a rack-scale form factor. As an engineered system, all hardware and software components of the system Local storage resources are combined to create a virtual pool of block storage with varying performance tiers. VxFlex integrated rack provides you with the following key benefits as your scalable IaaS platform:

- Pre-tested, pre-integrated, hyper-converged system
- Fully integrated rack-scale fabric
- Simplified management and operation
- Software-defined for workload flexibility
- Lifecycle management and ease of upgrade
- Faster, more powerful handling of workloads with 14th Generation PowerEdge Servers

VxFlex OS is the key enabler and provides an unmatched combination of performance, resiliency and flexibility to address enterprise data center needs. The unique features of VxFlex OS make it an excellent complement to Kubernetes deployed in virtual machines or bare metal servers for stateful applications.

VxFlex OS supports a single, scalable block storage service across hypervisors, container platforms and other data center services. VxFlex OS offers true block storage as a service:

- Provisioned natively through Kubernetes
- Dynamically create and delete volumes on demand
- Support quality of service and security context through container storage interface
- Dynamically scale storage service to match demand
- Support fully non-disruptive updates without future fork-lift migrations

1.3.1 **VxFlex OS components**

**Storage Data Client (SDC)**

- Provides front-end volume access to applications and filesystems
- Installed on servers consuming storage
- Maintains peer-to-peer connections to every SDS managing a pool of storage

**Storage Data Server (SDS)**

- Abstracts local storage, maintains storage pools, and presents volumes to the SDCs
- Installed on servers contributing local storage to the VxFlex OS cluster

**Metadata Manager (MDM)**

- Oversees storage cluster configurations, monitoring, rebalances, and rebuilds
• Highly available, independent cluster installed on 3 or 5 different nodes
• May reside alongside SDCs and/or SDSs, or on separate nodes
• Sits outside the data path

**Gateway**

• Performs installation and configuration checks
• Acts as an endpoint for API calls and passes them to MDM

### 1.4 Anthos overview

Anthos is an integrated platform that lets you modernize how you develop, secure, and operate hybrid cloud and cloud-native environments. Anthos is built on open-source technologies pioneered by Google Cloud, including Kubernetes, Istio, and Knative, enabling consistency between cloud and on-premises environments like VxFlex integrated rack. GKE, GKE On-Prem, and Anthos Config Management are the core building blocks of Anthos. In addition to these, Anthos includes with GCP Marketplace and integration with platform-level services such as Stackdriver, Cloud Build, and Binary Authorization.

Anthos puts all your IT resources into a consistent development, management, and control framework, automating away low-value and insecure tasks across your VxFlex integrated rack and GCP infrastructure.

Within the context of GCP, the term hybrid cloud describes a setup in which common or interconnected services are deployed across multiple computing environments, one based in the public cloud, and at least one being on premises. Anthos provides you with a consistent platform for building and managing applications across hybrid infrastructures and helps your developers become more productive across all environments. Anthos provides all the mechanisms required to bring your code into production reliably, securely, and consistently, with minimal risk.

![Anthos Diagram](image)

**Figure 1** Anthos

A hybrid cloud strategy lets you extend the capacity and capabilities of your IT, without up-front capital expense investments by using the public cloud, as well as preserve your existing investments by adding one or more cloud deployments to your existing infrastructure. For more information, refer to [Hybrid and Multi-Cloud Architecture Patterns](#).
Solution overview

2.1 Prerequisites

Deploying Anthos on VxFlex integrated rack requires the following prerequisite conditions be met:

- Fully configured and working VxFlex integrated rack system of RCM 3.5 or greater
- Google Account, GCP Service Account and a billing enabled GCP project
- vSphere 6.5 environment
- At least one VMFS datastore with 2 TB capacity
- Ability to create required DNS entries
- F5 BIG-IP Local Traffic Manager (LTM) v12 or v13 virtual appliance installed and licensed
- Minimum of 20 reserved IP Addresses (DHCP or static) in three networks
- Network access to Google Cloud (googleapis.com)
- Official GKE On-Prem installer downloaded (OVA file)

2.2 Physical design

This solution encompasses a fully functional VxFlex integrated rack system which is engineered for complete fault redundancy and ability to scale out across compute and storage dimensions. At a minimum, three controller nodes and HCI nodes running VMWare ESXi 6.5 or greater are required to deploy GKE On-Prem. Additionally, at least one outbound connection to googleapis.com is required to complete the GKE On-Prem cluster registration and deployment processes. Figure 2 describes how ANTHOS can manage multiple GKE On-Prem clusters across geographically dispersed VxFlex integrated rack systems through a single pane of glass with GCP console.
2.3 Logical design

From a logical perspective, traffic flows into and out of the on-premise system using an F5 BIG-IP Logical Traffic Manager (LTM) virtual appliance. The F5 BIG-IP LTM creates dynamic connections between the compute nodes and the external network interfaces. Current best practice for VxFlex integrated rack is to create three special-purpose networks used for management, internal, and external traffic:

- GKE On-Prem Admin Network
- GKE On-Prem Internal Network
- GKE On-Prem External Network

Figure 2 Physical design of Anthos with multiple VxFlex integrated rack systems geographically dispersed.
Create a new compute cluster with at least one compute server and one resource pool in the production workload vCenter to host the workloads in the production vCenter. This cluster requires DRS as well as one resource pool.

Figure 3 describes the logical configuration between GKE On-Prem clusters running in on premises data centers on VxFlex integrated rack, GKE clusters and Anthos hosted on GCP. In this architecture, applications running on GKE On-Prem cluster can be exposed internally or externally to the web without traffic passing through GCP.

**Note:** In Figure 3, the GKE On-Prem Admin Network connection from the on premises datacenter to Anthos is outbound only.

---

The production vCenter server hosts multiple virtual machines in the new resource pool that comprise a virtual GKE On-Prem compute cluster. The application workloads are processes that run inside one of the GKE On-Prem compute cluster virtual machines. When an application gets deployed into the GKE On-Prem compute cluster, no additional vSphere virtual machines are created. The application runs inside of the GKE
On-Prem compute cluster virtual machines. If additional workload capacity is required, the GKE On-Prem compute cluster is expanded using the `gkectl` command line utility or the Kubernetes Cluster API.

Here is a summary of the correlation between vSphere VMs and GKE On-Prem cluster servers:

- GKE On-Prem cluster is a collection of vSphere 6.5 VMs running as a collective instance.
- A GKE On-Prem cluster’s performance profile (for example: RAM, CPU, storage, and so on) is the additive sum of all vSphere 6.5 VMs running in the GKE On-Prem cluster.
- Multiple GKE On-Prem clusters can exist in a single vCenter deployment.
3 Deploying Anthos on VxFlex integrated rack

Deploying the Anthos software requires knowledge with vCenter, Linux command-line skills, and an understanding of the existing network topology. The time to deploy the code varies based on network speed and familiarity with the tools that are discussed below but can generally be completed within a few hours.

**Note:** Deployment process is subject to change in future releases of Anthos.

3.1 Deployment tasks

Full deployment of Anthos is not covered in this white paper, however, deployment may be broken down into the following high-level steps:

1. Gather prerequisite details.
2. Prepare vCenter.
3. Deploy and configure F5 BIG-IP LTM.
4. Deploy GKE On-Prem Admin VM on VxFlex integrated rack.
5. Create new GKE On-Prem Admin cluster on VxFlex integrated rack.
6. Register GKE On-Prem Cluster with GCP Console.

3.2 Register GKE On-Prem cluster with GCP Console

Once the GKE On-Prem cluster configuration is complete, the final step is to register the cluster in GCP. This section covers the process of registering a GKE On-Prem cluster though the GCP console.

**Note:** From this point onward, all activities are accomplished from the GCP Console unless specified. Exact wording in GCP console is subject to change. No additional work is needed on the VxFlex integrated rack infrastructure.

1. Go to [https://console.cloud.google.com](https://console.cloud.google.com) in your browser.
2. Log in to the Google Cloud Platform with your account.
3. Select the Kubernetes Engine from the Google Cloud Platform menu.
4. From Kubernetes Engine, click Clusters, and then REGISTER CLUSTER.
5. Follow the prompts and input the required information to register your previously created GKE On-Prem cluster.
a. Input the **Cluster name** and **Service account key**, and then click **CONTINUE**.

**Note:** The **Service account key** is created during GKE On-Prem cluster creation with `gkectl` from the GKE On-Prem Admin VM.

b. Click **DOWNLOAD GKE CONNECT MANIFEST** to download a local copy of the YAML configuration file. Then copy the manifest file to the GKE On-Prem Admin VM deployed on VxFlex and run the `kubectl` command shown in the GCP Console. After you see confirmation of the cluster connection in GCP console, click **CONTINUE**.
c. Next, set cluster labels for **Location** and **Provider**.

**Note:** The **Location** parameter for GKE On-Prem clusters is analogous to a Google Compute Engine (GCE) zone for GKE clusters on GCP. For more information on GCE regions and zones see: [https://cloud.google.com/compute/docs/regions-zones/](https://cloud.google.com/compute/docs/regions-zones/).
d. The new cluster name now appears in the list of GKE clusters alongside all other clusters.
5. After the GKE On-Prem cluster has been registered with GCP the Kubernetes cluster nodes are visible from GCP console by clicking on the cluster name and then clicking on **Nodes** from the Kubernetes cluster details page.

**Note:** The Kubernetes cluster nodes for GKE On-Prem clusters are running as vSphere VMs on VxFlex integrated rack.

**Note:** The Kubernetes nodes shown in the GCP console have the same name as the worker node VMs running on VxFlex integrated rack. GKE On-Prem Admin VMs, which can be seen from vSphere, are not shown to the user in GCP as these are being managed by Google Cloud.
Deploying Anthos on VxFlex integrated rack

GKE On-Prem Admin VMs

GKE On-Prem Cluster (worker) VMs
Deploying Anthos applications on VxFlex integrated rack

After the GKE On-Prem cluster has been deployed and registered, you can easily deploy workloads using the Google Cloud Console tool.

**Note**: Deploying applications to the GKE On-Prem cluster running on VxFlex integrated rack system is no different that deploying applications to GKE running on GCP. GKE On-Prem clusters can be selected as the workload destination cluster in the same way GCP datacenter would be selected. End-users use a single, simplified interface to deploy applications without requiring in-depth knowledge of the underlying infrastructure.

### 4.1 Sample workload: NGINX web server deployment

In this example, we will be deploying NGINX - a free, open-source HTTP server and reverse proxy that is well known for its high performance, stability, rich feature set, simple configuration, and low resource consumption. Since it is already in the Google market place, deploying NGINX requires just a few mouse clicks.

**Note**: For more information on NGINX, see [https://www.nginx.com/resources/wiki](https://www.nginx.com/resources/wiki)

### 4.1.1 Deploying NGINX

1. From the GCP Console, select the **Kubernetes Engine** icon under the Navigation Menu.
2. Select the **Workloads** icon on the left and then click **Deploy**. Then follow the prompts:

   ![Google Cloud Console](image)

   a. **Edit Container**
      
      i. Select **Existing container image** (for example: nginx:latest).
      ii. Click **Continue**.

   b. **Edit Configuration**
      
      i. Enter the **Application Name** (for example: nginx-test).
      ii. Enter the minimum number of Pods to start.
      iii. Enter the appropriate Namespace (for example: “nginx-deployment”).
iv. Select the appropriate Destination Cluster (ie: gke-on-prem-US)

3. Click Deploy.
4. After the new workload has been deployed, view it in the Workloads window of Google Cloud Platform console:

Note: Since the NGINX pods run inside the GKE On-Prem cluster virtual machines, you will not see any additional virtual machines in vCenter.
4.1.2 Scaling NGINX deployment

Once deployed, scaling the number of pods in the deployment to meet the demands of your application is accomplished easily with a few clicks.

1. Select the workload and click the **SCALE** icon at the top right corner of the Google Cloud Console.
2. Enter the desired number of pod replicas and click **SCALE**.
3. Refreshing the console window shows the new pods as they come online:

   Note: If you do not refresh right away the new pods may have passed the **PodInitializing** status and show **Running**.
Deploying Anthos applications on VxFlex integrated rack
5 Conclusion

Engineered together with Google Cloud team, this hybrid cloud solution provides you with best in class hyperconverged infrastructure from Dell EMC and managed Kubernetes by Google Cloud. You can leverage scale-out capabilities of GKE On-Prem for container-runtime and VxFlex integrated rack for scale-out infrastructure, with enterprise class reliability. You can add storage and/or compute on the fly with no downtime to achieve cloud-like scale on-premises.

This solution lets you streamline operations and provides portability between on-premises and cloud through a managed Kubernetes service, GKE On-Prem. Given the high-availability of VxFlex integrated rack combined with the HA of the GKE On-Prem, you have a best-in-class solution.

Additionally, you can use your private repositories allowing your teams to leverage them between these two environments for their development, testing and production deployments as needed. This simplifies your operations as you can leverage the same Kubernetes tooling on-premises and in the cloud. This includes CI/CD, configuration management, service-mesh, and monitoring and logging.