Using AMI to Back Up EC2 Instances

AWS stores system images in what are called Amazon Machine Images (AMIs). These images consist of the template for the root volume required to launch an instance. You can use the AWS Management Console or the `aws ec2 create-image` CLI command to back up the root volume as an AMI.

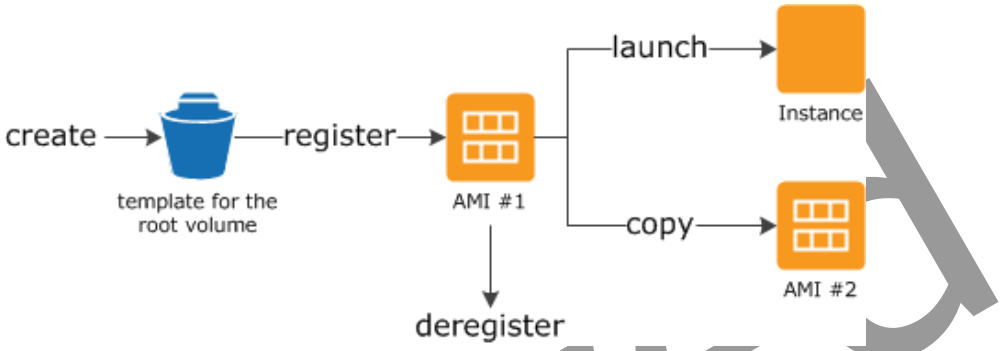

Figure 3: Using an AMI to Back Up and Launch an Instance

When you register an AMI, it is stored in your account using Amazon EBS snapshots. These snapshots reside in Amazon S3 and are highly durable.

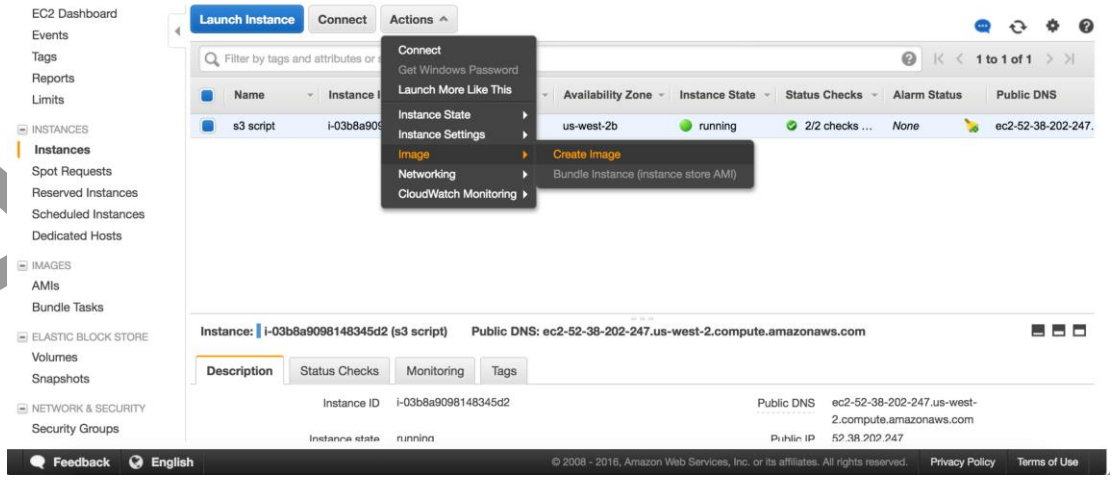

Figure 4: Using the EC2 Console to Create a Machine Image

After you have created an AMI of your Amazon EC2 instance, you can use the AMI to re-create the instance or launch more copies of the instance. You can also copy AMIs from one region to another for application migration or disaster recovery.

On-Premises to AWS Infrastructure

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

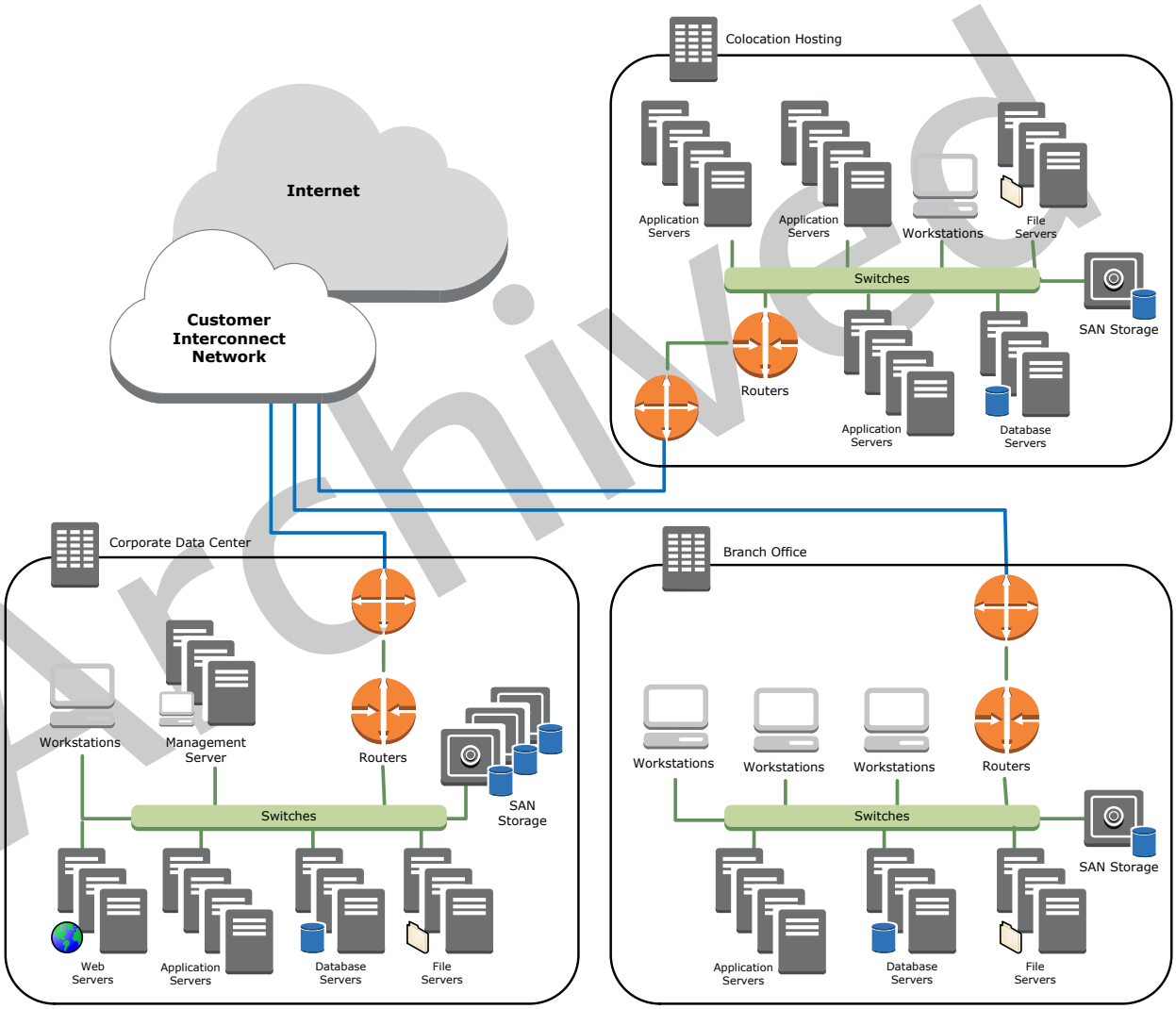

Figure 5: On-Premises Environment

By using AWS storage services in this scenario, you can focus on backup and archiving tasks. You don't have to worry about storage scaling or infrastructure capacity to accomplish the backup task.

Amazon S3 and Amazon Glacier are natively API-based and available through the Internet. This allows backup software vendors to directly integrate their applications with AWS storage solutions, as shown in the following figure.

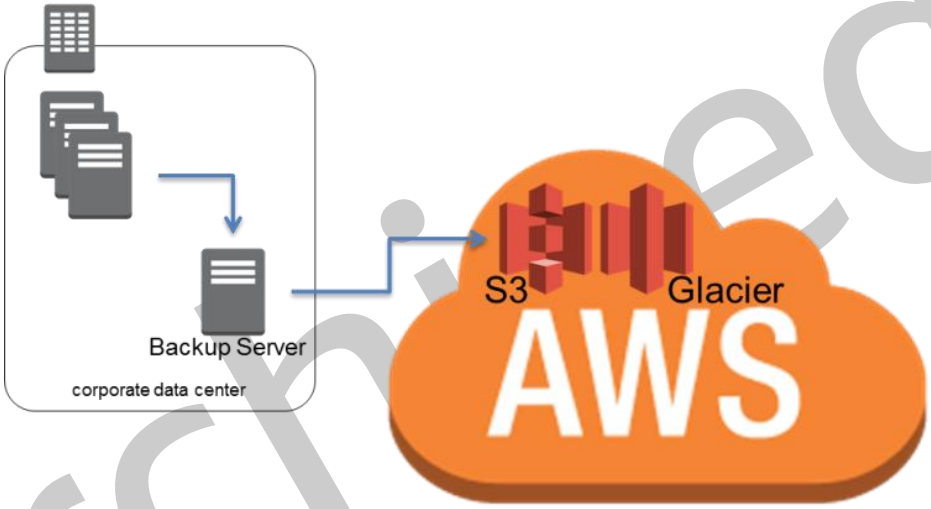

Figure 6: Backup Connector to Amazon S3 or Amazon Glacier

In this scenario, backup and archive software directly interfaces with AWS through the APIs. Because the backup software is AWS-aware, it will back up the data from the on-premises servers directly to Amazon S3 or Amazon Glacier.

If your existing backup software does not natively support the AWS cloud, you can use AWS storage gateway products. AWS Storage Gateway[13] is a virtual appliance that provides seamless and secure integration between your data center and the AWS storage infrastructure. The service allows you to securely store data

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

in the AWS cloud for scalable and cost-effective storage. 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.

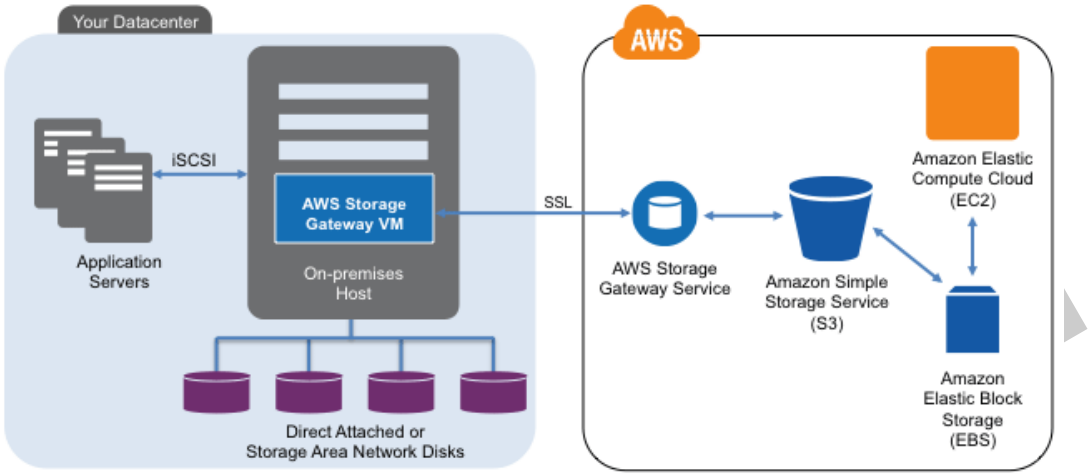

Figure 7: Connecting On-Premises to AWS Storage

AWS Storage Gateway supports the following configurations:

- **Volume gateways:** Volume gateways provide cloud-backed storage volumes that you can mount as Internet Small Computer System Interface (iSCSI) devices from your on-premises application servers. The gateway supports the following volume 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.

Hybrid Environments

The two infrastructure deployments discussed to this point, cloud-native and on-premises, can be combined into a hybrid scenario where the workload environment has on-premises and AWS infrastructure components. Resources, including web servers, application servers, monitoring servers, databases, Active Directory, and more are hosted either in the customer data center or AWS. Applications running in the AWS cloud are connected to applications running on-premises.

This is becoming a common scenario for enterprise workloads. Many enterprises have data centers of their own and use 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[14], you can establish private, dedicated connectivity from your premises to AWS. This provides the bandwidth

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

and consistent latency to upload data to the cloud for the purposes of data protection and consistent performance and latency for hybrid workloads.

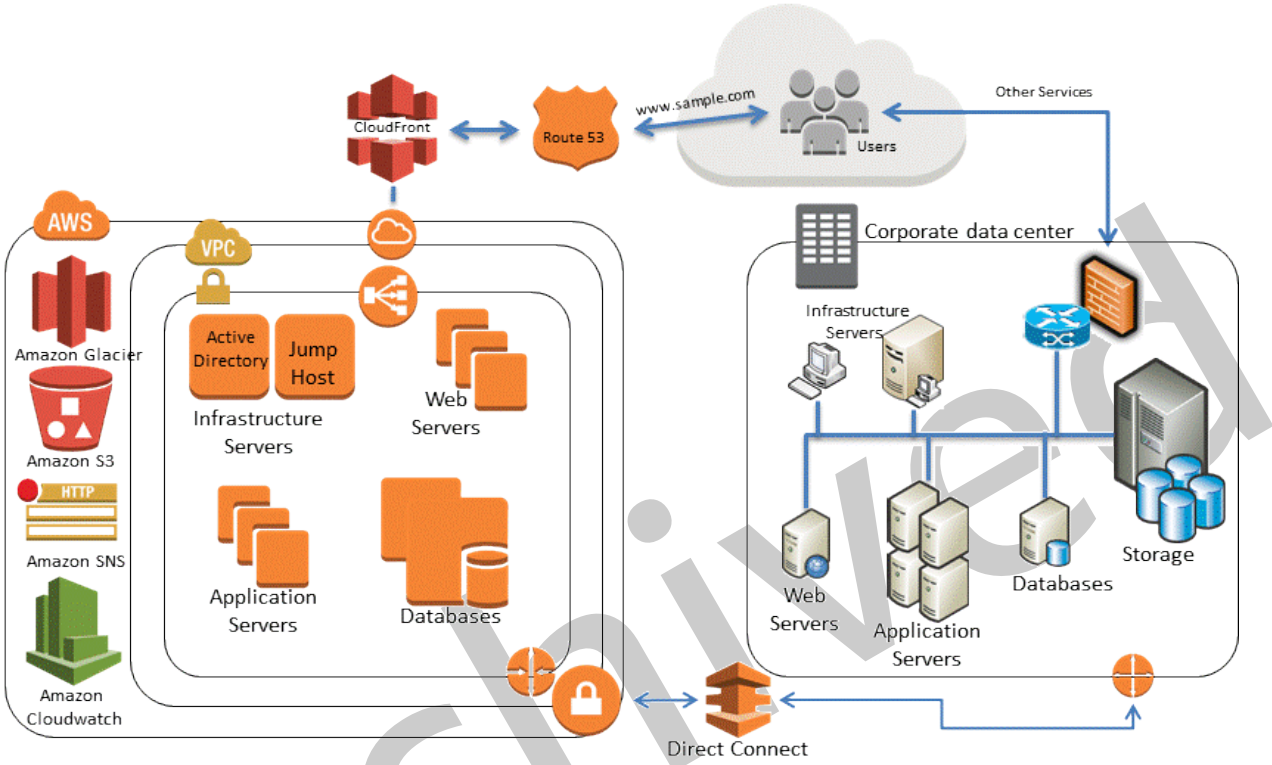

Figure 8: A Hybrid Infrastructure Scenario

Well-designed data protection solutions typically use a combination of the methods described in the cloud-native and on-premises solutions.

Backing Up AWS-Based Applications to Your Data Center

If you already have a framework that backs up data for your on-premises servers, then it is easy to extend it to your AWS resources over a VPN connection or through AWS Direct Connect. You can install the backup agent on the Amazon EC2 instances and back them up per your data-protection policies.

Migrating Backup Management to the Cloud for Availability

Depending on your backup architecture, you may have a master backup server and one or more media or storage servers located on-premises with the services it's protecting. In this case, you might want to move the master backup server to an Amazon EC2 instance to protect it from on-premises disasters and have a highly available backup infrastructure.

To manage the backup data flows, you might also want to create one or more media servers on Amazon EC2 instances. Media servers near the Amazon EC2 instances will save you money on internet transfer and, when backing up to S3 or Amazon Glacier, increase overall backup and recovery performance.

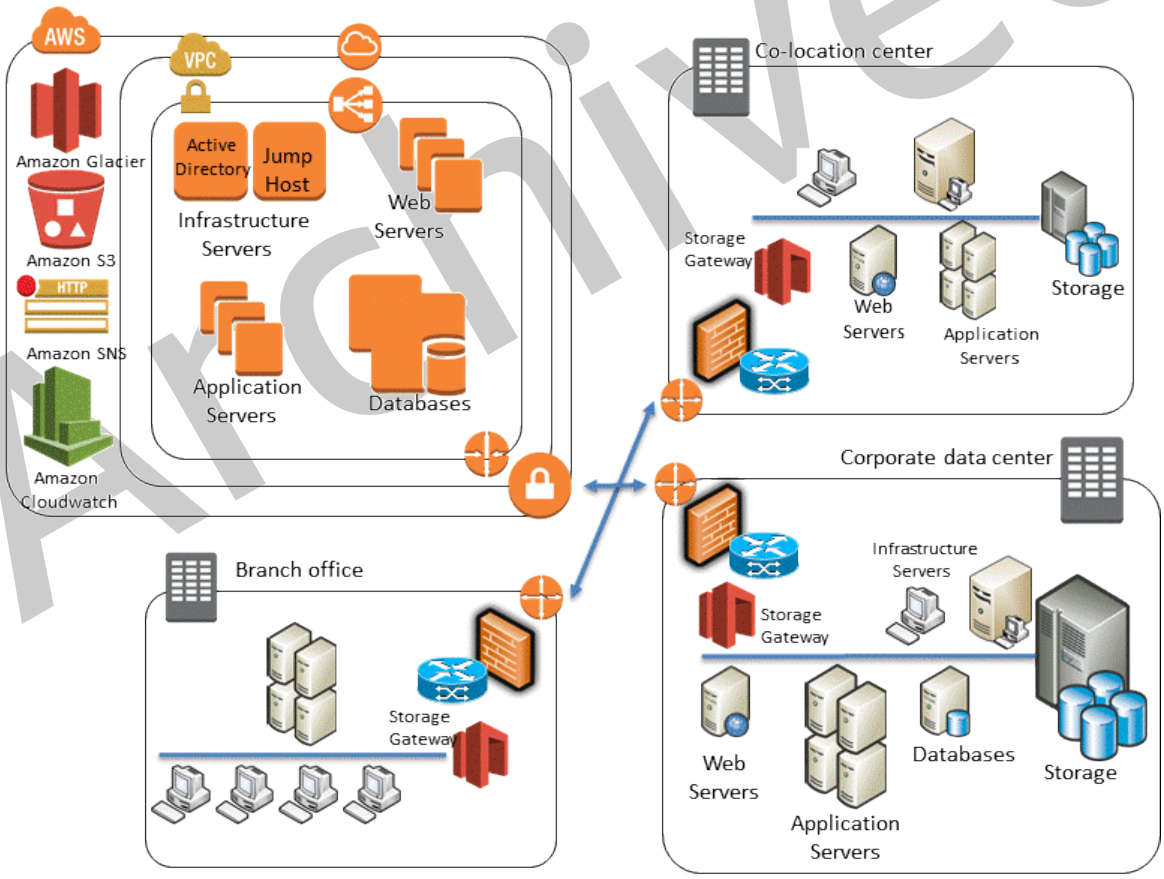

Figure 9: Using Gateways in the Hybrid Scenario

Example Hybrid Scenario

Assume that you are managing an environment where you are backing up Amazon EC2 instances, standalone servers, virtual machines, and databases. This environment has 1,000 servers, and you back up the operating system, file data, virtual machine images, and databases. There are 20 databases (a mixture of MySQL, Microsoft SQL Server, and Oracle) to back up.

Your backup software has agents that back up operating systems, virtual machine images, data volumes, SQL Server databases, and Oracle databases (using RMAN). For applications like MySQL that your backup software does not have an agent for, you might use the mysqldump client utility to create a database dump file to disk where standard backup agents can then protect the data.

To protect this environment, your third-party backup software most likely has a global catalog server or master server that controls the backup, archive, and restore activities as well as multiple media servers that are connected to disk-based storage, Linear Tape-Open (LTO) tape drives, and AWS storage services.

The simpliest way to augment your backup solution with AWS storage services is to take advantage of your backup vendor's support for Amazon S3 or Amazon Glacier. We suggest you work with your vendor to understand their integration and connector options. For a list of backup software vendors who work with AWS, see our partner directory[15].

If your exising backup software does not natively support cloud storage for backup or archive, you can use a storage gateway device, such as a bridge, between the backup software and Amazon S3 or Amazon Glacier.

There are many third-party gateway solutions. You can also use AWS Storage Gateway virtual appliances to bridge this gap because it uses generic techniques such as iSCSI-based volumes and virtual tape libraries (VTLs). This configuration requires a supported hypervisor (VMware or Microsoft Hyper-V) and local storage to host the appliance.

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

Archiving Data with AWS

When you need to preserve data for compliance or corporate reasons, you archive it. Unlike backups, which are usually performed to keep a copy of the production data for a short duration to recover from data corruption or data loss, archiving maintains all copies of data until the retention policy expires.

A good archive meets the following criteria:

- Data durability for long-term integrity
- Data security
- Ease of recoverability
- Low cost

Immutable data stores can be another regulatory or compliance requirement.

Amazon Glacier provides archives at low cost, native encryption of data at rest, 11 nines of durability, and unlimited capacity.

Amazon S3 Standard - Infrequent Access is a good choice for use cases that require the quick retrieval of data. Amazon Glacier is a good choice for use cases where data is infrequently accessed and retrieval times of several hours are acceptable.

Objects can be tiered into Amazon Glacier either through lifecycle rules in S3 or the Amazon Glacier API. The Amazon Glacier Vault Lock feature allows you to easily deploy and enforce compliance controls for individual Amazon Glacier vaults with a vault lock policy. You can specify controls such as “write once, read many” (WORM) in a vault lock policy and lock the policy from future edits. For more information, see Amazon Glacier.

Securing Backup Data in AWS

Data security is a common concern. AWS takes security very seriously. It’s the foundation of every service we launch. Storage services like Amazon S3 provide strong capabilities for access control and encryption both at rest and in transit. All Amazon S3 and Amazon Glacier API endpoints support SSL encryption for

data in transit. Amazon Glacier encrypts all data at rest by default. With Amazon S3, customers can choose server-side encryption for objects at rest by letting AWS manage the encryption keys, providing their own keys when they upload an object, or using AWS Key Management Service (AWS KMS)[16] integration for the encryption keys. Alternatively, customers can always encrypt their data before uploading it to AWS. For more information, see Amazon Web Services: Overview of Security Processes.

Conclusion

Gartner has recognized AWS as a leader in public cloud storage services[17]. AWS is well positioned to help organizations move their workloads to cloud-based platforms, the next generation of backup. AWS provides cost-effective and scalable solutions to help organizations balance their requirements for backup and archiving. These services integrate well with technologies you are using today.

Contributors

The following individuals contributed to this paper:

- Pawan Agnihotri, Solutions Architect, Amazon Web Services
- Lee Kear, Solutions Architect, Amazon Web Services
- Peter Levett, Solutions Architect, Amazon Web Services

[16] http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingKMSEncryption.html

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

Document Revisions

Updated May 2016