Backup and Restore to AWS

Working with APN Partners

August 2018

Notices

This document is provided for informational purposes only. It represents AWS's current product offerings and practices as of the date of issue of this document, which are subject to change without notice. Customers are responsible for making their own independent assessment of the information in this document and any use of AWS's products or services, each of which is provided "as is" without warranty of any kind, whether express or implied. This document does not create any warranties, representations, contractual commitments, conditions or assurances from AWS, its affiliates, suppliers or licensors. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

Contents

Abstract

Today, many storage and backup administrators are looking for ways to extend their backup environments to Amazon Web Services (AWS). This paper outlines options for utilizing existing or leveraging new partner solutions to extend or fully migrate backup environments to AWS, as well as protect workloads running on AWS with partner solutions.

Introduction

Data is continuing to grow, which is driving the need to reconsider traditional backup environments. Storage administrators, backup administrators, and IT organizations are looking for the ability to extend data center backups to AWS and are looking to integrate backup solutions to help protect workloads running on AWS.

This whitepaper will explore various partner-based backup solutions and how they integrate with AWS. For information about AWS backup strategies, see the AWS Backup and Restore Whitepaper.

What is Backup?

Backup and Restore solutions protect data from physical or logical errors, such as system failure, application error, or accidental deletion. Backup involves storing point-in-time copies of data. This data is often indexed to allow searching to find specific content, which can be at a granular level such as a virtual machine (VM) or a particular file.

Every backup solution is a slightly different, but many include similar components. The following are logical components of many popular backup software offerings. Sometimes these components are on a single server or appliance, and sometimes they can be distributed and scaled individually. Components may go by different names in each solution but maintain the same basic functions.

- **Catalog/Database** – The catalog or database generally holds the details of what has been backed up and where it is stored. It often also holds information like backup schedules, client, and server configuration.
- **Master Server** – The master server generally controls the backup environment. It is the main server and often hosts the backup database.
- **Media/Storage Server** – The Media or Storage server generally is responsible for connecting to the storage media disk, tape or object storage that stores the backup data.
- **Agent/Client** – The clients are the individual servers, storage, endpoints, and applications that are being backed up.
- **Proxy** – Some backup applications include proxies for accessing specific types of platforms, such as VMWare.

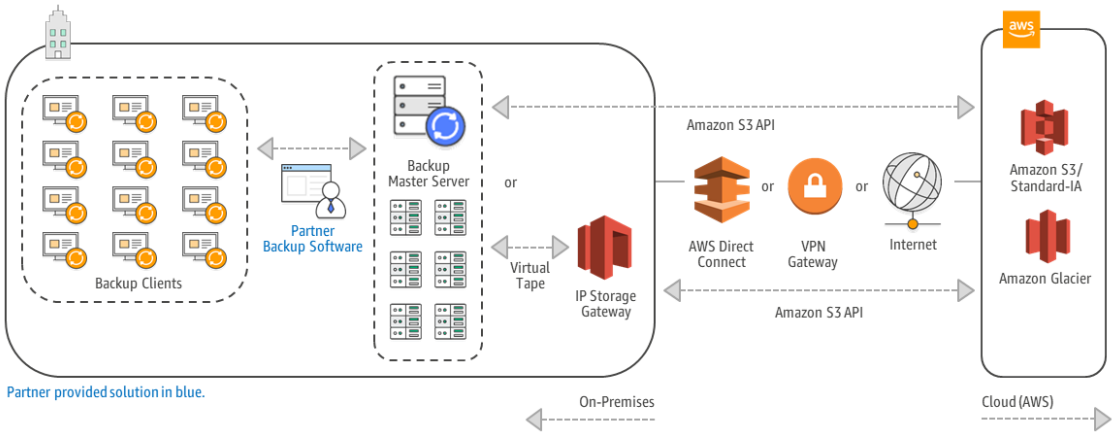

Figure 1: Backup to AWS using AWS Partner Network solutions

Traditional Backup

A traditional on-premises backup environment consists of a backup master and/or media server that typically points to some type of disk storage as a primary backup target. Due to its cost profile, disk storage is generally only used for short-term retention. Secondary copies often are stored on tape storage for longer term retention. Depending on the business requirements the ratio of disk to tape can vary. These storage tiers are usually in a single datacenter, which is the same datacenter that hosts the primary data. Since the entire environment may reside in a single datacenter, many customers have a requirement to store a copy of the data in an offsite location. Due to the offsite requirement, customers who don't have a second datacenter often send copies of their tape to a tape storage provider.

Hybrid Backup

When customers begin to use AWS, backup workloads are often the first workloads customers move to the AWS Cloud. These customers also often want to extend their current on-premises backup solutions to AWS. Each Backup and Restore AWS Partner Network (APN) technology partner offers different methods to connect to AWS Cloud storage. The details of some of APN's Backup and Restore partners are below. In general, these backup solutions run in part or wholly on-premises. The software points to Amazon Simple Storage Service (Amazon S3) and/or Amazon Glacier to either tier backup data, create a copy of backups, or act as the primary storage for backups.

Cloud Backup

As customers start moving their workloads to the AWS Cloud or launch new applications on AWS Cloud, they often turn to APN partner solutions to protect these workloads. To support this, many APN partner solutions can run on Amazon Elastic Compute Cloud (Amazon EC2). These backup solutions often work in very similar ways as they do on-premises and can allow

customers to manage backups for their AWS workloads the same way they manage their on-premises environment.

Backup versus Replication

For many customers with large on-premises storage systems, replication can be a means of providing an offsite copy of data. Replication can be combined with snapshots on both the source and target array to provide point-in-time restores for data. This type of backup often has limitations, such as requiring the same storage system on both the source and target side and does not including granular indexing. This type of solution is often used for disaster recovery purposes, and is therefore out of scope for this document.

Cloud Connectors

Many APN partner Backup and Restore solutions include connecters for directly reading/writing to AWS storage, such as Amazon S3 or Amazon Glacier. These connectors can be used with either existing on-premises installations or installations on Amazon EC2, where supported. Depending on the product, there are various levels of support, including tiering data, cloning data, or using AWS storage as a main data repository.

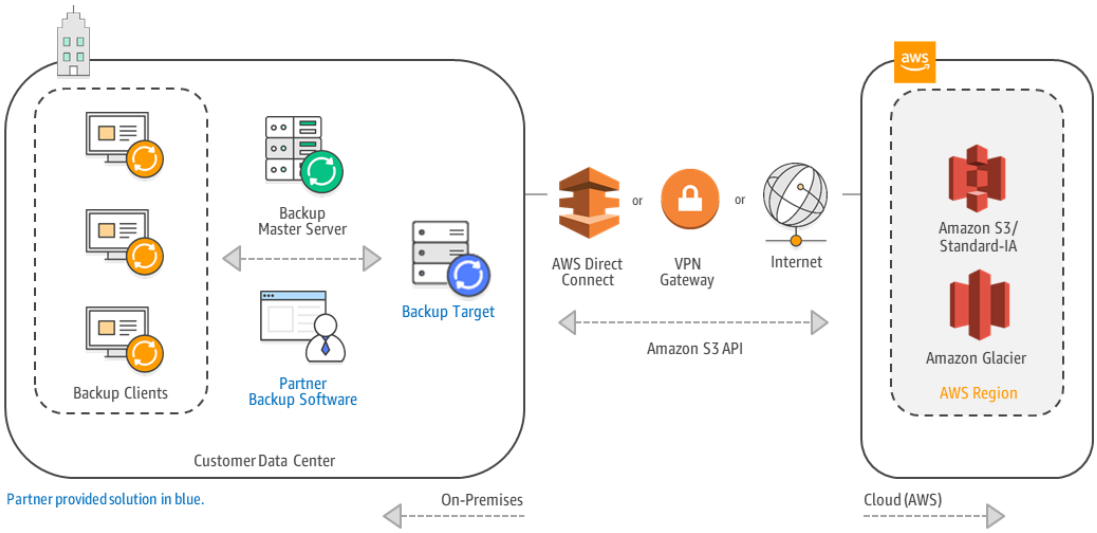

Figure 2: Back up to AWS from on-premises using cloud connectors

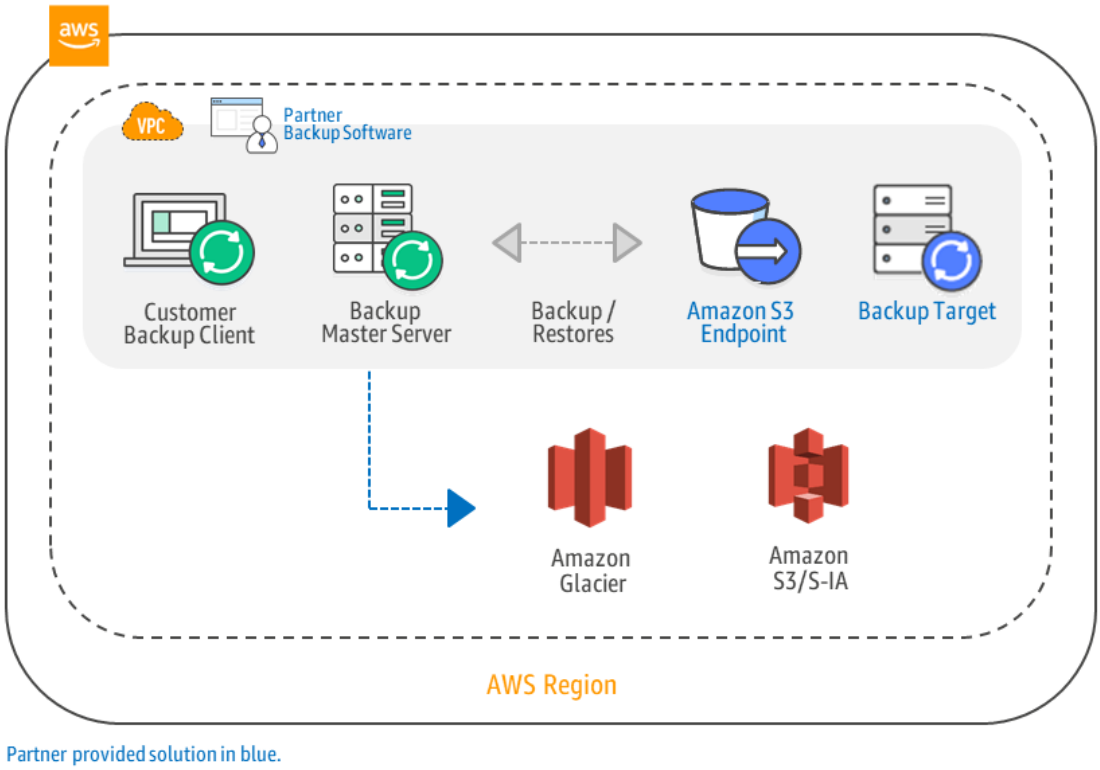

Figure 3: Back up of Amazon EC2 instances using cloud connectors

Arcserve UDP

Arcserve Unified Data Platform (UDP) supports backup to Amazon S3 directly from on-premises as well as running the UDP server on Amazon EC2. Arcserve UDP supports source-side global deduplication, encryption and compression and provides several deployment methods with AWS. These methods include backing up to Amazon S3, copying backups and individual files to Amazon S3, running a server on Amazon EC2, and replicating data between a server running on-premises and one running on Amazon EC2. Arcserve also supports a function called Instant Virtual Machine, which allows you to quickly create an Amazon EC2 instance from a backup stored on Amazon S3. Additional information can be found in the Arcserve deployment guide[1] and the Arcserve solutions guide[2].

CloudBerry Backup

CloudBerry Backup (CBB) supports backing up to Amazon S3-Standard, Standard-Infrequent Access (S3-IA) and Amazon Glacier. CloudBerry Backup can be configured to support Amazon S3 transfer acceleration. It also supports using Amazon S3 lifecycle policies, which can be managed in the CloudBerry client to support transitioning between Amazon S3 and Amazon Glacier. CloudBerry Backup also can back up directly to Amazon Glacier.

Cloudberry Backup supports encryption, compression, and deduplication. It also has a wide range of support for various clients and types of backups like image based, block level, application aware, and network shares. CloudBerry operates on a per-client basis, with the clients directly talking to Amazon S3 to store the backups. For information on setting up CloudBerry Backup, see the installation and configuration guide.[3]

Commvault

Commvault's architecture consists of a Commserve server and media agents. Media agents connect directly to Amazon S3, Amazon Glacier, and Glacier Vault Lock for write once read many (WORM). Commvault has integrations with AWS Snowball Edge and Snowball for off-line sync to cloud. You can orchestrate snapshots, backup and restore from Amazon Elastic Block Store (Amazon EBS) snapshots, deduplicate, compress, and encrypt data both in transit and at rest.

Commvault can be deployed both on-premises and on AWS using Amazon EC2 instances for all components. For more information, visit the Commvault AWS microsite[4].

Dell EMC NetWorker

For Dell EMC NetWorker® to use AWS storage, there is an appliance called CloudBoost. The CloudBoost appliance acts as a global deduplication engine. There is CloudBoost client built into the NetWorker client in current versions. The clients are able to directly handle encryption, deduplication, compression and upload to object storage the net new bits. With this setup, the CloudBoost server only handles metadata operations so it can add additional clients without having to scale significantly.

NetWorker also supports cloning backups to a CloudBoost in which case backups on the clients would go to a NetWorker storage node and backups would be cloned and deduped on CloudBoost appliance and sent to the Amazon S3 storage from the appliance. In this configuration, instead of each client sending to Amazon S3, the customer can have that filtered through the appliance to control bandwidth and be able to direct network routes for the specific IP, which some customers use in conjunction with AWS Direct Connect. For more information see the CloudBoost integration guide.[5]

IBM Spectrum Protect

IBM Spectrum Protect, formerly known as Tivoli Storage Manager (TSM), supports three main deployment patterns with AWS.

The first deployment pattern involves an IBM Spectrum Protect server that is installed on premises or on an Amazon EC2 instance, with primary backup and archive data landing on Amazon S3. This pattern could involve use of a direct-to-cloud architecture with accelerator

cache or a small disk container pool with immediate tiering to a second cloud-container storage pool without accelerator cache.

The second deployment pattern would make use of AWS as the secondary site. Much like the first deployment pattern, here the IBM Spectrum Protect server at the secondary site could make use of a direct-to-cloud topology with a cloud pool featuring accelerator cache, or it could use a small disk container pool landing spot with immediate tiering to a cloud pool backed by object storage.

The third deployment pattern features specific use of disk-to-cloud tiering, available with IBM Spectrum Protect V8.1.3 and later, to allow for operational recovery data to reside on faster performing disk storage. Data that is older, archived, or both would be tiered to cloud-based object storage after a specified number of days. This deployment also could be performed at an on-premises site or within a cloud compute instance. However, the additional cost of having a larger capacity disk container pool should be factored into cost estimates with an in-the-cloud solution.

For more information on cloud-container see the IBM wiki.[6]

N2W Software Cloud Protection Manager

N2W Software (N2WS) Cloud Protection Manager (CPM) runs on AWS and supports backing up Amazon EC2 instances, Amazon Relational Database Service (Amazon RDS) instances, Amazon Redshift clusters and Amazon DynamoDB tables. N2WS is now a Veeam company. CPM can back up instances to Amazon S3 in the same AWS region or in other regions. CPM is available on the AWS Marketplace.[7] For more information, visit the N2WS AWS backup site.[8]

Rubrik Cloud Data Management

Rubrik is most commonly deployed as an on-site appliance, which can back up data locally and then move data, by policy, to Amazon S3, Amazon Glacier including support for Vault Lock (one of two APN partners in this document that support WORM on AWS). Customers can configure policies that lifecycle the data to cloud storage as it ages out. Rubrik has client support for Microsoft SQL, Oracle, Microsoft Windows, Linux, Network Attached Storage (NAS) Shares and also virtual infrastructure support for VMware, Microsoft Hyper-V and Nutanix AHV. Rubrik can also be run as an Amazon EC2 instance, which can used to restore workloads backed up from an on-premises instance in an Amazon S3 bucket to AmazonEC2. For more information, see the Rubrik AWS solution brief.[9]

Rubrik Datos IO RecoverX

Rubrik Datos IO RecoverX is a scale-out, elastic, software-only data management platform that runs on-premises or natively on AWS and delivers scalable and fully featured point-in-time backup and restore. RecoverX also provides data mobility to, from, and within AWS public

cloud for traditional applications and cloud-native applications. RecoverX can create application-consistent backups of databases running either on-premises or on Amazon EC2 and store the backups in Amazon S3.

Veritas Backup Exec

Backup Exec is the Veritas solution for small and midsize businesses (SMB) and mid-market customers who are looking for a compelling backup solution that can span across the customers' diverse infrastructure requirements. Backup Exec has three main integration methods, using Amazon S3 as a storage target directly, using AWS Storage Gateway, and deploying on AWS to protect workloads running on AWS, as seen in Figure 9. Veritas Backup Exec is available in the AWS Marketplace[10]. For more information, visit the Veritas Backup Exec AWS microsite.[11]

Veritas NetBackup

Veritas NetBackup includes several options for integrating with AWS services. All components of the NetBackup solution, which include a master server and media server(s), can run on Amazon EC2.

Media server(s) that run on Amazon EC2 can store deduplicated backups on block storage, which is known as Media Server Deduplication Pool (MSDP). On AWS, the block storage would be Amazon EBS volumes attached to the Amazon EC2 instance.

Media servers also can be configured with a cloud connector. This enables the servers to directly store data onto Amazon S3 or Amazon Glacier. Data stored with the cloud connector is compressed. Amazon Glacier is supported via the use of a zero-day lifecycle policy[12].

Lastly, NetBackup CloudCatalyst can be used as a gateway between the media server and Amazon S3. When NetBackup CloudCatalyst is deployed, it handles deduplication of the data before it is sent to Amazon S3. At this time, NetBackup CloudCatalyst only supports Amazon S3 storage and does not support Amazon Glacier. NetBackup CloudCatalyst can be deployed as a physical or virtual appliance on-premises, not only reducing the amount stored in Amazon S3 but also reducing the amount of data sent over the wire. NetBackup CloudCatalyst also can be deployed on an Amazon EC2 instance.

Storage Gateways

Storage Gateways often are used in conjunction with backup software. Gateways can provide specialized functionality like protocol conversion, compression, deduplication, and caching. Different gateways support different front-end and back-end protocols and may offer just some or all of the aforementioned features.

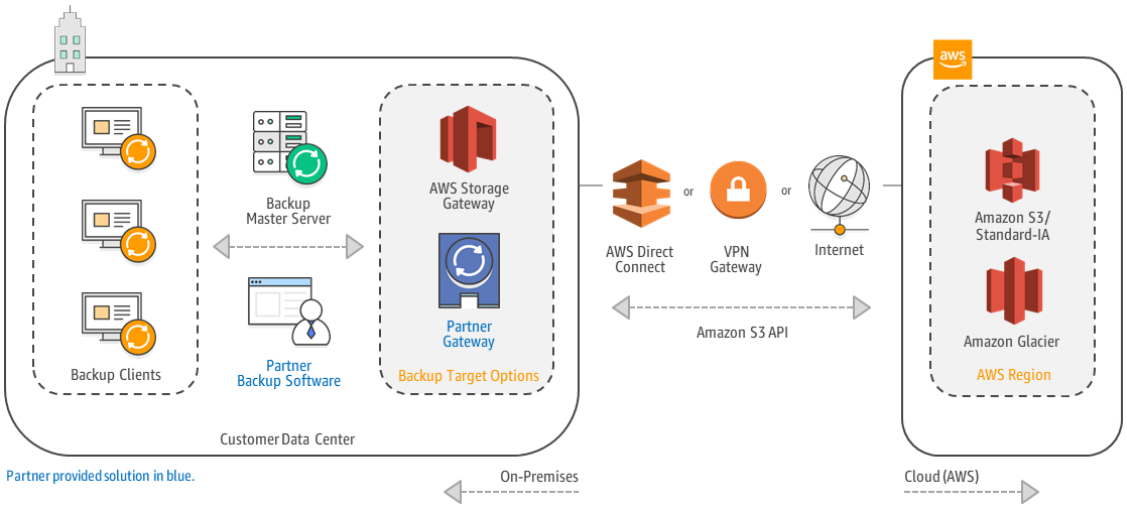

Figure 4: Back up to AWS using storage gateways

AWS Storage Gateway

AWS Storage Gateway is a hybrid storage service that enables your on-premises applications to seamlessly use AWS cloud storage. You can use the service for backup and archiving, disaster recovery, cloud bursting, storage tiering, and migration. Your applications connect to the service through a gateway appliance using standard storage protocols, such as NFS and iSCSI. The gateway connects to AWS storage services, such as Amazon S3, Amazon Glacier, and Amazon EBS, providing storage for files, volumes, and virtual tapes in AWS. The service includes a highly-optimized data transfer mechanism, with bandwidth management, automated network resilience, and efficient data transfer, along with a local cache for low-latency on-premises access to your most active data. For more information check the AWS storage gateway page.[13]

Figure 5: AWS Storage Gateway Options

The below backup applications are currently supported with Storage Gateway Virtual Tape Library (VTL) Mode. For the latest list check the VTL requirements document on the AWS site.[14]

Backup Application	Medium Changer Type
Arcserve Backup	AWS-Gateway-VTL
Commvault V11	STK-L700
Dell EMC NetWorker V8.x or V9.x	AWS-Gateway-VTL
Micro Focus (HPE) Data Protector 9.x	AWS-Gateway-VTL
Microsoft System Center 2012 R2 Data Protection Manager ***Note:*** *Data Protection Manager doesn't display barcodes for virtual tapes created in AWS Storage Gateway.*	STK-L700
NovaStor DataCenter/Network 6.4 or 7.1	STK-L700
Quest NetVault Backup 10.0	STK-L700
Veritas Backup Exec 2012	STK-L700
Veritas Backup Exec 2014	AWS-Gateway-VTL
Veritas Backup Exec 15	AWS-Gateway-VTL
Veritas Backup Exec 16	AWS-Gateway-VTL
Veritas NetBackup Version 7.x	AWS-Gateway-VTL
Veeam Backup & Replication V7	STK-L700
Veeam Backup & Replication V8	STK-L700
Veeam Backup & Replication V9 Update 2 or later	AWS-Gateway-VTL

Dell EMC Data Domain

Dell EMC Data Domain is an appliance that can be deployed physically or virtually. The virtual appliance is known as Data Domain Virtual Edition (DDVE) and has some different limitations from the physical appliance, such as maximum capacity. Data Domain supports multiple front-end protocols such as CIFS/NFS and its own DDboost protocol. Data Domain can integrate with and be supported by many popular backup software offerings.

Data Domain uses disk storage with deduplication. The primary storage for Data Domain is called the active tier. Data Domain also supports a secondary tier called cloud tier. The cloud tier is a separate deduplication domain from the active tier. There is a cache disk that is set up for the cloud tier that is separate than the disk storage used for the active tier. The data is either moved from the active tier to the cloud tier based on age of data or via an application policy from a supported backup application, such as Dell EMC NetWorker. Data Domain supports uploading to Amazon S3 and supports up to two times the size of the data in the active tier.

DDVE can run on-premises in a virtual environment or on AWS as an Amazon EC2 instance. When running on-premises, cloud tier is supported similar to how it is on the physical appliance. The version of DDVE that runs on AWS does not currently support cloud tier but supports using Amazon S3 for the active tier.

NetApp AltaVault

NetApp AltaVault is a storage gateway that can integrate with many popular backup applications. AltaVault supports protocols like CIFS and NFS on the front end and can tier data to either Amazon S3 or Amazon Glacier. AltaVault supports a large cache for customers that need more active data stored on-premises. Data on AltaVault and stored on object storage is deduplicated globally. For more information, see the Netapp Altavault site.[15]

StorReduce

StorReduce is a storage gateway specifically designed to sit in front of Amazon S3 and handle deduplication. StorReduce, unlike some other gateways, is not designed to do protocol conversion but instead provides S3 protocol on both the front end and the back end. StorReduce is mainly designed to sit in front of Amazon S3 and handle global deduplication, often providing greater deduplication levels then some backup software can achieve with its own deduplication algorithms. StorReduce can be used with nearly any backup software that supports Amazon S3.

StorReduce can be deployed as an on-premises appliance or on AWS. StoreReduce is available in the AWS Marketplace[16] and also has an AWS Quick Start.[17]

Backup as a Service

Backup as a Service is a Software as a Service (SaaS) offering that backs up a myriad of clients and devices without the need to manage any server or storage infrastructure as part of the backup environment. Agents are installed on the clients, but all other components of the backup environment run in the partner's AWS account.

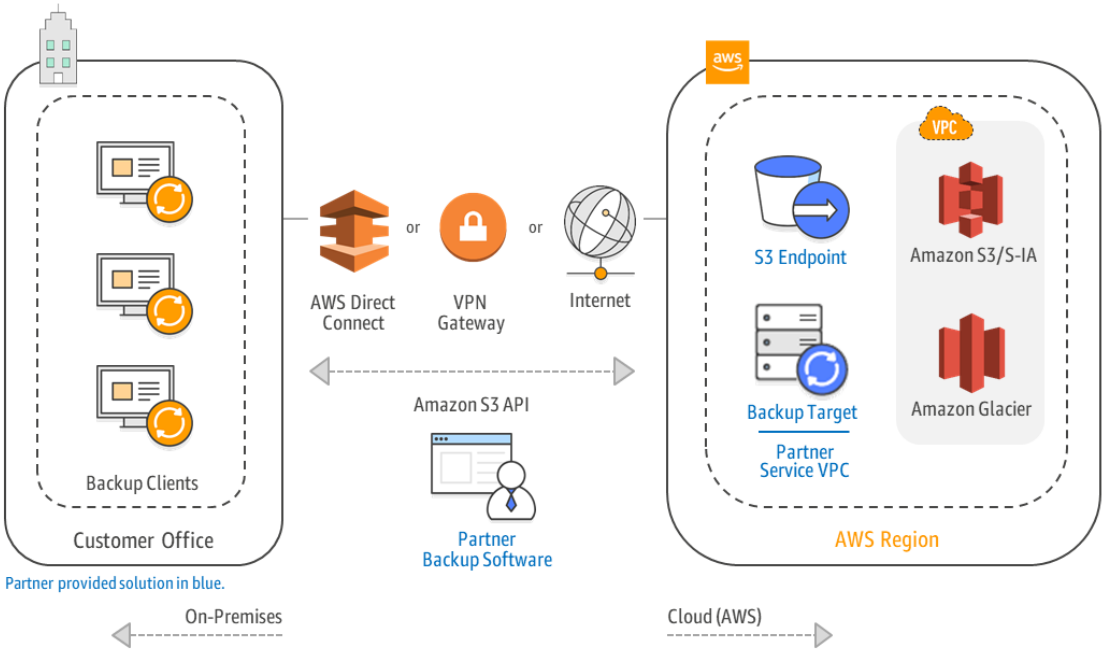

Figure 6: Software as a Service backup model

Druva inSync

Druva inSync provides a single pane of glass for protecting, preserving and discovering information across endpoints and cloud applications. Druva provides global deduplication and utility pricing where customers pay for data usage post deduplication. More information about Druva inSync can be found on the Druva inSync site.[18]

Druva Phoenix

Druva Phoenix is designed to back up physical and virtual servers. In addition to backing up to Amazon S3, it supports moving archival and long-term retention backups to Amazon Glacier. More information about Druva Phoenix can be found on the Druva Phoenix site.[19]

Conclusion

There are many options that you can use to back up your on-premises workloads to AWS or to protect your workloads running on AWS. Almost any major backup software has some option to store backups on AWS storage. It is important to understand the options that each backup software supports and how the software integrates other partner solutions like gateways to create a comprehensive solution that meets virtually any backup requirement.

Contributors

The following individuals and organizations contributed to this document:

- Henry Axelrod, partner solutions architect, Amazon Web Services
- Peter Kisich, partner solutions architect, Amazon Web Services
- Doug Cliche, partner solution architect, Amazon Web Services

Further Reading

For additional information, see the following:

- AWS Whitepapers[20]
- Amazon Simple Storage Service [21]
- Backup and Recovery Approaches using AWS[22]
- Backup and Recovery Partner Solutions [23]

Document Revisions

Date	Description
June 2018	First publication

Notes

[1] https://s13937.pcdn.co/wp-content/uploads/2018/02/Arcserve-UDP-On-AWS-Cloud-v3.1_Final.pdf

[2] https://s13937.pcdn.co/wp-content/uploads/2017/03/Arcserve-Solutions-for-AWS.pdf

[3] https://www.cloudberrylab.com/download/Backup_Installation_and_Configuration_Guide.pdf

[4] https://www.commvault.com/solutions/by-technology/virtual-machine-and-cloud/amazon-web-services/backup-and-archive-to-the-cloud

[5] https://www.emc.com/collateral/TechnicalDocument/docu81525.pdf

[6] https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Tivoli%20Storage%20Manager/page/Cloud-container%20storage%20pools%20FAQs

[7] https://aws.amazon.com/marketplace/seller-profile?id=b1157be6-ad44-4ba2-9db1-b95a423fd270

[8] https://n2ws.com/product/aws-backup

[9] https://www.rubrik.com/wp-content/uploads/2018/02/Solution-Brief-Rubrik-and-Amazon-Web-Services-AWS.pdf

[10] https://aws.amazon.com/marketplace/pp/B075S3PKVR

[11] https://www.veritas.com/product/backup-and-recovery/backup-exec/amazon-web-services

[12] https://www.veritas.com/support/en_US/doc/58500769-127471507-0/v126612396-127471507

[13] https://aws.amazon.com/storagegateway/

[14] https://docs.aws.amazon.com/storagegateway/latest/userguide/Requirements.html#requirements-backup-sw-for-vtl

[15] https://cloud.netapp.com/altavault

[16] https://aws.amazon.com/marketplace/seller-profile?id=3dc14b6c-c05e-4d38-9de3-3ae3641df4d8

[17] https://aws.amazon.com/quickstart/architecture/storreduce/

[18] https://www.druva.com/products/insync/

[19] https://www.druva.com/products/phoenix/

[20] https://aws.amazon.com/whitepapers/

[21] https://aws.amazon.com/documentation/s3/

[22] https://d1.awsstatic.com/whitepapers/Storage/Backup_and_Recovery_Approaches_Using_AWS.pdf

[23] https://aws.amazon.com/backup-recovery/featured-partner-solutions/