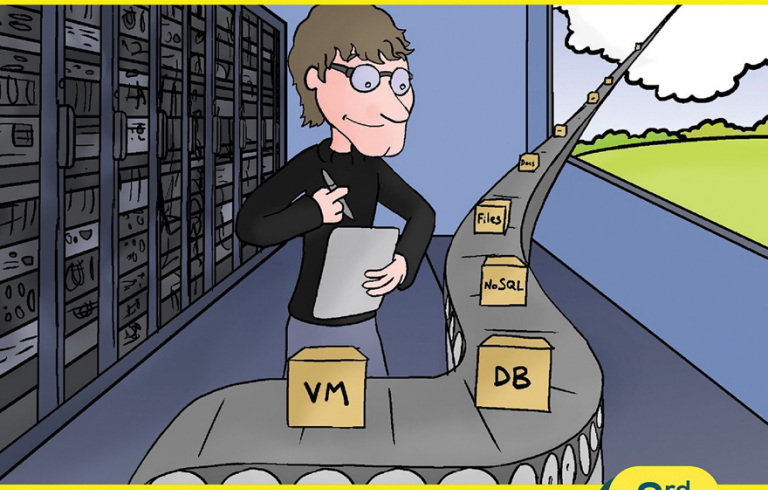
ConversationalGeek

Conversational AWS Data Protection

Jayendra Patil (AWS Certified Developer and Solutions Architect)



3rd MINI Edition

Learn about:

- Building a Robust Data Protection Strategy for AWS
- Ways to Enhance Your AWS Data Protection, and Develop a Holistic View of Modern Data Management

Sponsored by Cohesity

Cohesity ushers in a new era in data management that solves a critical challenge facing businesses today: mass data fragmentation. The vast majority of enterprise data — backups, archives, file shares, object stores, and data used for dev/test and analytics — sits in fragmented infrastructure silos that make it hard to protect, expensive to manage, and difficult to analyze. Cohesity consolidates silos onto one web-scale platform, spanning on-premises, cloud, and the edge, and uniquely empowers organizations to run apps on that platform — making it easier than ever to back up and extract insights from data. Cohesity is a 2019 CNBC Disruptor and was named a Technology Pioneer by the World Economic Forum.

COHESITY

For more information, visit

www.cohesity.com

twitter.com/cohesity

www.linkedin.com/company/cohesity

www.facebook.com/cohesity

www.cohesity.com/blog

Conversational AWS Data Protection (3rd Mini Edition)

by Jayendra Patil

ConversationalGeek

Conversational AWS Data Protection (3rd Mini Edition)

Published by Conversational Geek® Inc.

www.ConversationalGeek.com

Trademarks

Warning and Disclaimer

Additional Information

For general information on our other products and services, or how to create a custom Conversational Geek book for your business or organization, please visit our website at www.ConversationalGeek.com.

Publisher Acknowledgments

All of the folks responsible for the creation of this book:

Author:	Jayendra Patil
Project Editor:	Hope Crocker
Copy Editor:	Nick Cavalancia
Content Reviewer(s):	Raj Dutt
	Michelle Garcia

The “Conversational” Method

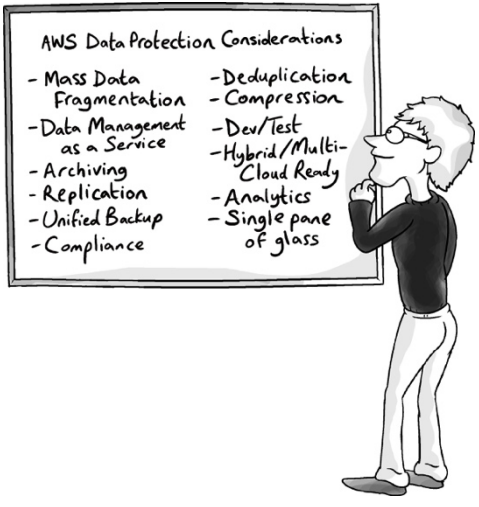
We have two objectives when we create a “Conversational” book. First, to make sure it’s written in a conversational tone so that it’s fun and easy to read. Second, to make sure you, the reader, can immediately take what you read and include it into your own conversations (personal or business-focused) with confidence.

“Geek in the Mirror” Boxes

We infuse humor and insight into our books through both cartoons and light banter from the author. When you see one of these boxes it’s the author stepping outside the dialog to speak directly to you. It might be an anecdote; it might be a personal experience.

Within these boxes I can share just about anything on the subject at hand from the author or other SMEs. Read ’em!

Data Protection in Amazon Web Services (AWS)



"Data is a precious thing and will last longer than the systems themselves." – Tim Berners-Lee (inventor of the World Wide Web)

Data… including, big data… is information that must be collected, consolidated, organized, transformed, managed, protected, and secured. How we

accomplish that in modern times is an ever-moving target.

In today's data-driven world, data is available in various structured and unstructured formats – including streaming data, files and objects, and relational and non-relational data – and comes from a variety of source systems – including cloud-native, SaaS, and on-premises applications.

This data fuels organizations' business intelligence and decision making, so it is critical to ensure uptime and continuity. Data backup is essential as it helps to restore your environment, meet the SLAs for fault tolerance and disaster recovery, and support compliance and regulatory requirements for security and data retention.

However, data backup requires that you not only manage the data itself but also the backup infrastructure and all the related operations. This may include managing tapes, tape drives, backup storage targets, backup servers, and backup software tools, or sending tapes offsite, creating backup policies, ensuring the backup data is secure,

meeting compliance requirements for data retention, and performing restores.

With the advent of hybrid cloud offerings like AWS Outposts, VMware Cloud on AWS, and others, more companies are now looking for decoupled, cloud-agnostic, centralized options for data management and protection.

Moreover, data protection needs to provide data security and privacy. Hacking and ransomware attacks are ever-present threats, so organizations must use analytics to identify and secure sensitive data.

According to Flexera's 2022 State of the Cloud Report, 89% of businesses have a multi-cloud strategy and 80 percent are already taking a hybrid approach by combining the use of both public and private clouds.

AWS provides a secure, high-performance, flexible, cost-effective, and easy-to-use cloud computing platform, and with it, multiple data backup services, which eliminate the manual processes and heavy

lifting required for data backups and data protection.

AWS Data Protection Techniques

AWS provides various storage services, ranging from object storage, block storage volumes, file systems to relational and NoSQL databases. Although it's easy to get started, there are data protection concerns to address, including cost, replication, isolation, point-in-time backup/recovery, versioning, and centralized management across the services. You also need to plan your data protection strategy carefully to fully enable disaster recovery and to meet business and regulatory compliance requirements.

AWS provides zonal or regional data durability for most of its fully managed storage services, like S3. However, as per the shared responsibility model, users are still responsible for the protection and management of their data resident on services like EBS volumes, RDS databases, EFS, and DynamoDB tables, as well as any on-premises data center resources and AWS Outposts.

AWS Backup

AWS Backup provides a fully managed automated, centralized, single interface for the backup of data across AWS services and on-premises data centers using AWS Storage Gateway. AWS Backup provides you with an on-demand or policy-based backup solution, as well as supporting cross-account and cross-region backup. This helps meet both business continuity and compliance requirements, including storing backups at a minimum distance away from the production data.

AWS Backup covers resources like Amazon EC2 instances, Amazon EBS volumes, Amazon RDS databases (including Aurora), Amazon DynamoDB tables, Amazon EFS file systems, Amazon FSx for Linux, Amazon FSx for Windows, and AWS Storage Gateway volumes. AWS Backup provides organizations with solutions for:

- centralized backup management
- policy-based backup
- tag-based backup policies
- on-demand backups
- automated backup schedules

- automated retention management using lifecycle policies
- incremental backups
- all AWS accounts within AWS organizations
- data protection using encryption at rest and in transit
- protection and recovery of critical data from ransomware events and account compromise
- immutability, with the protection of backups from deletion or changes to their lifecycle by inadvertent or malicious changes
- backup activity monitoring and auditing across AWS services and on-premises data centers using AWS Storage Gateway

AWS Elastic Disaster Recovery

AWS Elastic Disaster Recovery service (DRS) provides a cost-effective disaster recovery site on AWS for any source (on-premises or another cloud provider) with server-hosted applications and server-hosted databases. AWS DRS provisions a site

with minimal version of resources and in the event of a failover, scales it to full capacity.

Simple Storage Service (S3) Replication

Amazon S3 provides secure, durable, highly scalable object storage at a very low cost. S3 comes with unlimited storage and is the most widely used storage option from AWS. S3 is usually the starting point for migration, backups, and data capture in big data, IoT solutions.

AWS S3 ensures data availability by replicating the data in multiple availability zones (AZs) – as per the storage class – within the same region. S3 is designed to sustain data in the event of an entire S3 AZ loss.

Replication and durability for S3 are good. However, it does not protect against point-in-time failure or overwrite. Versioning (turned off by default) needs to be turned on, but it gives you recoverability based on the history for that one given object. There is no way to recover all objects based on the state/version on an exact date/time due to a lack of backup.

Due to compliance and regulatory challenges, Amazon S3 does not, by default, replicate the data across regions. If you need to back up data across regions, AWS S3 supports both same-region and cross-region replication (CRR) – at an additional cost – and automatically replicates data within the same or between different AWS regions. With CRR, objects remain encrypted throughout the process. Encrypted objects are transmitted securely via SSL from the source region to the destination region. S3 now supports S3 Replication Time Control (S3 RTC) to help meet compliance or business requirements for data replication and provides visibility into S3 replication times.

Turning on Backup for Elastic Cloud Compute (EC2) Servers

Enterprise adoption of EC2 servers falls into many use cases, from full production environments, to test and development deployments. AWS by default will not turn on backups of these servers, and administrators have some options to consider. Your workflow can allow you to go in and enable the backups manually per server or by using a wizard to set up a backup schedule via SLAs of your choice.

When using manual or wizard-based options for SLA management of backups it is easy to "miss" backing up a server. This solution also doesn't offer a single management interface across all of your systems which is necessary for simplicity and efficiency.

Amazon Data Lifecycle Manager

Amazon EBS volumes provide block level storage in AWS. With EC2, they are the basic building blocks for almost all of the services in AWS.

It's important to note that most of the Amazon EBS volumes are designed for an annual failure rate (AFR) of between 0.1% and 0.2%, "where failure refers to a complete or partial loss of the volume, depending on the size and performance of the volume."

goto.cg/44477MA

Amazon Data Lifecycle Manager allows you to automate the creation, retention, and deletion of snapshots taken to back up the Amazon EBS volumes. By automating snapshot management, you can benefit from:

- data protection enforced by a regular backup schedule
- retention, as required by auditors or internal compliance
- lifecycle management, to reduce storage costs by deleting outdated backups
- monitoring and auditing capabilities

Relational Database Service (RDS) Backup and Replication

Amazon Relational Database Service (RDS) is a managed service that provides the ability to setup, operate, monitor, and scale relational databases in the cloud.

RDS hosts data that usually forms the core and critical parts of any business, and performs backup and restore operations to maintain a defined RTO and RPO SLA. RDS also provides multiple service options, including backups, or synchronous and asynchronous replication.

Furthermore, similar to EBS volumes, RDS provides two different ways for backing up the DB instances – automated backups and snapshots.

With automated backups, RDS performs a full daily snapshot of the DB instance and the transaction logs, helping point-in-time recovery. Automated backups are retained for a limited time (7-35 days) and are deleted when the DB instance is deleted. DB snapshots are manually initiated by the user, are

independent of the DB instance, and are not deleted when the DB instance is deleted.

Finally, RDS allows cross-region copying of snapshots to enable the restoration of DB instances in a different region for disaster recovery. Data protection can be augmented by enabling encryption of the RDS instance.

Storage Gateway for Remote Backup

With on-premises data centers, there has always been a need for remote or offsite backups, which involves replicating the data to a remote data center or shipping backup tapes to an off-site storage facility. Public clouds have become a viable, cost-effective option for remote backup storage in recent years due to the need of ever-expanding, low-cost storage and various techniques to maintain security and compliance.

AWS Storage Gateway service enables hybrid storage between on-premises environments and AWS, enabling existing on-premises backup applications to store backups on Amazon S3 and mirror data to cloud-based compute resources. It's

an on-premises software appliance with cloud-based storage-to-storage integration between an organization's on-premises data center and AWS. AWS Storage Gateway supports industry-standard storage protocols that work with existing applications. AWS Storage Gateway is not backed up by default, so administrators need to do this independently.

We need to look at more than just moving data to the cloud; we have to consider the recoverability of that data.

Questions that are worth considering include: once data is deleted from on-premises solutions, what is the retention policy? Is it automatically removed from the cloud? Does it get aged out?

AWS Data Protection Considerations

As elaborated in the previous section, Amazon Web Services provides a number of cost-effective and scalable solutions to help organizations balance their requirements for backup and archiving.

Although AWS covers a large number of backup requirements, with the current market evolving it does present some challenges.

While AWS provides the ability to back up all types of data, the data is backed up as-is; AWS does not perform data deduplication and compression. De-duplicating and compressing data can help save a significant amount of time and money during migrations and transfers and can reduce the cost of data storage.

AWS backup services also perform the backups and the data is stored within the same tenant account usually sharing the same access controls which is a security risk. A compromise in the account risks both the original and the backed-up data.

AWS also does not provide an out-of-the-box data management (consolidation, compression, instant searchability, security, and analytics) solution for the data across all its services. Customers need to use individual marketplace solutions which help with data compression, deduplication, management, and protection. These tools will benefit organizations with large amounts of data, as

they will help reduce both data storage and data transfer egress costs.

Again, we need to look beyond backup considerations and examine recoverability. How easily can you recover the RIGHT data from the RIGHT point in time?

Problems with recoverability include backup data being aged off and deleted, or having bad data replicated and overwriting good data. These are areas where you may wish to look for more comprehensive solutions that protect data on AWS.

Leveraging a private cloud offering or a hybrid / multi-cloud environment is a leading trend across the industry. Most organizations either have, or are trying to implement, a hybrid / multi-cloud environment to take advantage of the specialized services offered by other cloud vendors like VMware, Azure, and GCP as well as other SaaS offerings like Salesforce.

Modern organizations are looking for solutions that can provide them with a single centralized interface

for managing their distributed data environments, getting visibility into their entire data estate, compressing it appropriately to rein in storage and migration costs, allocating appropriate compute resources, upholding governance, managing security, and monitoring activity and status (all without any vendor lock-in and tight coupling).

The same scenario is true for backups, as organizations prefer a common management tool with which to interface, manage backups, and search and restore across multiple clouds, on-premises, and co-located or edge environments. Organizations can turn to marketplace solutions to support this multi-cloud requirement.

A lot of organizations also use licensed third-party software, which delivers more comprehensive backup and recovery features. Some also provide a single interface across hybrid and multi-cloud environments, which is not commonly supported by AWS backup services. Customers need either to rely on the third-party solution to provide a backup solution, or to rely on a marketplace solution.

Most organizations also have existing backup solutions for their on-premises environment and prefer not to use AWS backup solutions, as they are tightly integrated with AWS services and can introduce a vendor lock-in scenario.

There is a small handful of data management vendors that bridge the gap between the on-premises and public cloud world by providing a platform-oriented solution that does backups, DR, cross-cloud data compression, and optimizes data archival and migration. These offerings work as well on AWS as they do on-premises (say, on VMware), on other public clouds, or on edge locations.

Organizations and the market in general view these vendors favorably, as they provide not only a point solution for each use-case, but also a holistic offering for data management across locations. From an ownership standpoint, they offer the flexibility to be deployed and operated by the customer, as a service from public cloud, or fully managed by a partner such as a managed service provider.

It's not just vendor lock-in we should be concerned about; it's the problem of siloed data as well. Organizations want on-premises and cloud solutions to interact better, and to share data for innovation. AWS backup copy could potentially create another silo in the cloud which requires time and effort to maintain. Instead, a hybrid cloud backup solution can simplify data management and enable more agility and innovation.

The Need for Cloud-based Data Security & Management

It is unreasonable – perhaps even impossible – to expect your IT team to perform all the aforementioned tasks of reining in the problem of siloed data. It would require a broadly cross-functional and multifarious, senior-level team across backup, security, storage, cloud, and devops to truly rein in all your data and start seeing top and bottom-line results. And heavy infrastructure investments to go with it.

However, just like AWS ushered in a new paradigm with Infrastructure as a Service (IaaS), and then Platform as a Service (PaaS), data management is likely to serve you best when consumed as a fully managed software service, as a simple line item on your monthly cloud bill.

Cloud-based data security & management bridges the chasm between infrastructure/platform and analytics/machine learning. To be able to protect and put your data to good use, it's not enough that it's available and resilient, but it also needs to be controllable, isolated, visible, and secure.

Since we are talking about AWS data protection, c loud-based data security & management offerings will typically start with Backup as a Service. It is something you need to do anyway, but instead of just being an insurance policy, it also becomes an excellent data collector for all types of data from on-premises apps, cloud-native services (EBS, RDS, EFS), and SaaS (M365) – including backups of AWS workloads and data.

With all this data collected and consolidated in a cloud service that you can plug into AWS (and

associated infrastructure services, such as data warehouses and data lakes) on one side, and to your analytics platform and developers on the other, then you have yourself a fully optimized, fully operational data engine.

The most robust data engines will have Backup as a Service (BaaS), Security as a Service, Disaster Recovery as a Service (DRaaS), Cyber Vaulting as a service, Threat Detection as a service, all deduplicated, secure and available on a single platform, consumed as a service from the AWS marketplace.

The Big Takeaways

In recent years, there has been tremendous growth in the amount of data businesses store and produce, as well as a fundamental change in the nature and type of that data. This includes the creation and capture of machine-generated data from a variety of sources.

There is a need for cost-effective, durable, isolated, scalable, and secure backup solutions. At the same time, backup is only the first step, and it's critical that organizations have an overarching data security and management strategy that helps them get ahead of siloed data.

The purview of data management includes data compression, isolation, visibility, searchability, security, SLA delivery, compliance, and availability, across clouds and deployments.

AWS, already a cloud leader, is well-positioned to help organizations move their workloads to the cloud-based platforms that would be an essential part of next-generation backup. AWS does, indeed, provide cost-effective and scalable solutions to help

organizations balance their requirements for backup, archive, and restore. However, there are gaps that a third-party solution can address.

Choosing a solution that will allow your enterprise to manage its entire data protection strategy from a single pane of glass to break down work silos is just the beginning.

A solution that can see all of your data allowing for true insights will ensure that you are securing and optimizing your backups. Hosting data in completely different isolated, controlled, air-gapped tenants help against internal and external cybersecurity and ransomware incidents. Data deduplication and compression will ensure that your storage costs are optimized by storing only unique file segments, and a modern solution will not only allow you to backup your on-premises workloads, and other SaaS solutions, but your cloud-based ones as well.

Finally, a truly modern solution will not only allow you to manage your organization's data everywhere, but also offer the flexibility to consume it as a cloud service.

COHESITY

Migrate, protect, and do more with your data

Get enterprise-class backup and recovery for your AWS workloads and cloud-native apps. Seamlessly extend and migrate legacy workloads and datasets to the cloud.

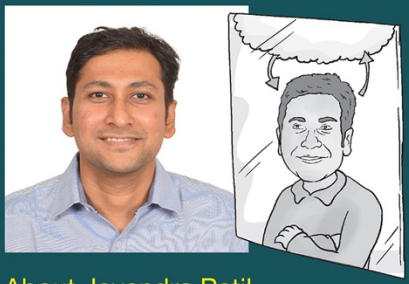
Migrate, failover, and archive with ease

Enterprise-grade protection

Hybrid cloud made easier

Request free trial ›

Amazon Web Services (AWS) provides a robust cloud computing platform – one that includes a variety of different data types that require protection. In this eBook, the focus will be on the different ways AWS provides protection for that data and areas for enhancement.



About Jayendra Patil

Jayendra is an AWS Certified Developer (Associate) and Solutions Architect (Associate). He is an experienced Big Data and Cloud Architect with a long history working in the consulting and finance industries.