

Provisioning and Managing AWS Storage Gateway

UNDERSTANDING AWS STORAGE GATEWAY

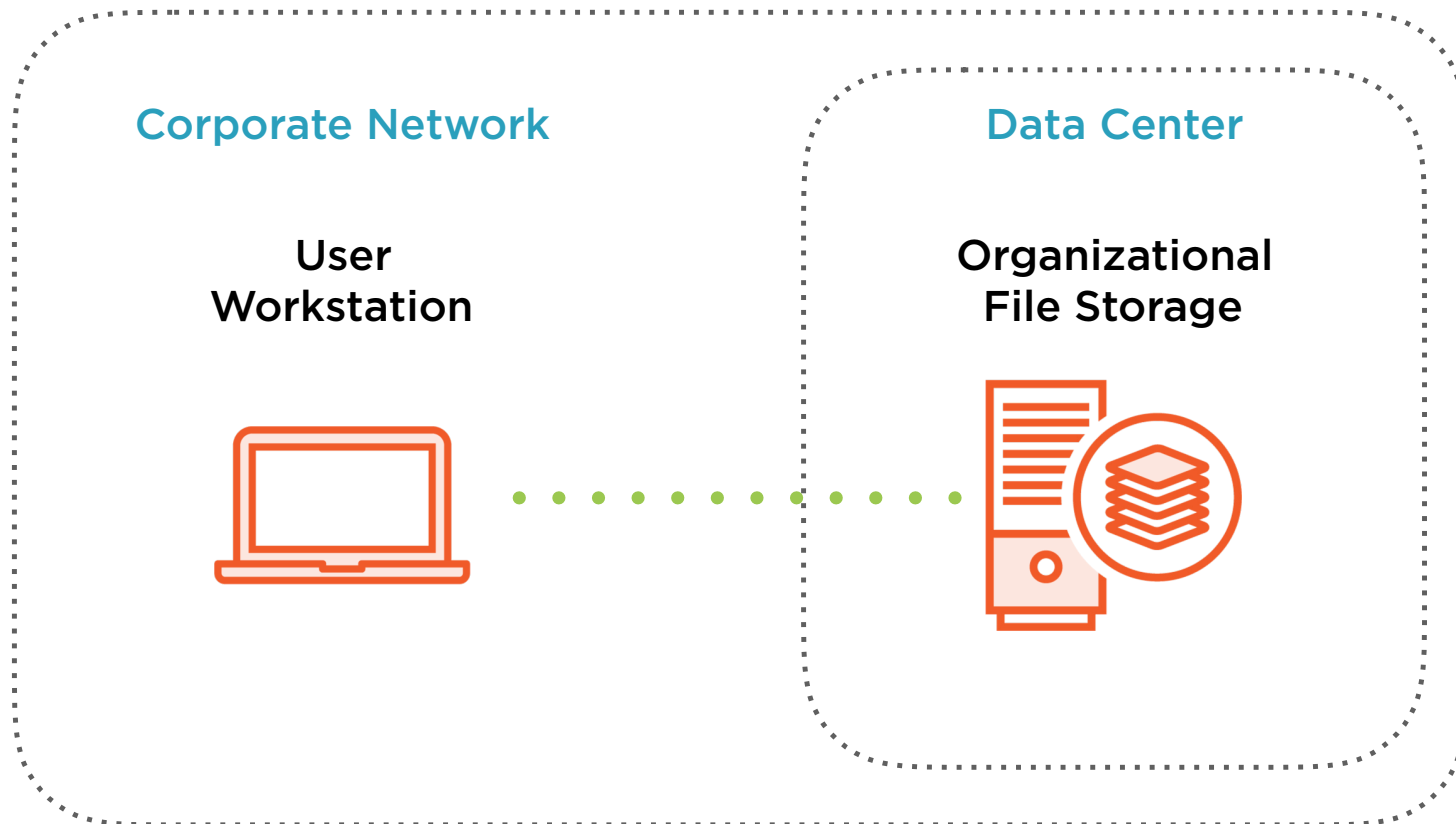


David Tucker

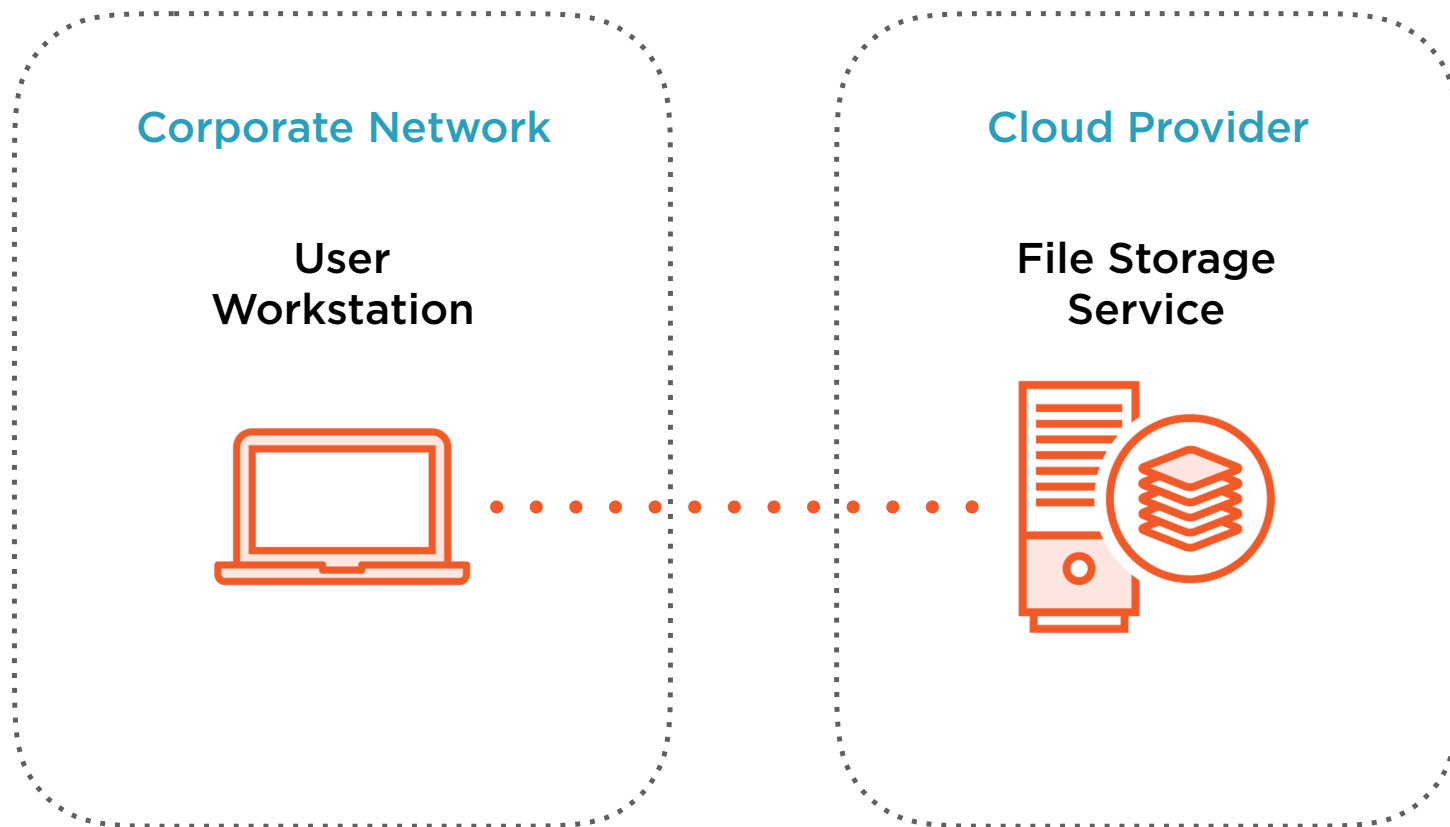
TECHNICAL ARCHITECT & CLOUD CONSULTANT

@_davidtucker_ davidtucker.net

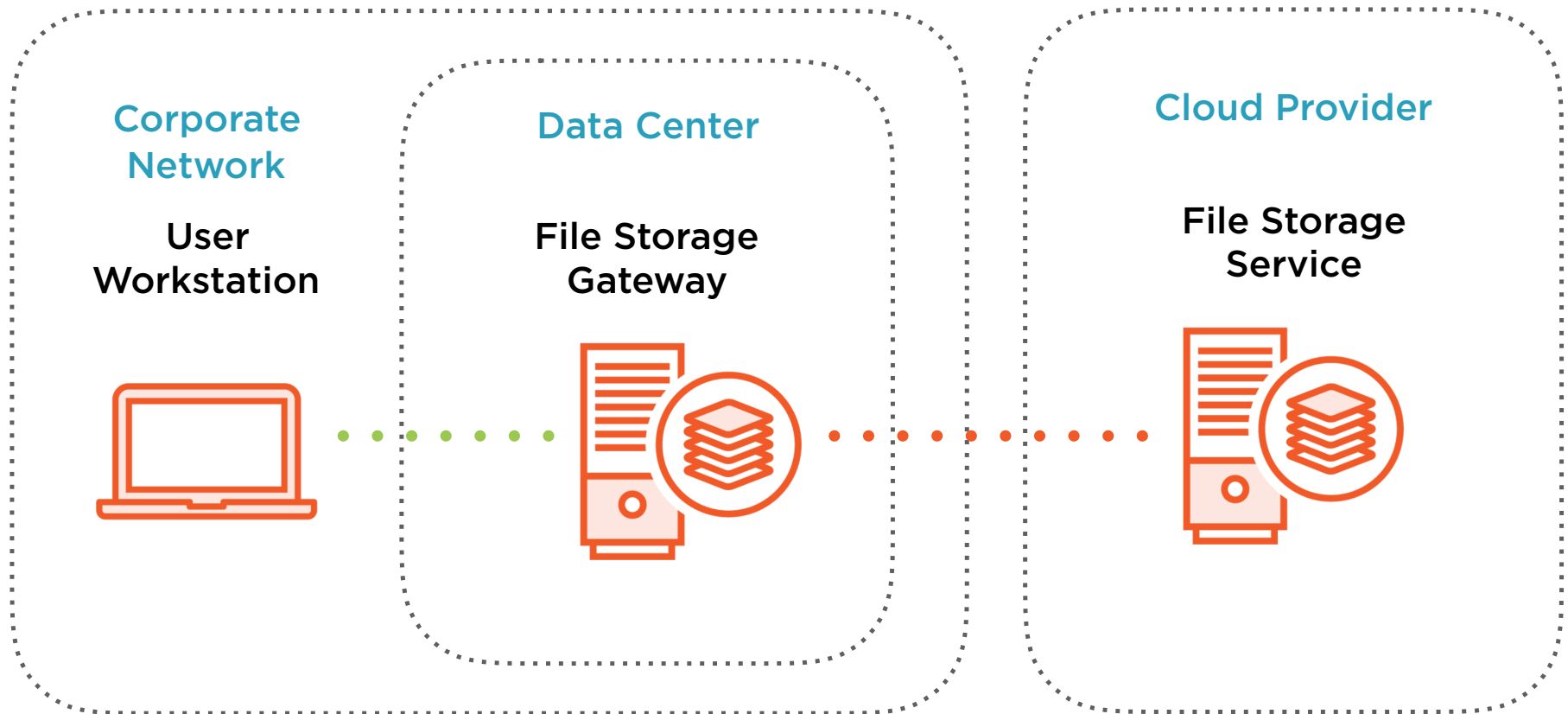
Traditional Data Center



Cloud Storage Service



Hybrid Cloud Approach



Overview

Understanding the AWS Storage Gateway service and its capabilities

Reviewing the three different gateway interface types

Examining the key components of the AWS Storage Gateway architecture

Exploring the different options for launching a gateway instance

Configuring the demo architecture we will be using

Introduction to AWS Storage Gateway

AWS Storage Gateway



Enables hybrid cloud storage to connect on-premises environments and AWS

Integrates existing workflows and solutions with AWS storage services

Leverages industry standard storage protocols including:

- NFS
- SMB
- iSCSI, iSCSI-VTL

AWS Storage Gateway Use Cases

Backup & Archiving

Disaster Recovery

**Transfer for Cloud
Workflows**

Tiered Storage

AWS Storage Gateway Interfaces



File

Store and retrieve files in S3 using SMB & NFS protocols



Volume

Utilize block storage for applications using the iSCSI protocol



Tape

Enable virtual tape library for use with a backup application

AWS Storage Gateway Benefits

Caches frequently used data for low latency access on-premises

Stores data securely and durably in AWS services

Optimizes transfer between your data center and AWS

Provides native integration with many AWS services

Enables organizations to leverage their existing file based workflows

AWS Storage Gateway Encryption

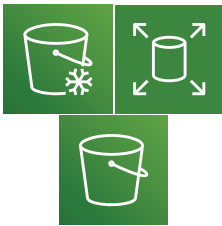
Data transferred to AWS is encrypted with SSL, and data is stored in S3 with service keys (SSE-S3). You can also leverage customer managed keys with AWS KMS.

AWS Storage Gateway Compliance

HIPAA Eligible

PCI DSS Compliant

AWS Integrations



Storage



Monitoring



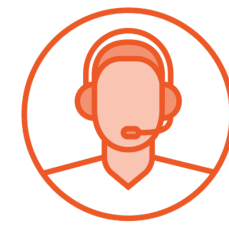
**Identity
Management**



Security



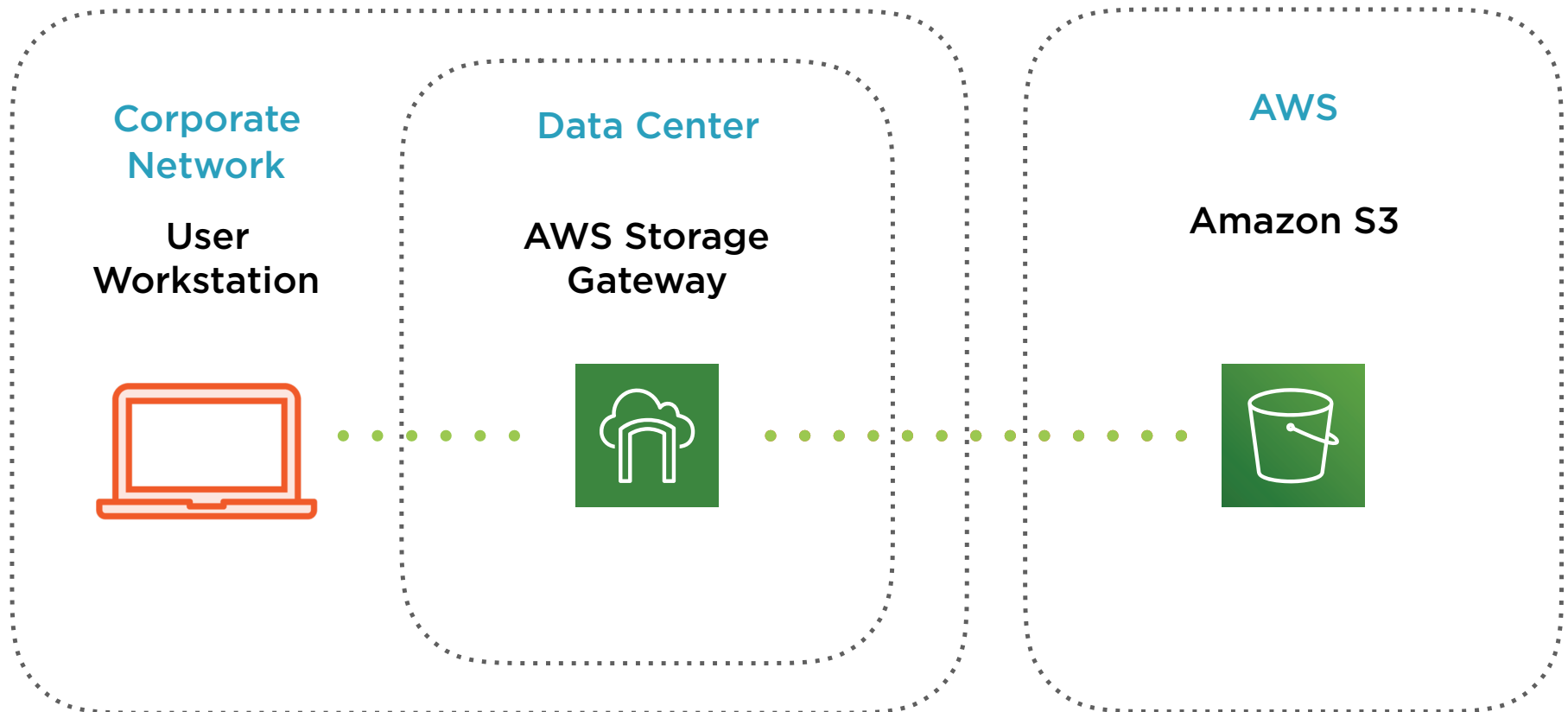
Backup



Support

Storage Gateway Components

AWS Storage Gateway



AWS Storage Gateway

The gateway software runs as a VM either in the cloud or on-premises. The VM is stateless so you can create and manage new instances based on changing demand.

Supported Hypervisors

VMWare ESXi

Microsoft Hyper-V

EC2 Instance

Hardware Appliance

Launching an AWS Storage Gateway



Service is accessed through the AWS Console and the gateway itself

Virtual appliance gateway can be downloaded for supported hypervisors

Gateway is associated with your AWS account through the activation process

Console provides monitoring of the data transfer and storage interfaces

Storage Gateway Interfaces

AWS Storage Gateway Interfaces



File

Store and retrieve files in S3 using SMB & NFS protocols



Volume

Utilize block storage for applications using the iSCSI protocol



Tape

Enable virtual tape library for use with a backup application

File Gateway Overview

Enables you to store and retrieve objects in Amazon S3 using NFS and SMB

Allows you to access uploaded objects in Amazon S3 and utilize S3 API

Enables you to leverage your existing file-based workflows and apps

Volume Gateway Overview

Provides block storage for your applications using iSCSI

Stores block data in Amazon S3

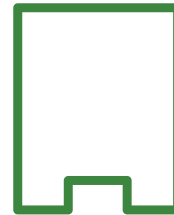
Allow you to take EBS snapshots of volume and launch within EC2

Integrates with AWS Backup

Modes for Volume Gateways



Cached Mode
Primary data is
written to S3 with a
local cache



Stored Mode
Primary data is stored
locally with async
backup to S3

Tape Gateway Overview

Provides your backup application with an iSCSI virtual tape library (VTL)

Data is stored within Amazon S3

Tapes can be archived to Amazon S3 Glacier and Glacier Deep Archive

Integrates with backup applications to support existing backup process

Hardware Appliance

Hardware Appliance Overview

Utilizes a Dell EMC PowerEdge R640 Server with a validated configuration

Includes storage gateway pre-installed

Supports management from the management console or API

Enables creation of all volume types and storage interfaces

Supports running only one gateway at a time

Hardware Appliance Use Cases



Can be leveraged for remote office outside of corporate data center



Can be leveraged when there is not an existing virtual server infrastructure available



Can be leveraged when the experience necessary to manage virtual server infrastructure is not available

How to Purchase Appliance

Only available as a purchase, no rental options

You fully own the appliance once purchased

Can be shipped to US and European addresses

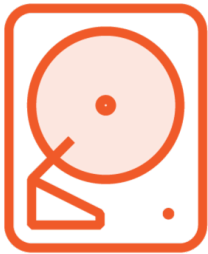
You will be redirected to Amazon.com to purchase the appliance

Comes with a 3 year warranty from Dell

Not supported in all regions

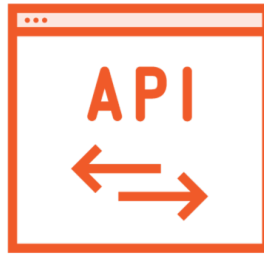
Service Pricing

Billing Considerations



Storage

You pay for the amount of storage you use



Requests

You pay for requests made to AWS services such as Amazon S3



Data Transfer

You are charged for data written to AWS from your gateway

Storage Costs

File Gateway

Billed for the space you
take up in S3

Volume & Tape Gateway

Billed for the amount of
space you actually take
up in volumes

EBS Snapshots

Billed in the same manner
as other EBS snapshots

Archival Storage Costs



Virtual tapes in the tape gateway can be archived into Amazon S3 Glacier

Retrieval fee applies for archived virtual tapes

You will be charged an early deletion fee if you data is deleted too early

Data Transfer Costs

Writing Data to AWS

You are charged \$0.01 per GB with a max of \$125/month per gateway.

Reading Data from AWS

There are no charges for reading data from AWS.

Requests Cost



Amazon S3 requests are billed when using the File Gateway

This includes all Amazon S3 requests made by the gateway

Caching will reduce the number of Amazon S3 requests that are required

Preparing to Launch

Demo Configuration

We will leverage two separate regions for the demos

- One that will represent our data center (*us-west-2*)
- One that we will use to launch our storage gateway (*us-east-1*)

Each region will be configured with a VPC with a single public subnet

Demo

Configure the VPC for our data center region

Configure the VPC for our cloud region

Verify that we can launch storage gateway

Summary

Summary

Understanding the AWS Storage Gateway service and its capabilities

Reviewing the three different gateway interface types

Examining the key components of the AWS Storage Gateway architecture

Exploring the different options for launching a gateway instance

Configuring the demo architecture we will be using

Leveraging the File Gateway



David Tucker

TECHNICAL ARCHITECT & CLOUD CONSULTANT

@_davidtucker_ davidtucker.net

Overview

Reviewing file gateway capabilities and configuration options

Implementing a file gateway instance on Amazon EC2

Configuring an NFS file share for our file gateway

Configuring an SMB file share for our file gateway

Examining how to leverage Amazon S3 alongside our file gateway

File Gateway Deep Dive

File Gateway Use Cases



Migration of corporate data from on-premises data center while still providing local file system access



Backup of corporate data at an object level



Migration of local data for cloud-based processing workflows

Amazon S3 Integration with File Gateway

**Access Secured with
IAM Role**

**One to One Mapping
of Files and Objects**

**Enables Lifecycle,
Versioning, and Replication**

**Supports POSIX-style file
metadata on objects**

File Gateway Limits

Max File Size

Max size of a single file is
5 TB (same as S3)

Share Limits

Max of 10 file shares on
a gateway

Gateway Size

Returns a total gateway
size of 8 EB

Supported S3 Storage Classes

S3 Standard

S3 Standard - Infrequent Access (IA)

S3 One Zone - Infrequent Access (IA)

NFS Share Capabilities

Supports versions 3 and 4.1

Can limit access to specific NFS clients

Can limit access to specific networks

Supports read-only or read-write

Supports user permission squashing

SMB Share Capabilities

Supports versions 2 and 3

Can specify Active Directory (AD) users only

Can specify authenticated guest user access

Can specify read-only or read-write and limit to specific AD users and groups

Supports both on-premises Microsoft AD and in-cloud Active Directory

Can export with just a username and password

Data Transfer Optimization

Listing files only requires a query of object metadata

Utilizes multi-part uploads of file data

Data only downloaded when accessed by a client

Only changed data is uploaded to Amazon S3

File Gateway Cache

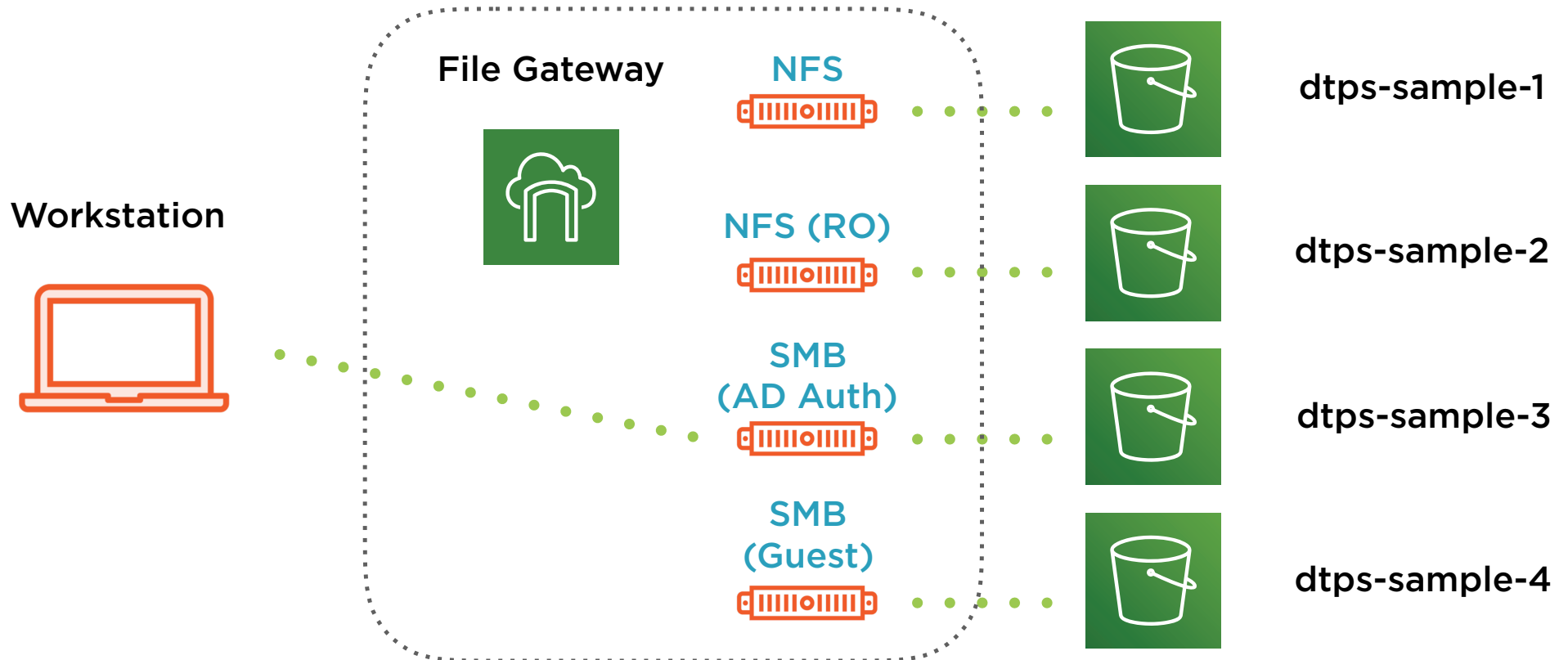
File gateway manages its own cache

Data is only evicted from the cache when it runs out of space

Considerations for cache size:

- Size of working dataset where low latency access is needed
- Size of files written to the gateway by your applications

Sample File Gateway Configuration



Readers & Writers

**Only leverage a single writer
for your S3 bucket**

**Multiple readers per bucket
may be leveraged**

**File gateway does not report
on multi-write conflicts**

**File gateway does not block
multiple writer share config**

Limitations

Symbolic and hard links are not supported

Renaming will appear atomic to the client but is not instantaneous

Objects managed by the gateway should only be modified through the gateway

Concurrent modifications from multiple NFS clients can cause issues

Additional Configuration Options

Cross-account

Buckets in other AWS accounts may be used

Default Metadata

Default file metadata can be configured per share

MIME Types

Guessing of mime types can be enabled per share

File Gateway Demo Introduction

Globomantics Call Center



GLOBOMANTICS

Current call center software exports each call as an MP3 to a network file share

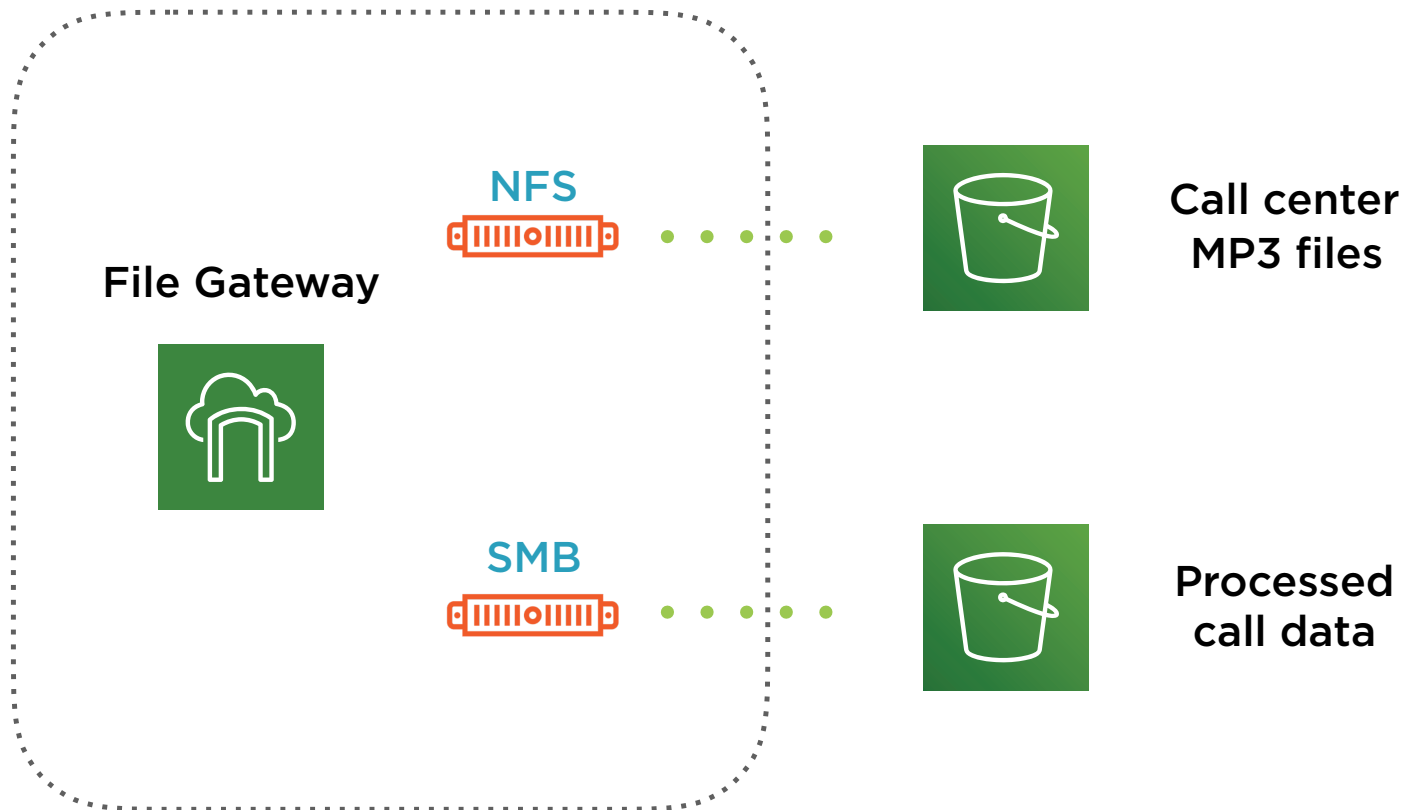
Call center leadership wants to leverage Amazon Transcribe for transcription

Call center leadership interested in using Amazon Comprehend sentiment analysis

IT team would like additional durability of data after a recent hardware failure

Consultant team needs access to review a selection of support calls

File Gateway Configuration



Demo

Provision a file gateway instance on Amazon EC2

Configure an NFS file share for our file gateway

Configure SMB file shares for our file gateway

Examine how file gateway data is stored within Amazon S3

Implement lifecycle rules to transition data to Amazon S3 Glacier

Summary

Summary

Reviewed file gateway capabilities and configuration options

Implemented a file gateway instance on Amazon EC2

Configured an NFS file share for our file gateway

Configured an SMB file share for our file gateway

Examined how to leverage Amazon S3 alongside our file gateway

Utilizing the Volume Gateway



David Tucker

TECHNICAL ARCHITECT & CLOUD CONSULTANT

@_davidtucker_ davidtucker.net

Overview

Reviewing volume gateway capabilities and configuration options

Connecting to iSCSI volumes from the volume gateway

Performing a manual backup of a volume to an EBS snapshot

Creating automated backups with volume snapshot schedule

Automating backup & scheduled backup deletion with AWS Backup

Volume Gateway Deep Dive

Volume Gateway & Amazon S3

Data for volumes is stored within S3

Bucket is managed by AWS Storage Gateway

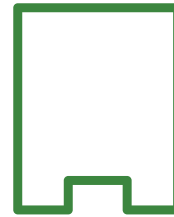
Bucket does not support lifecycle and versioning

Data cannot be accessed by the S3 API

Modes for Volume Gateways



Cached Mode
Primary data is
written to S3 with a
local cache



Stored Mode
Primary data is stored
locally with async
backup to S3

Volume Gateway Limits

Volume Limit

Gateway can support up to 32 volumes

Cached Mode

Supports 32TB per volume up to 1PB

Stored Mode

Supports 16TB per volume up to 512TB

Volume Snapshot Use Cases

Cached Mode

Leveraging snapshots can preserve versions of your data in volume

Stored Mode

Utilizing snapshots provides durable backups of data in S3

Snapshots & Volume Gateway

**Can be taken without
unmounting volume**

**Stores changed data since
previous snapshot**

**No limits on the number of
snapshots**

**Can be restored as gateway
volume or EBS volume**

Backup Approaches for Volume Gateway

Manual

Snapshots can be taken
of volumes at any time

Snapshot Schedule

Within storage gateway a
schedule can be added

AWS Backup

Service provides
additional automation

AWS Backup & Volume Gateway



Provides completely separate process from snapshot schedule

Provides greater detail in configuring when backups occur

Enables deletion of snapshots based on specific criteria

Requires that you are in the same region where your gateway is activated

Data Optimization for Volume Gateway

Compressed Transfer

Data is compressed when transferred between gateway & S3

Compressed Storage

Data remains compressed when stored in S3

Volume Gateway Demo Introduction

Media Storage



GLOBOMANTICS

Marketing team leverages volumes for media storage

IT support team would like additional durability on the data

Marketing team needs a low latency access to pull files for local workloads

Production team would like backups of these volumes for versions of their work

Demo

Provision a volume gateway instance on Amazon EC2

Configure a volume on the gateway instance

Connect to the volume

Perform manual backup and restoration of the data

Enable automated backup and deletion of backups

Summary

Summary

Reviewed volume gateway capabilities and configuration options

Connected to iSCSI volumes from the volume gateway

Performed a manual backup of a volume to an EBS snapshot

Created automated backups with volume snapshot schedule

Automated backup & scheduled backup deletion with AWS Backup

Implementing the Tape Gateway



David Tucker

TECHNICAL ARCHITECT & CLOUD CONSULTANT

@_davidtucker_ davidtucker.net

Overview

Reviewing tape gateway capabilities and configuration options

Provisioning a tape gateway instance in Amazon EC2

Connecting to the tape gateway from a client

Implementing a one time backup of data

Restoring data after being archived in Amazon S3 Glacier

Tape Gateway Deep Dive

Virtual Tape Library (VTL)

A VTL is a virtualized data storage solution that allows hard disk storage to integrate with existing tape based backup software as if it were a collection of tapes.

Virtualization on Tape Gateway

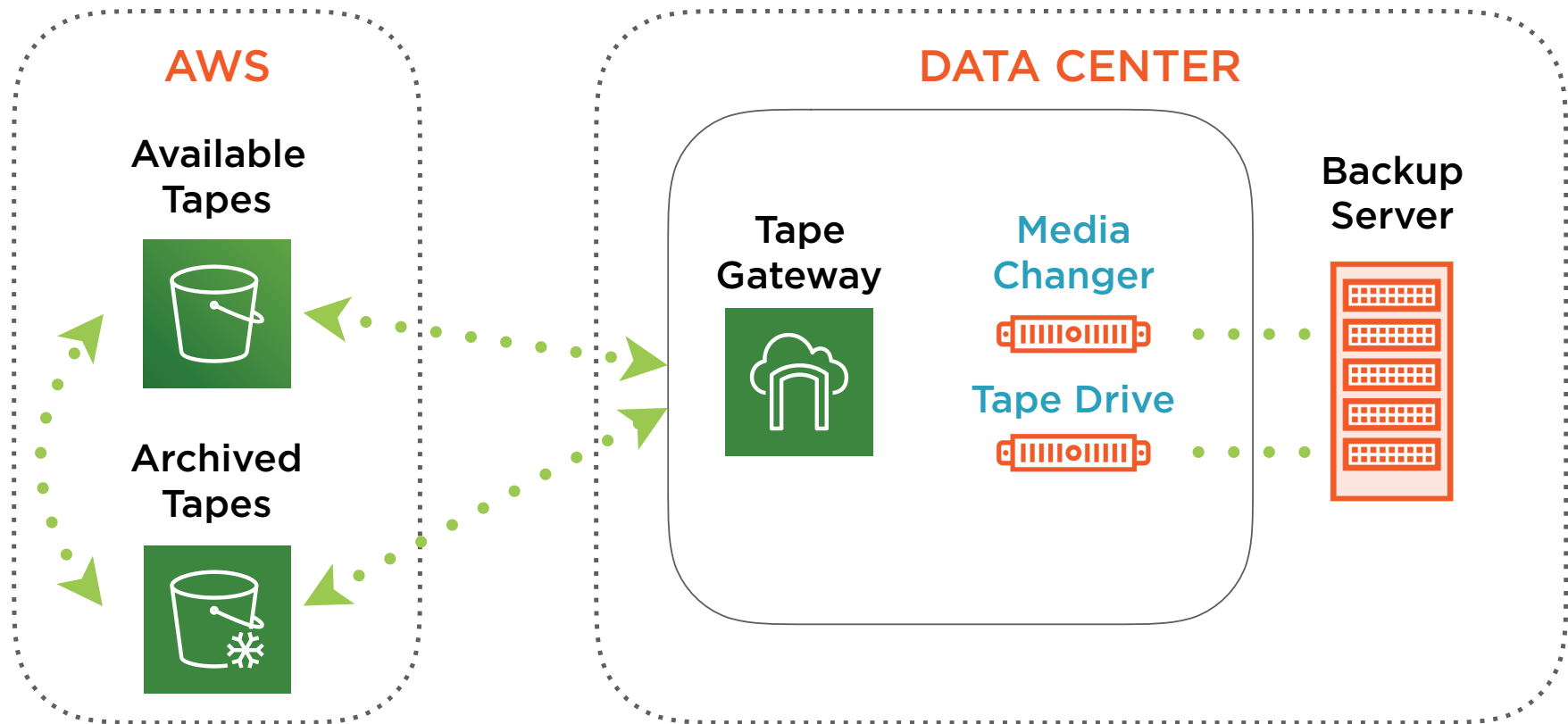
Media Changer

Different media changer virtualization types are supported such as STK-L700

Tape Drive

Gateway currently supports the IBM-ULT3580-TD5 device type through emulation

Tape Gateway Configuration



Tape Gateway Benefits

11 9's of durability

Industry standard encryption

Tiered storage

No physical transport risk

Amazon S3 Storage Classes

S3 Standard

S3 Glacier

S3 Glacier Deep Archive

Tape Gateway Limits

Tape Limit

Gateway can support
1,500 virtual tapes

Size Limit

Gateway supports total
volume up to 1PB

Archives

Gateway has no limits for
archived tapes

Archived Tape Retrieval

S3 Glacier

Archived tapes can be retrieved in
3-5 hours

S3 Glacier Deep Archive

Archived tapes can be retrieved in
12 hours

Supported Backup Applications

Arcserve Backup

Bacula Enterprise

Commvault

Dell EMC NetWorker

IBM Spectrum Connect

Micro Focus (HPE) Data Protector

**Microsoft System Center Data Protection
Manager**

Supported
Backup
Applications

NovaStor DataCenter

Quest NetVault Backup

Veeam Backup & Replication

Veritas Backup Exec

Veritas NetBackup

Tape Gateway Demo

Data Backup Strategy



GLOBOMANTICS

IT team currently leverages a traditional tape backup system

Current backup software is Veritas Backup Exec

IT team would like increased durability of backup data

Finance would like to reduce costs of backup hardware and offsite tape storage

Demo

Provision a tape gateway instance on Amazon EC2

Create a tape for the tape gateway

Connect to the tape gateway and leverage it as a VTL

Perform a backup of data to a virtual tape through the gateway

Archive a tape to Amazon S3 Glacier

Retrieve a tape and its data from Amazon S3 Glacier

Summary

Summary

Reviewed tape gateway capabilities and configuration options

Provisioned a tape gateway instance in Amazon EC2

Connected to the tape gateway from a client

Implemented a one time backup of data

Restored data after being archived in Amazon S3 Glacier

Scaling Your Storage Gateway



David Tucker

TECHNICAL ARCHITECT & CLOUD CONSULTANT

@_davidtucker_ davidtucker.net

Administering a Storage Gateway



Provisioning Config

Setting initial configuration for the storage gateway



Monitoring

Analyzing metrics and events provided by the gateway



Scaling

Adjusting configuration for real world performance

Overview

Analyzing performance configuration for EC2 gateways

Reviewing metrics & events provided by AWS Storage Gateway

Analyzing EC2 gateway scaling factors

Reviewing best practices around gateway monitoring

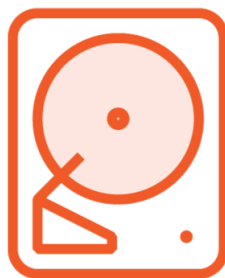
Performance for EC2 Gateways

Performance Considerations



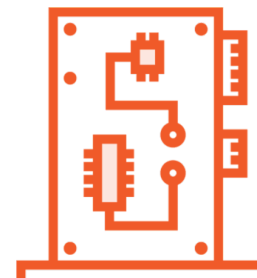
Instance Type

Not all instance types
can be leveraged for a
gateway instance



Storage

Storage is the most
critical aspect of
gateway performance



Networking

Efficient transfer
between gateway and
AWS is critical

Instance Types

File Gateway

AWS recommends c5.4xlarge or i3.4xlarge instance types

Volume & Tape Gateway

AWS recommends c5.4xlarge or c5d.9xlarge instance types

Storage Considerations

Cache configuration can include multiple disks for increased IOPS

Tape gateways should leverage 1MB block size if backup software supports it

Tape gateways should leverage at least 4 tapes for simultaneous parallel backups

For high performance, leverage NVMe SSD drives

Ephemeral storage may be leveraged for highest IOPS but higher risk

Ephemeral Storage for Gateway

Provides temporary block storage

Offers IOPS greater than EBS volumes

Can lead to data loss if cache is interrupted

Must follow process for restart to avoid data loss

Upload Buffer Sizing

$$\left(\begin{array}{c} \text{Application} \\ \text{Throughput} \\ \text{(MB/s)} \end{array} - \begin{array}{c} \text{Network} \\ \text{Throughput} \\ \text{to AWS (MB/s)} \end{array} \right) \times \begin{array}{c} \text{Compression} \\ \text{Factor} \end{array} \times \begin{array}{c} \text{Duration of} \\ \text{Writes} \\ \text{(s)} \end{array} = \begin{array}{c} \text{Upload} \\ \text{Buffer} \\ \text{(MB)} \end{array}$$

55 MB/s 10 MB/s 2 43,200 seconds (12 Hours) 1,512,000 MB (1.512 TB)

Cache Sizing

$$\begin{array}{ccccc} \text{Upload Buffer} & & \text{Cache} & & \text{Cache} \\ \text{Size} & \times & \text{Factor (1.1)} & = & \text{Size} \\ \\ 1.512 \text{ TB} & & 1.1 & & 1.6632 \text{ TB} \end{array}$$

Network Considerations

Instance Type

Ensure instance type supports needed throughput

Data Center

Verify the speed between gateway and application servers or workstations

Multiple Interfaces

Leverage multiple network interfaces for high performance

File Gateway Config 1

Instance type of c5.4xlarge

Root disk is 80 GB io1 with 4,000 IOPS

**Cache disk is 512 GiB EBS cache, io1,
1,500 provisioned IOPS**

Minimum network performance is 1 Gbps

This configuration can achieve 125 MB/s

File Gateway Config 2

Instance type of i3.4xlarge

Root disk is 80 GB io1 with 4,000 IOPS

Cache disk utilizes two 1.9 TB NVMe caches (ephemeral)

Minimum network performance is 5 Gbps

Configuration can achieve 500 MB/s

Tape Gateway Config 1

Instance type of c5.4xlarge

Root disk is 80 GB io1 with 4,000 IOPS

Cache disk utilizes 150 GB EBS volume

Upload buffer utilizes 150 GB EBS volume

Minimum network performance is 10 Gbps

Configuration can achieve 2.3 Gbps write and 3.2 Gbps read from cache

Configuration can achieve 1.2 Gbps write and 0.6 Gbps read from AWS

Tape Gateway Config 2

Instance type of c5d.9xlarge

Root disk is 80 GB io1 with 4,000 IOPS

Cache disk utilizes 450 GB NVMe volume

Upload buffer utilizes 450 GB NVMe volume

Minimum network performance is 10 Gbps

Configuration can achieve 2.7 Gbps write and 3.9 Gbps read from cache

Configuration can achieve 1.3 Gbps write and 0.7 Gbps read from AWS

Monitoring Your Gateway

Types of Monitoring Information

CloudWatch Metrics

Metrics around gateway performance are published to CloudWatch Metrics

CloudWatch Events

Events around upload completion and cache refresh are pushed to CloudWatch Events

Scope of Metrics

Gateway

Metrics apply to the gateway as a whole

Volume

Metrics apply to a single volume on a specific gateway instance

Gateway Metrics

Cache Metrics

These metrics do not apply to the Stored Mode for a Volume Gateway

CacheHitPercent

CachePercentUsed

CachePercentDirty

CacheFree

CacheUsed

Cloud Metrics

CloudBytesDownloaded

CloudBytesUploaded

CloudDownloadLatency

Upload Buffer Metrics

These metrics do not apply to the File Gateway.

UploadBufferPercentUsed

UploadBufferFree

UploadBufferUsed

Read & Write Metrics

QueuedWrites

ReadBytes

ReadTime

WriteTime

WriteBytes

TimeSinceLast
RecoveryPoint

Working Storage Metrics

These metrics only apply to the Stored Mode of the Volume Gateway.

WorkingStorageFree

WorkingStorageUsed

Volume Metrics

Cache Metrics

These metrics only apply to Volume Gateway volumes when used in the Cached Mode.

CacheHitPercent

CachePercentUsed

CachePercentDirty

Read & Write Metrics

QueuedWrites

ReadBytes

ReadTime

WriteTime

WriteBytes

Events

Gateway CloudWatch Events

NotifyWhenUploaded

Event fired when files have finished their async upload process to store

RefreshCache

Event fired when the cache has finished refreshing

Adding a Dashboard and Alarms

Demo

Configure an alarms based on cache volume metrics

Implement a dashboard to monitor the health of our gateway

Scale our gateway instance based on metrics

Scaling Your EC2 Gateway

Scaling Aspects for Storage Gateway

Instance Type

Network Bandwidth

Cache Size

Buffer Size

Storage Bandwidth

File Gateway Considerations

If CacheHitPercent remains low add additional cache volumes

If CloudBytesDownloaded remains high, consider increasing overall cache size

If NotifyWhenUploaded event is lagging, increase the bandwidth for your gateway

Volume Gateway (Cached) Considerations

If UploadBufferPercentUsed remains high, increase the overall buffer size

If CacheHitPercent remains low, increase the overall cache size

If QueuedWrites remains high, increase bandwidth to AWS

If overall throughput is less than desired, add additional Volume Gateway

Volume Gateway (Stored) Considerations

If UploadBufferPercentUsed remains high, increase the overall buffer size

If QueuedWrites remains high, increase bandwidth to AWS

If overall throughput is less than desired, add additional Volume Gateway

If WorkingStoragePercentUsed remains high, increase root volume size

Tape Gateway Considerations

If UploadBufferPercentUsed remains high, increase the overall buffer size

If CacheHitPercent remains low, increase the overall cache size

If QueuedWrites remains high, increase bandwidth to AWS

If throughput remains less than desired, leverage NVMe SSD drives

Summary

Summary

Analyzed performance configuration for EC2 gateways

Reviewed metrics & events provided by AWS Storage Gateway

Analyzed EC2 gateway scaling factors

Reviewed best practices around gateway monitoring