Backup, Archive, and Restore Approaches Using AWS

Pawan Agnihotri

AWS Certified Solutions Architect – Professional

Amazon Web Services

November 2014

Contents

Abstract

Over the past couple of years enterprise data has grown substantially, and the growth is accelerating. The need to protect the data has grown along with it. The increase in data also brings an increase in the complexity of methods used for the backing up the data. Questions such as durability and scalability of the backup solution are now commonplace. A common question is: How does cloud help my backup and archival needs?

This document aims to answer this question and propose some solutions around using cloud to back up your data. It discusses best practices of protecting your data using cloud services from AWS. This guide for backup, archive and restore will assist enterprise solution architects, backup architects, and IT administrators who are responsible for the design and deployment of the data protection for their corporate IT environments.

Introduction

As a backup architect or engineer, you are responsible for backups and archive for your enterprise. You have to manage the infrastructure as well as the backup operations. This may include managing tapes, sending tapes offsite, managing tape drives, managing backup servers, managing backup software, creating backup policies, insuring the backup data is secure, meeting compliance requirements for data retention, and performing restores. Furthermore, cost cutting puts pressure on your budgets and, with business open for more hours, your window to perform the backup is getting smaller.

These are some of the challenges that are faced by backup teams across many enterprises. The legacy environments are hard to scale, you need more tape and tape drives, and more storage capacity to back up the avalanche of data that the business is producing.

For those of you dealing with backups and restores, you may be employing many different systems, processes, and techniques available in the market. Additionally, you may have to support multiple configurations. With AWS, organizations can obtain a flexible, secure, and cost-effective IT infrastructure in much the same way that national electric grids enable homes and organizations to plug into a centrally managed, efficient, and cost-effective energy source. When freed from creating their own electricity, organizations were able to focus on the core competencies of their business and the needs of their customers. Please review some of the terms related to backup and archiving in the appendix which will be used throughout this whitepaper.

Why Use AWS

Amazon Web Services (AWS) is a secure, high-performance, flexible, cost-effective, and easy-to-use cloud computing platform. AWS takes care of the undifferentiated heavy lifting and provides the user with the necessary tools to accomplish the task of backing up the vast amounts of data from a variety of sources.

The first question asked by many customers is about security: Will my data be secure in the cloud? Amazon Web Services takes security very seriously; every service that we launch focuses on security as the foundation. Our storage services like Amazon Simple Storage Service[1] (Amazon S3) provide strong capabilities for access control and encryption both at rest and in transit. For encryption at rest, customers can use their own encryption keys[2] with the Amazon S3 server side giving them control over their data.

Switching to AWS offers many advantages:

- **Durability** – Amazon S3 and Amazon Glacier[3] are designed for 99.999999999% durability for the objects stored in them.
- **Security** – AWS provides a number of options for access control and encrypting data in transit and at rest.
- **Global Infrastructure** – Amazon Web Services are available across the globe so you can back up and store data in the region that meets your compliance requirement.
- **Compliance** –AWS infrastructure is designed and managed in alignment with regulations, standards and best-practices including (as of the date of this publication) SOC, SSAE 16, ISO 27001, PCI DSS, HIPPA, and FedRamp so you can easily fit the backup solution into your existing compliance regimen.
- **Scalability** – With AWS, you don't have to worry about capacity. You can scale your consumption up or down as your needs change.
- **Lower TCO** – The AWS scale of operations drives service costs down and helps lower the overall TCO of the storage. AWS often passes these cost savings on to the customer. As of the date of this publication, AWS has lowered prices 45 times since they began offering web services.

[1] http://aws.amazon.com/s3/

[2] http://aws.amazon.com/blogs/aws/s3-encryption-with-your-keys/

[3] http://aws.amazon.com/glacier/

Backup and Archive

Developing a comprehensive strategy for backing up and restoring data is not a simple task. In some industries, regulatory requirements for data security, privacy, and records retention can be important factors to consider when developing a backup strategy. A good backup process can be defined based on the objectives:

1. Backing up file data
2. Backing up database
3. Backing up machine images

In the following sections we describe the backup and archives approaches based on the organization of your infrastructure. IT infrastructure can broadly be categorized into the following scenarios - Cloud native, on premises, and hybrid.

Cloud Native

This scenario describes a workload environment that exists entirely on AWS. This includes web servers, application servers, databases, Active Directory, monitoring servers, etc. See Figure 1: AWS Native Scenario.

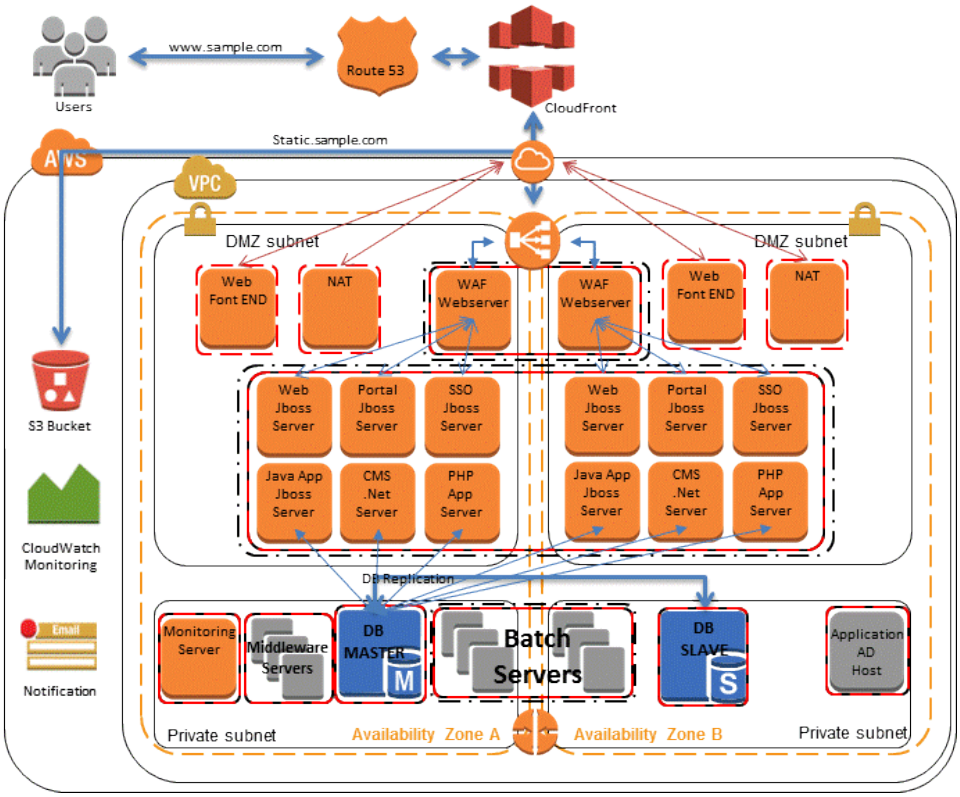

Figure 1: AWS Native Scenario

With all the services in AWS, you can leverage many of the built-in features to accomplish the backup-archive tasks.

Snapshot Options for Amazon EBS

In AWS "file data" can be stored on either Amazon S3 or Amazon Elastic Block Store[4] (Amazon EBS) volumes. Let's take a look at how you can backup data on these.

Amazon Elastic Compute Cloud[5] (Amazon EC2) can use Amazon EBS volumes to store block-based data. You can use this block storage for databases, or formatted into any OS-supported file system. Amazon EBS provides the ability to create snapshots (backups) of any Amazon EBS volume and write a copy of the data in the volume to Amazon S3, where it is stored redundantly in multiple Availability Zones. You can create the Amazon EBS snapshot by using the AWS Management Console, the command line interface (CLI), or the APIs. Using the Elastic Block Store Volumes page[6] of the Amazon EC2 console, click **Actions** and then click **Create Snapshot** to commence the creation of a snapshot that is stored in Amazon S3.

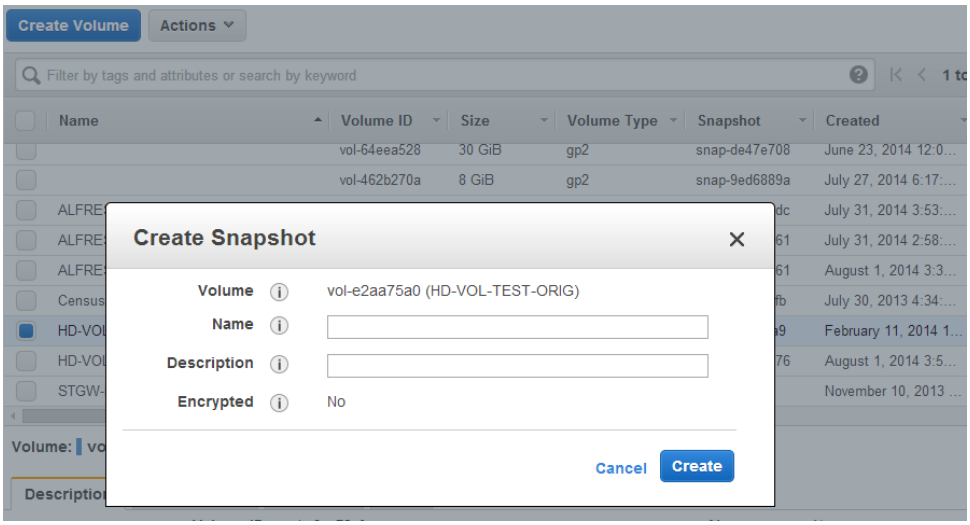

Figure 2: Creating a Snapshot from Amazon EBS Using the Console

You can also create the snapshot using the CLI command `ec2-create-snapshot`.

When you create a snapshot, you protect your data directly to durable disk-based storage. You can schedule and issue the commands on a regular basis. And due to the economical pricing of Amazon S3, you can retain many generations of data. Further,

[4] http://aws.amazon.com/ebs/

[5] http://aws.amazon.com/ec2/

[6] https://console.aws.amazon.com/ec2/v2/#Volumes

because snapshots are block-based, you consume space only for changed data after the initial snapshot is created.

To restore data from a snapshot, use AWS Management Console, the command line interface (CLI), or the APIs to create a new volume from an existing snapshot.

For example, to restore a volume to a prior point-in-time backup, you could use the following sequence:

1. Create a new volume from the backup snapshot by using the following command:

   ```
   > ec2-create-volume –z us-west-1b –snapshot MySnapshotName
   ```

2. Within the Amazon EC2 instance, unmount the existing volume (e.g., by using `umount` in Linux or the Logical Volume Manager in Windows).
3. Detach the existing volume from the instance by using the following command:

   ```
   > ec2-detach-volume OldVolume
   ```

4. Attach the new volume that was created from the snapshot by using the following command:

   ```
   > ec2-attach-volume VolumeID –I InstanceID –d Device
   ```

5. Remount the volume on the running instance.

This process is a fast and reliable way to restore full volume data as needed. If you need only a partial restore, you can attach the volume to the running instance under a different device name, mount it, and then use operating system copy commands to copy the data from the backup volume to the production volume.

Amazon EBS snapshots can also be copied between AWS regions using the Amazon EBS snapshot copy capability that is available from the console or command line, as explained in the *Amazon Elastic Compute Cloud User Guide*.[7] You can use this feature to store your backup in another region without having to manage the underlying replication technology.

Creating Consistent or Hot Backups

When you back up a system, the ideal is to have the system in a quiet state where it is not performing any I/O. From a backup perspective, the ideal state is a machine that is

[7] http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-copy-snapshot.html

accepting no traffic. But this ideal is increasingly rare as 24/7 IT operations become the norm.

Consequently, it is necessary to quiesce the file system or database in order to make a clean backup. How you do this depends on your database and/or file system, so due diligence is required. To summarize the process for a database:

- If possible, put the database into hot backup mode. Alternatively, create a read replica copy of the database; this is a copy of the database that is up to date but runs on a separate instance. Keep in mind that on AWS you can run this instance for the duration required to perform the backup and then close it down, saving resources.
- Issue the relevant Amazon EBS snapshot commands.
- Take the database out of hot backup mode or, if using a read replica, terminate the read replica instance.

Backing up a file system works similarly and depends highly on the capabilities of the particular operating system or file system. An example of a file system that can flush its data for a consistent backup is xfs (see `xfs_freeze`).[8] If the file system in question does not support the ability to freeze, you should unmount it, issue the snapshot command, and then remount the file system. Alternatively, you can facilitate this process by using a logical volume manager that supports freezing of I/O.

Because the snapshot process is fast to execute and captures a point in time, the volumes you are backing up only need be unmounted for a matter of seconds. This ensures that the backup window is as small as possible and that outage time is predictable and can be effectively scheduled. While the data copy process of creating the snapshot may take longer, the snapshot activity requiring the volume to be un-mounted is very quick. Don't confuse the two processes when structuring your backup regime.

Multivolume Backups

In some cases, you may stripe data across multiple Amazon EBS volumes by using a logical volume manager in order to increase potential throughput. When using a logical volume manager (e.g., mdadm or LVM), it is important to perform the backup from the volume manager layer rather than the underlying devices. This ensures all metadata is consistent and that the various subcomponent volumes are coherent. You can take a number of approaches to accomplish this, an example being the script created by alestic.com.[9]

[8] https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/xfsfreeze.html

[9] https://github.com/alestic/ec2-consistent-snapshot

You can also perform backups of this nature from the logical volume manager or file system level. In these cases, using a "traditional" backup agent enables the data to be backed up over the network. A number of agent-based backup solutions are available on the internet and the AWS Marketplace.[10] It is important to remember that agent-based backup software expects a consistent server name/IP address. As a result, using these tools in concert with instances deployed in a virtual private cloud (VPC)[11] is the best method to ensure reliability.

An alternative approach is to create a replica of the primary system volumes that exist on a single large volume. This simplifies the backup process, as only one large volume needs to be backed up, and the backup does not take place on the primary system. However, it is important to ascertain whether the single volume can perform sufficiently to maintain changes during the backup and whether the maximum volume size is appropriate for the application.

Backing Up Databases

AWS has many options for running databases. You can run your own database on an Amazon EC2 instance or use one of the managed services offering. If you are running your own database on an Amazon EC2 instance, you can back up data to files using database native tools (e.g., MySQL,[12] Oracle,[13, 14] MSSQL,[15] postgresSQL[16]) or create a snapshot of the volumes containing the data.

Backing up data for database differs from the web and application layers. In general, databases contain larger amounts of business data (tens of GB to multiple TB) in database-specific formats that must be retained and protected at all times. In these cases, you can leverage efficient data movement techniques such as snapshots to create backups that are fast, reliable, and space efficient.

For databases that are built upon RAID sets of Amazon EBS volumes (and have total storage less than 1 TB), an alternative backup approach is to asynchronously replicate data to another database instance built using a single Amazon EBS volume. While the destination Amazon EBS volume will have slower performance, it is not being used for data access, and you can easily send a snapshot to Amazon S3 using the Amazon EBS

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

[11] http://aws.amazon.com/vpc/

[12] http://dev.mysql.com/doc/refman/5.7/en/backup-and-recovery.html

[13] https://media.amazonwebservices.com/AWS_Amazon_Oracle_Backups.pdf

[14] http://docs.oracle.com/cd/E11882_01/backup.112/e10642/rcmbckba.htm#BRADV8003

[15] http://msdn.microsoft.com/en-us/library/ms187510.aspx

[16] http://www.postgresql.org/docs/9.3/static/backup.html

snapshot capability (see "Snapshot Options for Amazon EBS" earlier in this paper section).

Backups for Amazon Relational Database Service

The Amazon Relational Database Service (Amazon RDS)[17] includes automated backups. This means that you do not need to issue specific commands to create backups of your database.

Amazon RDS provides two different methods for backing up and restoring your DB Instance(s); automated backups and database snapshots (DB snapshots).

- **Automated backups** enable point-in-time recovery of your DB Instance. When automated backups are turned on for your DB instance, Amazon RDS automatically performs a full daily backup of your data (during your preferred backup window) and captures transaction logs (as updates to your DB instance are made).

  When you initiate a point-in-time recovery, transaction logs are applied to the most appropriate daily backup in order to restore your DB instance to the specific time you requested. Amazon RDS retains backups of a DB instance for a limited, user-specified period of time called the retention period, which, as of the date of this publication, by default is one day but can be set to up to thirty-five days.

  You can initiate a point-in-time restore and specify any second during your retention period, up to the Latest Restorable Time. You can use the `DescribeDBInstances` API call to return the latest restorable time for your DB instance(s), which is typically within the last five minutes. Alternatively, you can find the Latest Restorable Time for a DB instance by selecting it in the AWS Management Console and looking in the **Description** tab in the lower panel of the console.

- **DB snapshots** are user-initiated and enable you to back up your DB instance in a known state as frequently as you wish, and then restore to that specific state at any time. DB snapshots can be created with the AWS Management Console or by using the `CreateDBSnapshot` API call. The snapshots are kept until you explicitly delete them with the console or the `DeleteDBSnapshot` API call.

Note that when you restore to a point in time or from a DB snapshot, a new DB instance is created with a new endpoint. (If you want, you can delete the old DB instance by using the AWS Management Console or a `DeleteDBInstance` call.) You do this so you can create multiple DB instances from a specific DB snapshot or point in time.

[17] http://aws.amazon.com/rds/

Backup of the Amazon Machine Image (AMI)

Next we look at "machine images." AWS stores system images in what are called Amazon Machine Images or AMI for short. These images consist of the template for the root volume required to launch an instance. To save your instance's as a machine image you simply backup the root volume as an AMI.

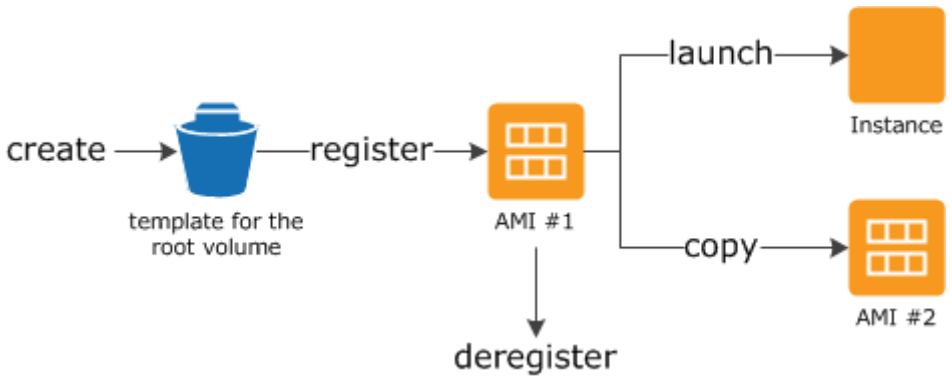

Figure 3: Using AMI to backup and launch an instance

An AMI that you register is automatically stored in your account using Amazon EBS snapshots. These snapshots reside in Amazon S3 and are highly durable. This means that the underlying storage mechanism for the AMIs is protected from multiple failure scenarios.

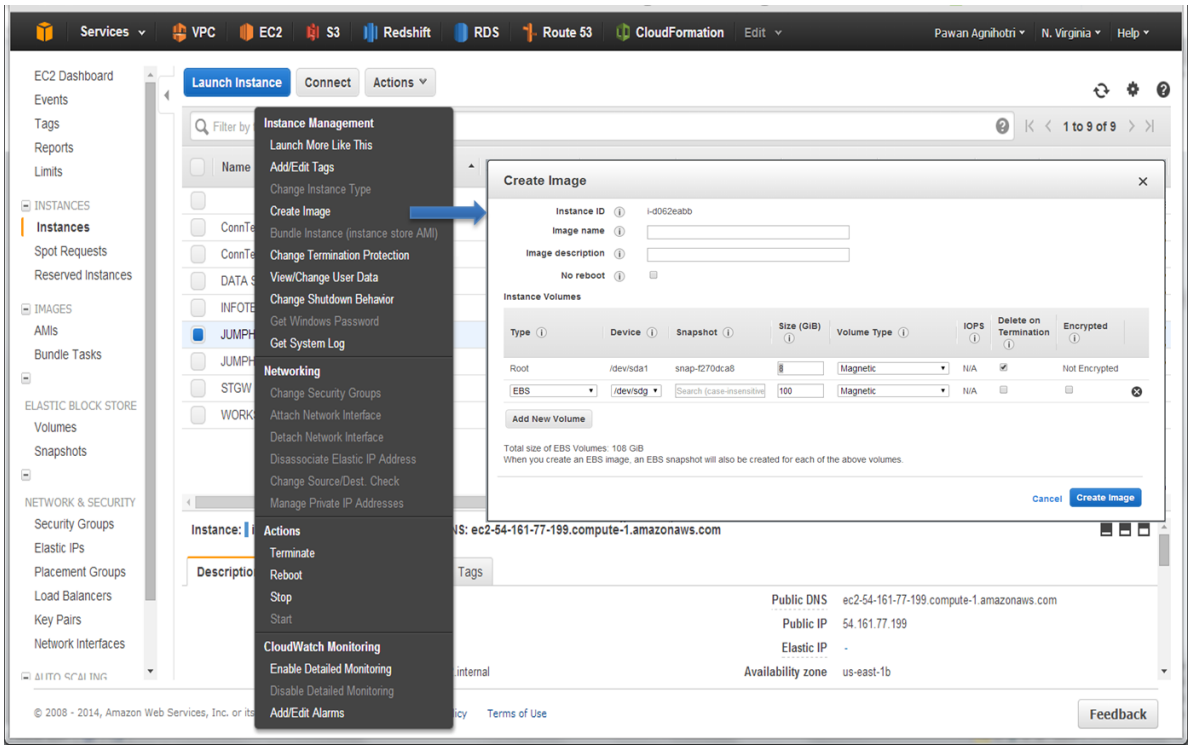

Figure 4: Using the EC2 console to create a machine image

Once you have created an AMI of your Amazon EC2 instance you can use the AMI to recreate the instance or launch more copies of the instance. It is also possible to copy AMIs from one region to another. Consequently, you can save a copy of a system image to another region.

On Premises

This scenario describes a workload environment with no component in the cloud. All resources, including web servers, application servers, databases, Active Directory, monitoring, and more, are hosted either in the customer data center or colocation. See the following figure.

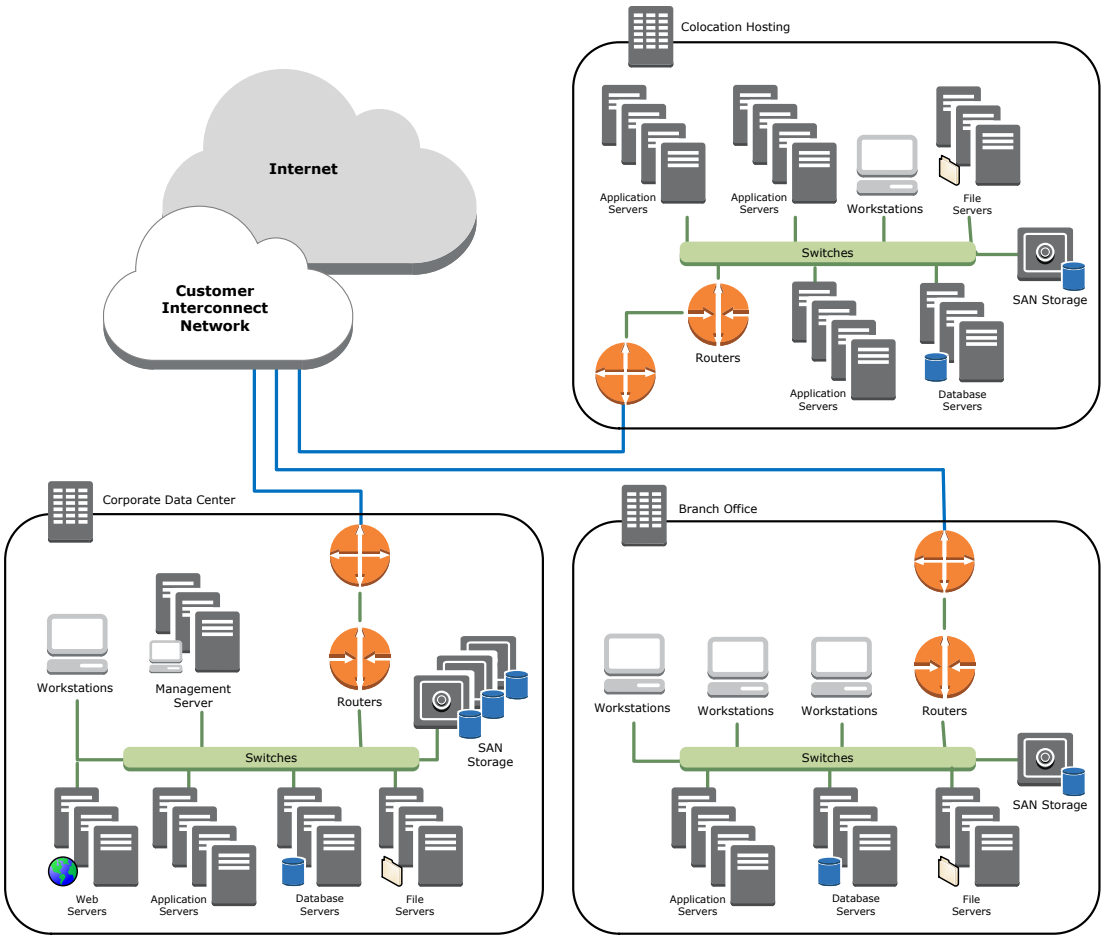

Figure 5: On-premises environment

AWS can be leveraged very nicely for this scenario to help with backup and archiving. Using AWS storage services lets you focus on the backup and archiving task, leaving the heavy lifting on the storage side to AWS. With AWS you do not have to worry about storage scaling or infrastructure capacity to accomplish the backup task.

Amazon storage services such as Amazon S3 and Amazon Glacier are natively API based and available via the Internet. This allows backup software vendors to directly integrate their applications with storage solutions provided by AWS as represented in the following figure. You can look at our partner directory[18] for the backup software vendors who work with AWS.

The primary solution in this scenario is to use backup and archive software that directly interfaces with AWS through the APIs. Here the backup software is AWS aware and will

[18] http://www.aws-partner-directory.com/PartnerDirectory/PartnerSearch?type=ISV

back up the data from the on premises servers directly to Amazon S3 or Amazon Glacier.

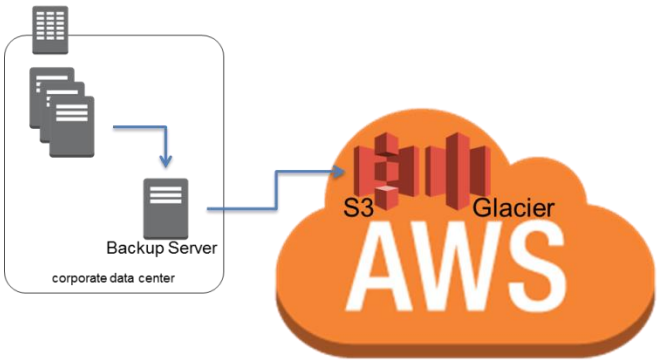

Figure 6: Backup connector to Amazon S3 or Amazon Glacier

If your existing backup software does not natively support the AWS cloud, the alternate solution is to use our storage gateway products. AWS Storage Gateway[19] is a virtual appliance that provides seamless and secure integration between your data center and AWS's storage infrastructure. The service allows you to securely store data in the AWS cloud for scalable and cost-effective storage. The AWS Storage Gateway supports industry-standard storage protocols that work with your existing applications while securely storing all of your data encrypted in Amazon S3 or Amazon Glacier.

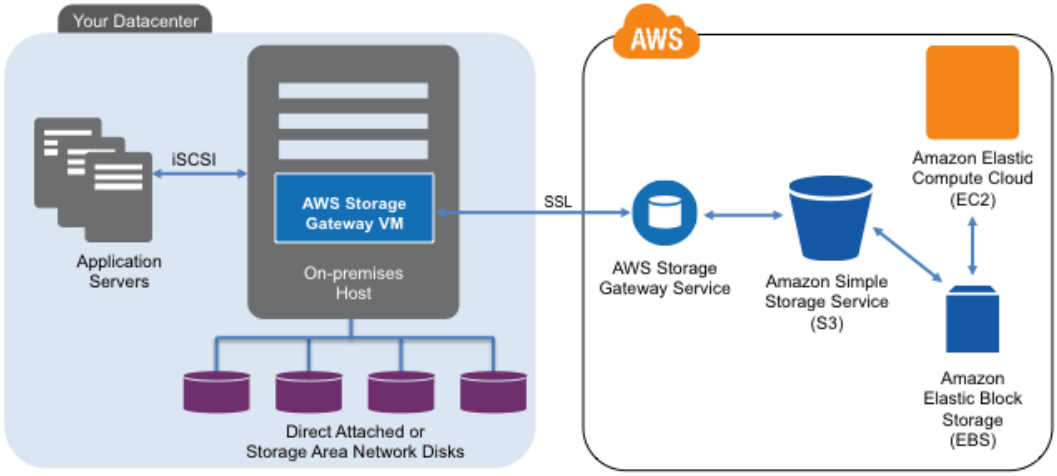

Figure 7: Connecting on-premises to AWS storage

[19] http://aws.amazon.com/storagegateway/

AWS Storage Gateway supports three configurations:

- **Gateway-cached volumes** – You can store your primary data in Amazon S3 and retain your frequently accessed data locally. Gateway-cached volumes provide substantial cost savings on primary storage, minimize the need to scale your storage on premises, and retain low-latency access to your frequently accessed data.
- **Gateway-stored volumes** – In the event you need low-latency access to your entire data set, you can configure your on-premises data gateway to store your primary data locally, and asynchronously back up point-in-time snapshots of this data to Amazon S3. Gateway-stored volumes provide durable and inexpensive off-site backups that you can recover locally or from Amazon EC2.
- **Gateway-virtual tape library (gateway-VTL)** – With gateway-VTL you can have a limitless collection of virtual tapes. Each virtual tape can be stored in a virtual tape library backed by Amazon S3 or a virtual tape shelf backed by Amazon Glacier. The virtual tape library exposes an industry standard iSCSI interface, which provides your backup application with online access to the virtual tapes. When you no longer require immediate or frequent access to data contained on a virtual tape, you can use your backup application to move it from its virtual tape library to your virtual tape shelf to further reduce your storage costs.

These gateways act as plug-and-play devices providing standard iSCSI devices, which can be integrated into your backup or archive framework. You can use the iSCSI disk devices as storage pools for your backup software or the gateway-VTL to offload tape-based backup or archive directly to Amazon S3 or Amazon Glacier.

Using this method your backup and archives are automatically offsite (for compliance purposes) and stored on durable media, eliminating the complexity and security risks of off-site tape management.

To offer flexibility to the customer, a multitude of third-party appliances and gateway devices work with Amazon storage services and can be found on our partner network website.[20]

Hybrid

The two infrastructure deployments addressed up to this point, “cloud native” and “on-premises,” can be combined into a hybrid scenario whose workload environment has infrastructure components in AWS as well as on premises. Resources, including web servers, application servers, databases, Active Directory, monitoring, and more, are

[20] http://www.aws-partner-directory.com/

hosted either in the customer data center or AWS. Applications running in the AWS are connected to applications running in the customer premises.

This scenario is emerging as a very common case for enterprise workloads. Many enterprises have data centers of their own while leveraging AWS to augment capacity. These customer data centers are often connected to the AWS network by high capacity network links. For example, with AWS Direct Connect[21] you can establish private dedicated connectivity from your premises to AWS.

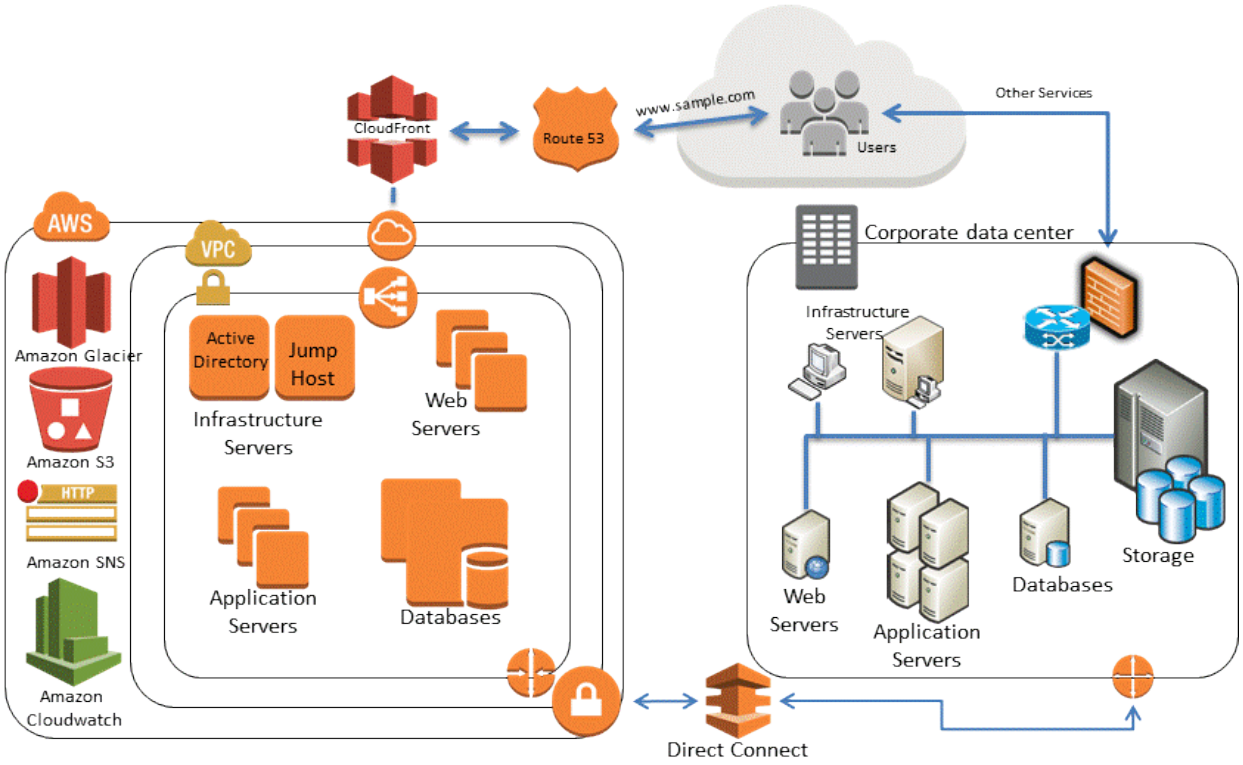

Figure 8: A hybrid infrastructure scenario

You can leverage AWS to help with backup and archiving for this scenario as well. The techniques you use are a combination of the methods described previously in cloud-native and on-premises solutions.

Hybrid Techniques

If you already have an existing framework that backs up data for your on-premises servers, then it is easy to extend that framework to your AWS resources over a VPN

[21] http://aws.amazon.com/directconnect/

connection or AWS Direct Connect. You will install the backup agent on the Amazon EC2 instances and back them up per the existing data protection policies.

Depending on your backup framework setup, you may have a master backup server along with one or more media servers. You may consider moving the master backup server to an Amazon EC2 instance to automatically protect your master backup server against on-premises disasters and have a highly available backup infrastructure.

To manage the backup data flows, you may also consider creating one or more media servers on Amazon EC2 instances. This will help the cloud-based resources backup to a local media target rather than go over the network back to the on-premises environment.

You can also leverage the AWS Storage Gateway or other third-party storage gateways from the AWS Marketplace to connect your backup framework to Amazon storage services. The storage gateways are connected to the media servers allowing data to be securely and durably stored on Amazon S3 or Amazon Glacier.

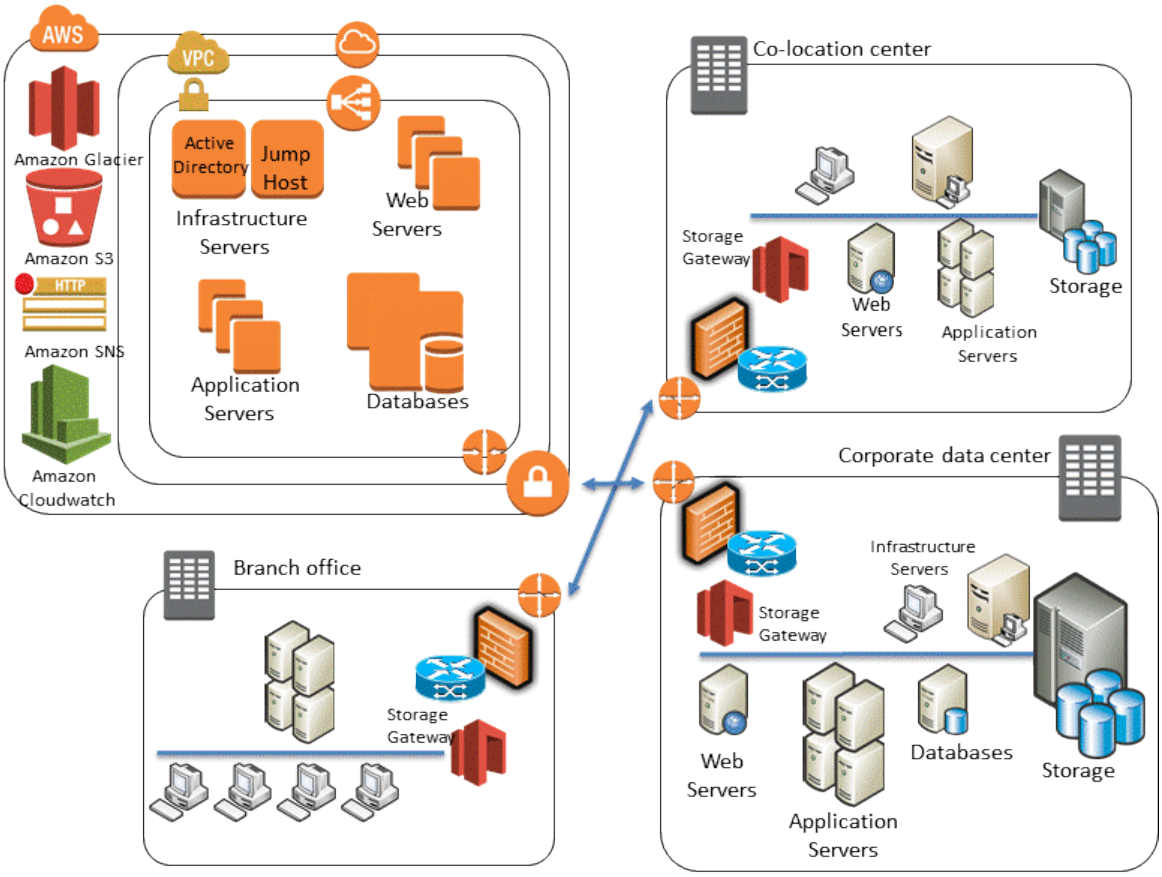

Figure 9: Leveraging gateways in the hybrid scenario

Use Cases

A use case can help explain the on-premises and hybrid scenarios: Assume that you are managing an environment where you are backing up a mixture of standalone servers, virtual machines, and database servers. This environment has 1,000 servers, and you backup operating system, file data, virtual machine images, and database backups. You have 20 databases to back up, which are a mixture of MySQL, MSSQL, and Oracle. You use “myqldump” to create a database dump file to disk for MySQL backups. Your backup software provides plugins or agents to back up operating system, virtual machine images, data, and MSSQL databases. Additionally this software has tight integration to backup Oracle database using RMAN.

To support the above environment, your backup software has a global catalogue server or master server that controls the backup, archive and restore activities as well as multiple media serves that are connected to disk-based storage and LTO tape drives.

Case 1: As the very first step, you check the vendor site to see if there is a plugin or built-in support for cloud storage backup and archive. If the software has cloud storage backup options, you can proceed to configure it. Many vendors support Amazon S3 as an option for cloud storage and Amazon Glacier for cloud archive. You can create the target bucket either from within the backup software or use the AWS Management Console to create a bucket in Amazon S3. Next you configure the media servers to create storage pools that use the Amazon S3 bucket. Once the storage pool is configured, the backup software starts using Amazon S3 to store the backup data.

Case 2: If your backup software does not natively support cloud storage for backup or archive, you can use a storage gateway device as a bridge between the backup software and Amazon S3 or Amazon Glacier. If you want to attach disk-based storage to your media server, you can download the gateway – cached volumes storage gateway. If you want to attach tape drives to your media server you can download the gateway–virtual tape library storage gateway. You can download the storage gateway from the AWS Management Console. Once the gateway is downloaded and activated, you can create iSCSI targets, which can be attached to the media servers. The media server sees these iSCSI targets as local disks or tape drives. You can then configure these into the storage pools and used for backups or archives.

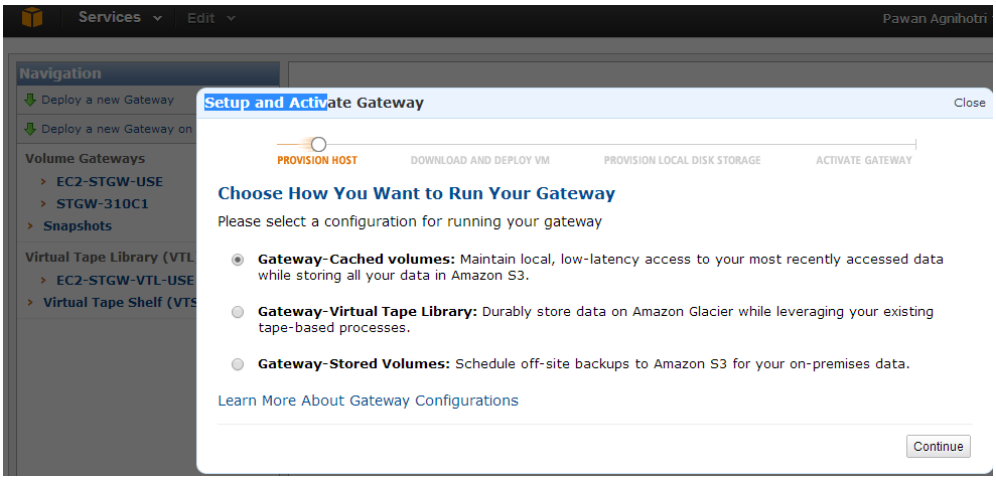

Figure 10: Choose and download the appropriate storage gateway

Cloud Paradigms

As cloud-based computing has evolved, so has the strategy for using it in backup and recovery.

Protecting Configurations Rather Than Servers

The Amazon EC2 service simplifies the backup and recovery of a standard server, such as a web server or application server. Traditionally, you would back up the complete server via a central backup server. With Amazon EC2 you can focus on protecting configuration and stateful data, rather than the server itself. This set of data is much smaller than the aggregate set of server data, which typically includes various application files, operating system files, temporary files, and so on. This change of approach means that regular nightly incremental or weekly full backups can take far less time and consume less storage space.

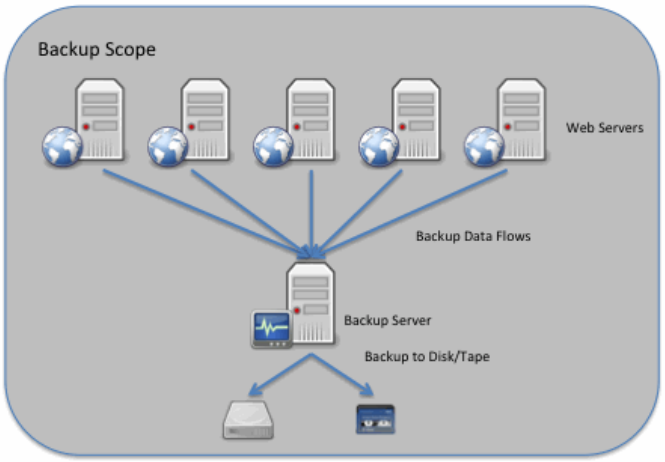

Figure 11: Traditional backup approach

When a compute instance is started in Amazon EC2, it is based upon an AMI and can also connect to existing storage volumes such as Amazon EBS. In addition, when launching a new instance, it is possible to pass "user data"[22] to the instance that can be accessed internally as dynamic configuration parameters.

A sample workflow is as follows:

- Launch a new instance of a web server, passing it the "identity" of the web server and any security credentials required for initial setup. The instance is based upon a prebuilt AMI that contains the operating system and relevant web-server application (e.g., Apache or IIS).
- Upon startup, a boot script accesses a designated and secured Amazon S3 bucket that contains the specified configuration file(s).

[22] http://docs.amazonwebservices.com/AWSEC2/latest/UserGuide/index.html?AESDG-chapter-instancedata.html

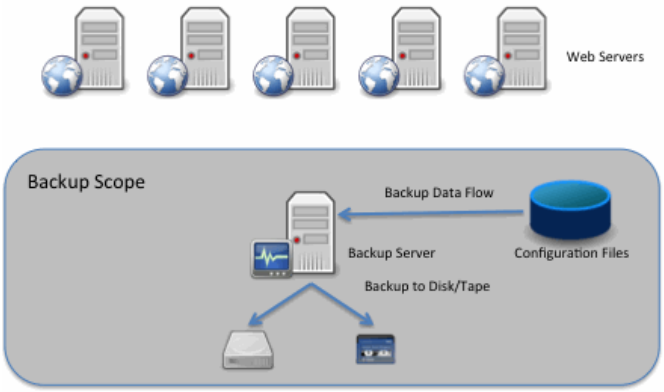

Figure 12: Amazon EC2 Backup Approach

- The configuration file contains various instructions for setting up the server (e.g., web server parameters, locations of related servers, additional software to install, and patch updates).
- The server executes the specified configuration and is ready for service. An open source tool for performing this process, called cloud-init,[23] is already installed on Amazon Linux AMIs and is also available for a number of other Linux distributions.

In this case, there is no need to back up the server itself. The relevant configuration is contained in the combination of the AMI and the configuration file(s). So the only components requiring backup and recovery are the AMI and configuration file(s).

Using Storage Fit for Purpose

In traditional storage model you have a choice of SAN or NAS based storage. Amazon Web Services provides many storage options. If your data and workflow need a file- or object-based store then Amazon S3 is the best solution. Amazon S3 is highly durable data store designed for 99.999999999% data durability. The data is internally replicated across multiple data centers. In addition Amazon S3 provides many features such as versioning. With versioning enabled, you can automatically save versions of objects that are overwritten.[24] This provides an automatic backup capability for the data stored in Amazon S3. Amazon S3 also offers data lifecycle management features so you can automatically archive or delete data when certain time criteria are met.

[23] https://launchpad.net/cloud-init

[24] http://docs.aws.amazon.com/AmazonS3/latest/dev/Versioning.html

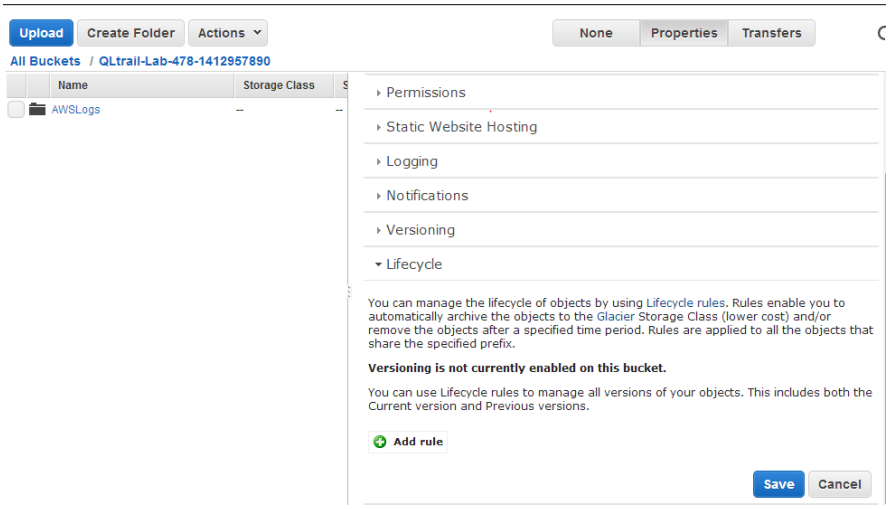

Figure 13: Using the console to enable versioning for an S3 bucket

If your data and workflow need long-term retention with low chances of retrieval then Amazon Glacier is the best solution. Amazon Glacier is a storage service optimized for infrequently used data, or "cold data." The service provides durable and extremely low-cost storage with security features for data archiving and backup. With Amazon Glacier, you can store your data cost-effectively for months, years, or even decades. With Amazon S3's lifecycle management, you can automatically move data from Amazon S3 to Amazon Glacier based on the lifecycle policy.

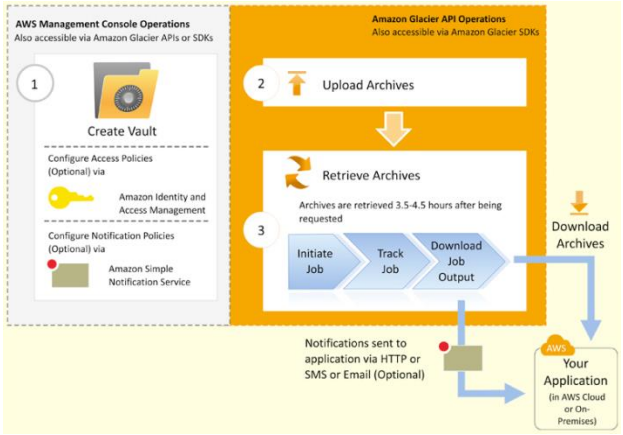

Figure 14: Amazon Glacier storage

If your data and workflow require a file system to store files or database data then Amazon EBS is the best storage option. Amazon EBS provides many features such as high durability and reliability, encryption, provisioned IOPS, and point-in-time snapshots amongst others. The built-in volume snapshot feature is a good option for backing up data.

Automating Infrastructure

One of the main advantages of using Amazon Web Services is that capacity is available to you on demand. You have no need to preprovision your infrastructure for future or backup use. Using tools such as AWS CloudFormation[25] and AWS OpsWorks,[26] you can automate the build out of your infrastructure, as explained in the Bootstrapping Applications whitepaper.[27] With this approach, you can operate your infrastructure as code. You are then not tied to a specific system image that you have to backup. The application can be backed up in code repositories and used to create a full-blown infrastructure at the time it is needed. Anytime you need to create a server, you launch an automated deployment of the application, which creates the infrastructure within minutes to host your application.

Conclusion

The growth in data and an explosion in the creation and use of machine-generated data are increasing the need for robust, scalable and secure backup solutions. At the same time, organizations are struggling to deal with an explosion in retained data for compliance or business reuse. Providing IT teams with services and solutions that are optimized for usability in backup and archival environments is a critical requirement.

Amazon Web Services provides cost-effective and scalable solutions to help organizations balance their requirements for backup and archiving. These services integrate well with new as well as existing technologies the customers are working with today. Gartner has recognized AWS as a leader in providing public cloud storage services[28]. AWS is well positioned to help organizations move their workloads to the cloud-based platforms that are the next generation of backup.

Notices

[25] http://aws.amazon.com/cloudformation/

[26] http://aws.amazon.com/opsworks/

[27] https://s3.amazonaws.com/cloudformation-examples/BoostrappingApplicationsWithAWSCloudFormation.pdf

[28] http://www.gartner.com/technology/reprints.do?id=1-1WWKTQ3&ct=140709&st=sb

customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

Appendices

Terms

Some important terms often used in backup and restore discussions

- **Archive** – A strategy for long-term retention of data for use in case of compliance, regulatory, or historical records requirement.
- **Backup** – A strategy of copying files or databases for short-term retention for use in case of failure or corruption.
- **Backup Frequency** – The time between consecutive backups.
- **Data Lifecycle Management** – The process of managing data information throughout its lifecycle, from requirements through retirement.
- **Data Versioning** – Maintaining multiple versions of data for backup purposes.
- **File / Data Backup** – The process of copying individual data files to a backup medium so that they will be preserved.
- **Image Backup** – An exact copy of a drive or storage device containing the complete contents and structure representing the operating system and all the data associated with it, including the system state and application configurations.
- **Off-Site Backups** – The process of storing the copy of data in a geographically different location from the source.
- **Restore** – A process that involves copying backup files from secondary storage (tape, zip disk, or other backup media) to hard disk. A restore is performed in order to return data to its original condition if files have become damaged or to copy or move data to a new location.
- **Retention** – The amount of time that a given set of data remains available for restore. Some backup products rely on daily copies of data and measure retention in terms of days. Others retain a number of copies of data changes regardless of the amount of time.
- **RPO** – The maximum tolerable period in which data might be lost from an IT service due to a major incident.
- **RTO** – The targeted duration of time and a service level within which a business process must be restored after a disaster (or disruption) in order to avoid unacceptable consequences associated with a break in business continuity.

Partner Solutions

- **Avere** – Hybrid cloud NAS and AWS: http://www.averesystems.com/amazon-web-services
- **Commvault** – Cloud Integration with Amazon Web Services: http://www.commvault.com/resource-library/1843/commvault-amazon-web-services-solution-brief.pdf
- **CTERA** – CTERA Cloud Storage Services Platform and Amazon Web Services: http://www.ctera.com/amazon-aws-cloud-storage-platform
- **NetApp Riverbed SteelStore™** – cloud storage gateway: http://www.riverbed.com/partners/find-a-partner/find-a-partner-tool/aws-partner.html#Cloud_Storage
- **Symantec Solutions for Amazon Web Services** – Symantec Netbackup Platform: http://www.symantec.com/content/en/us/enterprise/fact_sheets/b-nbu-DS-solutions-for-amazon-web-services-21281095.en-us.pdf
- **Zmanda** – Backup to Amazon S3: http://www.zmanda.com/backup-Amazon-S3.html