Protecting and governing your data on AWS

Nancy Wang (she / her / hers)
General Manager, AWS Data Protection Services
Amazon Web Services

Satish Lakshmanan (he / him / his)
Managing Director, Specialized Sales, Business Development, and Solutions Architecture
Amazon Web Services
Creating and consuming data on AWS

Layers of data governance on AWS

Baseline controls for data protection on AWS

Real-world scenarios for AWS data protection and governance
Data workloads are the life blood of many companies

- Structured Data
  - Amazon Redshift
  - Amazon Relational Database Service (Amazon RDS)

- Semi-structured Data
  - Amazon DynamoDB
  - Amazon DocumentDB (with MongoDB compatibility)

- Unstructured Data
  - Amazon Simple Storage Service (Amazon S3)
  - Amazon Elastic File System (Amazon EFS)
  - Amazon FSx
Data explosion is creating a balancing act between access and control

All companies are Data companies. We need centralized and consistent data protection and governance to protect our businesses.
Data governance ensures businesses can trust their data

Organizations lack data knowledge for efficient and effective data governance activities; 30% of time spent on data governance is wasted.

IDC

Data governance is no longer optional for enterprise organizations. They are finally realizing the value of data as an asset that needs to be protected, managed, and maintained to increase asset value.

IDC

Source: IDC & Talend - Deliver Data with a “yes” - 2019
Tiered layers of data governance on AWS provide comprehensive protection

- App-specific controls
- AWS service-specific controls
- AWS Data Protection and Governance
  - Baseline assurance controls
Good data protection is a function of people and technology

**DATA DURABILITY**
*(FUNCTION OF TECHNOLOGY)*

- Important on both primary and secondary data
- Example: Amazon FSx for ONTAP snapshots on the same media as the primary storage + archived into different media with WORM lock on the backup vault)

**DATA RESILIENCE**
*(FUNCTION OF EXTERNAL FACTORS)*

- Protects data from human factors: errors, bad actors
- Example: version control for Amazon S3, WORM lock, snapshots, cross-Region and cross-account backup copy

*From new business opportunities to attack surfaces, data protection is a MUST HAVE for enterprises today*
These are baseline controls for data protection on AWS

Visibility

Resiliency and Durability

Security and Controls
Here are common data visibility goals

How do I implement the capability to review all actions taken in my environment?

How do I determine the current state and posture of my data platforms and how they've changed over time?

How do I ensure data usage is compliant with legal, regulatory, and/or contractual policies (e.g., DPAs)?

How do I do these and more, at scale?
These are AWS services that provide data visibility

- **AWS Config**: Continuously assess and evaluate your resource configs
- **AWS CloudTrail**: Establish centralized logging and risk auditing
- **Amazon CloudWatch**: Collect and monitor observability data (metrics, logs)
Track changes and continuously assess your Amazon S3 bucket

- S3 bucket
- Object-level logging
- AWS CloudTrail
- AWS Config
- Config rules
- AWS CloudTrail
- Usage Logs
- Amazon Athena
- Usage Insights
- History, Snapshot
- Posture Assessment
- Notifications

```
{
  "eventVersion": "1.03",
  "userIdentity": {
    "type": "IAMUser",
    "principalId": "1111222333",
    "arn": "arn:aws:iam::1111222333:user/myUserName",
    "accountId": "1112222333",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
    "sessionIssuer": "myUserName"
  },
  "eventTime": "2019-02-01T03:22:33Z",
  "eventSource": "s3.amazonaws.com",
  "eventName": "PutBucketACL"
}
```
Resiliency and Durability
These are core principles of data resiliency and durability

**Centralized**
- Fully-managed
cross-AWS support

**Intent-based**
- RPO, RTO
- Replication
- Archival / retention

**Policy-driven**
- Data protection as code
- Auditable & provable
AWS uses a shared responsibility model for data resiliency

CUSTOMER RESPONSIBILITY FOR RESILIENCY IN THE CLOUD

AWS RESPONSIBILITY FOR RESILIENCY OF THE CLOUD

SECURE DATA BACKUP

WORKLOAD ARCHITECTURE

CHANGE MANAGEMENT

FAILURE MANAGEMENT

NETWORKING, QUOTAS, AND CONSTRAINTS

HARDWARE AND SERVICES

COMPUTE

STORAGE

DATABASE

NETWORKING

AWS GLOBAL INFRASTRUCTURE

REGIONS

AVAILABILITY ZONES

EDGE LOCATIONS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Here are data resiliency and durability solutions that span multiple AWS services

Create central protection policies for your VMs, databases, and storage (object, file, block) on AWS

Failover VMs (Amazon EC2, Amazon EBS) from on-premises to AWS (disaster recovery)
Protect your S3 bucket with AWS Backup

Create automated continuous or periodic backups using a central backup plan.

- Create central protection policies that automatically backup your AWS application data.
- Control access for managed resources via IAM.
- Setup notifications/logging via Amazon SNS, AWS CloudTrail & AWS CloudWatch.
- Setup detailed compliance reporting with AWS Audit Manager.
- Scale through AWS Organizations.
- Backup to other accounts to protect against multiple threats.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Data resiliency in action: restoring an Amazon S3 bucket

- Restore S3 bucket, prefix, or object to a point-in-time with a single click
- Restore to source S3 bucket, another existing bucket, or new bucket
- Restore object data, tags, ACLs, and other user-defined metadata
AWS Backup protects in-cloud and hybrid application data

**AWS Backup**

Policy-based, centralized data protection and management

**Hybrid Workloads**

AWS Storage Gateway

**AWS Storage Services**

- EBS
- FSx
- EFS
- S3

**VMs and Applications**

- Amazon EC2
- Windows
- SQL Server

**Managed Databases**

- Amazon RDS
- Amazon Aurora
- Amazon DynamoDB
- Amazon Neptune
- Amazon DocumentDB (with MongoDB compatibility)
Create a scalable permissions model based on attributes
Use security controls to keep your data perimeter secure

- AWS Cloud
- Encrypting services
- Amazon S3
- AWS Secrets Manager
- Amazon EBS
- Amazon Athena
- Amazon RDS
- Amazon DynamoDB
- AWS KMS
- AWS Identity and Access Management (IAM)
- Amazon S3
- AWS CloudTrail
- Storage
- Audit

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Use VPCs to set up a secure data perimeter

- Consumer VPC
- Provider VPC
- Multi-account
- Simplified network architecture
- Sensitive data
- Services
- Secure AWS Marketplace integration
- Highly scalable up to 40Gbps
- AWS service endpoints
- 3rd-party SaaS/DaaS applications
- 3rd Party VPC
- 3rd Party VPC
- Amazon VPC
- Customer services
- us-east-1

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Create defense in depth using AWS data protection and governance

1st: Identity Constructs
- Users
- Role
- Permissions
- MFA token

2nd: Perimeter and Posture Monitoring
- VPC endpoints
- AWS Organizations
- Bucket
- Config
- Cloud-trail

3rd: Recovering Last Known Good Bits
- AWS Backup
- AWS Elastic Disaster Recovery
- Bucket policy
In practice: Using data protection and security to mitigate and recover from ransomware
Incorporating AWS data protection and governance best practices

1. **Data visibility** is critical for understanding where your data is stored, including business-critical data.
2. Proper **data durability and resiliency** standards keeps data available.
3. **Security and controls** prevent unauthorized access to data.
4. Ongoing data compliance **monitors** for breaches to data access and resiliency.
S3 Storage Lens Overview

- Interactive dashboard experience, in the S3 Console – free to all customers
- Organization-wide visibility
- Drill-down by Region, storage class, bucket, and prefix
- Granular usage & activity metrics
- Recommendations for cost efficiency & data protection best practices

Launched November 2020
Use AWS Backup Vault Lock to secure your backups

Cloud-native Backups

- Protect your critical data across AWS services

Compliance

- Business & regulatory compliance

Disaster Recovery

- Reduce risk of downtime and build foundation for business continuity

Vault Level Protection

WORM Storage

Malicious or Accidental Actions
Customer data protection journey

- Backup, restore, and archive
- Ransomware recovery
- Disaster recovery
Protect your applications data on AWS

Backup vendors: Backup on premises and in the cloud, advanced backup functionality, ability to use AWS Backup APIs to integrate with AWS services.

API: Backup for AWS services, API access to multiple AWS services.

Services:
- Amazon EFS
- AWS Storage Gateway
- Amazon RDS
- Amazon DynamoDB
- Amazon EBS
- Amazon S3
- Amazon S3 Glacier
Customer data protection journey

Backup, restore, and archive

Ransomware recovery

Disaster recovery
Cybersecurity is increasingly important

75% of IT organizations will face one or more RANSOMWARE THREATS by 2025

Gartner, 2021

RANSOMWARE ATTACKS will occur every 11 seconds in 2021 and will result in financial loss of $20T

Source: Sophos the state of Ransomware 2020 Report—Independent study of 5,000 IT managers

The Colonial Pipeline attack should be a wake-up call for hardening our cyber defenses

Ransomware Attacks Grow, Crippling Cities and Businesses

Hackers are locking people out of their networks and demanding big payments to get back in. New data shows just how common and damaging the attacks have become.

The Hill

DHS to issue first cybersecurity regulations for pipelines after Colonial hack

Two directives will seek oversight of the industry after ransomware attack upended gas availability in the Southeast for 11 days

The Washington Post

Ransomware attack on Bose exposes employee SSNs and financial information

The company was forced to notify New Hampshire officials after employees in the state had their information accessed

The San Diego Union-Tribune

Scripps enters fourth week of ransomware attack

The ransomware surge ruining lives

BBC NEWS
Ransomware attacks are more likely when...

**Technical maintenance is behind schedule**
- Patching not up to date
- Irregular untested data backups
- Manual time-consuming processes

**Employee behavior creates security risks**
- Security awareness low
- Vulnerable to social engineering

**Security strategy doesn’t prepare for attacks**
- Overly permissive credentials
- Open trust model allows malware to spread
- No clear governance model
There are trade-offs when creating your ransomware strategy

How much data can you afford to re-create or lose?

Recovery point objective (RPO)

How quickly must you recover? What is the cost of downtime?

Recovery time objective (RTO)

Time

Data loss

Disaster

Downtime
NIST Cybersecurity Framework is the industry standard, ...but many companies lack a recovery strategy.

**IDENTIFY**
Identify an organization’s critical functions, assets and processes and how cybersecurity risks could disrupt them.

**PROTECT**
Define safeguards necessary to protect critical infrastructure services.

**DETECT**
Implement the right measures to identify threats and cyber risks promptly.

**RESPOND**
Define the measures necessary to react to an identified threat.

**RECOVER**
Strategic plans to restore and recover any capabilities damaged during a cybersecurity incident.

- Create trust radius with IAM
- Multi-factor Authentication
- Automated monitoring (Amazon VPC, AWS Organizations organizations, accounts)
- Monitor resource configs (change to Amazon S3, AWS KMS keys restricted, metadata)
- Cross-Region, cross-account backups
- Active auditing and altering
- Vaults or Air Gap Backups
AWS services work together to mitigate ransomware

Identify → Protect → Detect → Respond → Recover

Automate

Investigate

- AWS Security Hub
- Amazon GuardDuty
- Amazon Macie
- Amazon Inspector
- Amazon CloudWatch
- AWS Lambda
- AWS Key Management Service
- AWS Secrets Manager
- AWS Shield
- AWS Firewall Manager
- AWS Identity and Access Management (IAM)
- AWS IoT Device Defender
- AWS Single Sign-On
- AWS WAF
- Amazon CloudWatch
- AWS CloudTrail
- AWS Backup
- AWS Elastic Disaster Recovery
- AWS Systems Manager
- AWS Config
- AWS Elastic Disaster Recovery

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.
High Level Architecture
Isolated Data Vault & Forensic Capabilities - Advanced Plus

**Services used for Security, Monitoring, Auditing, Logging and Alerting**

- **IAM:** Permissions Management
- **AWS KMS:** Data Encryption
- **CloudWatch:** Monitoring, Logging, Alerting
- **CloudTrail:** Control/Data Plane Auditing
- **Macie:** Data Access Threat Detection
- **GuardDuty:** AWS Account Threat Detection
# Anatomy of an actual ransomware event

## First 12 hours

### Infiltrate/access
- Employee hit by social engineering
- application_agent.exe <click-here>
- Gain Windows OS login/Active Directory admin rights
- Push malicious code via Windows update

### Event/issue
- Full event begins/users report pop-up to IT
- IP phones down
- AD destroyed/Windows logins disabled
- No email access

### Isolate
- Personal email used
- Cell phones for comms

## +21 days post-event (millions lost per day)

### FBI engaged
-PO issued

### Cybersecurity consultant

### Scrap
- New hardware ordered
- 4,000 laptops /$3+ million in servers

### Recover
- Local backups not impacted/restores initiated
- No NAS backups; only snapshots
- Snapshot rollback initiated
- Dangerously high IOPS/extreme latency

### Rebuild/restore
- New hardware/rack and stack of new servers
- Manual rebuild of DNS
- Restores resume/Active Directory restored
- New laptops, reimaging begins
- 3+ weeks manual recovery to roll back snapshots

---

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Customer data protection journey

Backup, restore, and archive

Ransomware recovery

Disaster recovery
Strategies for disaster recovery

Active/passive

- Backup & restore
  - RPO/RTO: hours
  - • Lower-priority use cases
  - • Provision all AWS resources after event
  - • Restore backups after event
  - • Cost: $

- Pilot light
  - RPO/RTO: 10s of minutes
  - • Data live
  - • Services idle
  - • Provision some AWS resources after event
  - • Cost: $$

- Warm standby
  - RPO/RTO: minutes
  - • Always running, but smaller
  - • Business-critical
  - • Scale AWS resources after event
  - • Cost $$$

- Multi-site active/active
  - RPO/RTO: real time
  - • Zero downtime
  - • Near-zero data loss
  - • Mission-critical services
  - • Cost $$$$
Atlassian uses Amazon S3 as its main analytics and ML datastore

Atlassian uses AWS central data protection policies (via AWS Backup) to automate backup, restore, and disaster recovery of its application data

By securing and protecting their Amazon S3 buckets, Atlassian is also bolstering its ransomware recovery strategy

“We are happy users of AWS Backup today, it has simplified our backup operations and we look forward to making use of the S3 [bucket] support for data protection] when it is launched.”

- Disaster Recovery Team, Atlassian
What’s your data protection and governance strategy?

- Data protection is about people and technology

- Use AWS services such as AWS Config, Systems Manager, and CloudTrail to get a bird’s eye of your data estate

- Use AWS Backup to centralize data protection policies

- For data sets deemed critical, enable data trails so you have full auditability into how those data platforms are being used

- Use features like S3 Lens to gain data visibility and optimize TCO

- Establish a tagging best practice for your data (e.g., leverage projects like Cloud Custodian to enforce tagging discipline)

- Combine tagging best practices with attribute-based access controls for power policies that scale with data creation

- Air gaps, Vault Lock, and CloudFormation templates with Golden Copies can improve time to recovery

- Establish a tagging best practice for your data (e.g., leverage projects like Cloud Custodian to enforce tagging discipline)
Workshop

re:Invent sessions
STG311: Backup for applications running on the AWS Cloud
STG312: Building a disaster recovery solution with AWS storage services

Chalk talk

re:Invent sessions
STG341: AWS Backup to protect cloud-native resources and hybrid environments
STG332: Data protection and recovery strategies from ransomware events

Breakout

re:Invent sessions
STG203: Backup, disaster recovery, and ransomware protection with AWS
STG211: Driving innovation and insight with cloud data on AWS storage
STG220: Protecting and governing your data on AWS

Want more? Come to these sessions
Thank you!

Nancy Wang  
@nancyzwang

Satish Lakshmanan  
@satish8100