VIRGINIA
IT AGENCY

Commonwealth of Virginia

Enterprise Architecture Standard (EA-225)

Enterprise Solution Architecture [ESA]

Data Availability

Vita.virginia.gov   February 15, 2023

Revision History

Data Availability: Version History		
Revision	**Date**	**Description**
1.0	04-11-2022	Initial document created
1.1	02-12-2023	*Rationale:* To provide for the distinction between public cloud-based SaaS provider requirements and on premise (QTS Richmond) private cloud storage and backup service support providers. Minor reorganization of content sections. *Definition Updates:* Backup Service Provider, Data replication, Recovery Time Objective (RTO), Recovery Point Objective (RPO), Sensitive as to Integrity *Service provider distinctions added:* DA-01, DA-02, DA-04, DA-08, DA-09, DA-12, DA-13, DA-14, DA-15, DA-16, DA-17, DA-18, DA-19, DA-20, DA-21, DA-22, DA-23, DA-24, DA-27, DA-29, DA-32, DA-33, DA-40, DA-41 *Added Data Availability subsections:* **Backup Exceptions** DA-54, DA-55 **Data Replication and Backup Requirements** **Data Risk Classification** DA-56, DA-57 *(Integration/Interoperability) added requirement for SaaS and IT solution providers:* DA-58

Review Process

This requirements document was posted on the Virginia Information Technologies Agency's (VITA) Online Review and Comment application (ORCA). All agencies, stakeholders, and the public were encouraged to provide their comments through ORCA. All comments were evaluated, and individual commenters were notified of action(s) taken.

Standards and Agency Exceptions

These standards are incorporated within the COV Enterprise Architecture Standard (EA-225), and the requirements defined within this document are mandatory for Executive Branch agencies. Agencies deviating from these requirements must request an exception for each desired deviation, and receive an approved *Enterprise Architecture Exception* via Archer, prior to developing, procuring, or deploying such technology, or not complying with a requirement specified in this document.

Glossary

As appropriate, terms and definitions used in this document are in the COV ITRM IT Glossary. The COV ITRM IT Glossary is available on the VITA website.

Contents

Introduction

The commonwealth's data is critical to supporting the business of the commonwealth. Data Availability ensures that data is available to end users and applications, when and where it is needed.

Data Availability Vision and Strategy
Vision
Access to commonwealth data will not be adversely impacted beyond acceptable levels by interruptions in the availability of the IT services, platforms and applications that support that business. Interruptions, if they do occur, will be measured, reported, and will be within predefined agreed upon maximum durations.
Strategy
Objective 1: Commonwealth data must be protected from loss, no matter where it is or how it is stored **Objective 2:** Services will minimize data loss and ensure recoverability of data **Objective 3:** Since not all data loss is preventable, the commonwealth must be prepared for mitigation for loss of data **Objective 4:** Non-DR restore of data must be available for all use cases (hardware failure, data corruption, data theft/encryption, stolen asset, etc.)

Purpose

The intent of these requirements is to guide the purchase, design, implementation, and on-going operation of COV IT services and utilized technologies. For further information on the perspectives, please reference the most recent version of the Enterprise Solution Architecture (ESA) Requirements document.

Authority

- Code of Virginia, §2.2-2007. Powers of the CIO
- Code of Virginia, §2.2-2007.1. Additional duties of the CIO relating to information technology planning and budgeting
- Code of Virginia, §2.2-2009(A). Additional duties of the CIO relating to security of government information
- Code of Virginia, §2.2-2012(A). Additional powers and duties related to the procurement of information technology

Scope

This standard is applicable to all Executive Branch state agencies (hereinafter collectively referred to as "agencies") that are responsible for the management, development, purchase and use of information technology resources in the Commonwealth of Virginia. This standard does not apply to research projects, research initiatives, or instructional programs at public institutions of higher education.

The Data Availability scope covers data backup architecture, backup capture, backup retention, recovery/restoration planning, and testing of backup and restore operational performance.

The requirements in this document apply to:

- Agencies who manage storage services
- Agencies who contract with suppliers that provide storage services
- Suppliers and managers of COV storage
- Cloud service providers
- Third party service partners offering SaaS, BaaS, and DRaaS solutions

This and related documents are located on the EA Library web page:
https://www.vita.virginia.gov/it-governance/enterprise-architecture/ea-library/

There are eight perspectives in this document:

1. Solution Business Requirements
2. Design/Architecture
3. Availability/Performance
4. Capacity
5. Continuity
6. Integration/Interoperability
7. Technology
8. Security

This and related documents are located on the EA Library web page:
https://www.vita.virginia.gov/it-governance/enterprise-architecture/ea-library/

Solution Business Requirements

DA-01 All COV SaaS and storage service providers shall include data backup and restoration solutions capable of meeting the customer recovery objectives.

Note: this includes local attached storage associated with physical servers

DA-02 All COV SaaS and storage service providers shall assist in the resolution of errors and recovery of lost COV data.

DA-03 Backup schedules shall be maintained in alignment with recovery time and recovery point objectives. These schedules shall be published and be available to agency customers.

DA-04 All COV SaaS and storage service providers shall provide backups of changed production data at least once every 24 hours.

DA-05 All COV on premise storage backups shall be cataloged and the catalog shall be available to the customer agencies.

DA-06 Backups that are not executed per the approved schedule shall be documented. That documentation shall be accessible by the customer agencies.

DA-07 Storage services and technologies used for backups shall be capable of completing scheduled backups within a 12 hour window.

DA-08 COV on premise storage services and technologies shall be capable of performing a backup at any time and of any storage device unless documented otherwise

DA-09 COV on premise storage backup and restore services shall be capable of recovering and restoring captured data backups of any system within the enterprise.

DA-10 Backups shall not be relied on for records retention.

Note: Backups are done to ensure data availability, not for records retention. Attempting to use backups for records retention risks non-compliance with retention requirements under the Virginia Public Records Act and the related regulations and schedules. Agencies who have questions about backups and records retention should consult with their counsel and their assigned Library of Virginia records analyst.

DA-11 Automated reports for backup job completion, failure or other issues shall be made available to agencies.

DA-12 COV on premise storage backup and restore services shall include the capability to backup and restore open files.

DA-13 COV on premise storage backup and restore services shall provide self-service capabilities to allow agency customers to request/perform backup and recovery of data.

DA-14 COV on premise storage backup and restore services shall include a reporting dashboard that shows backup inventory, status, captures performance, identifies the system associated with each backup, and can search the backup repositories for files and data.

- Backup reports shall include the size of the backup as well as system specific increases or decreases in usage.
- The dashboard shall include an inventory of all servers and storage that is to be covered by the backup processes.

DA-15 COV on premise storage backup and restore services shall protect system state including, custom configurations, and application deployments.

DA-16 COV on premise storage backup and restore services shall provide support for database backup automation, self-service and reporting for all COV database instances.

DA-17 COV on premise storage backup and restore services shall configure database backups and associated dependent files to be written to designated external storage devices.

Examples: Transaction Logs, Archive logs

DA-18 All COV SaaS and storage providers shall provide backup and restore solutions that are resilient to the failure of local production backup storage.

Backup Exceptions

DA-54 If COV data is not being backed up by the supplier of the application or storage service, the consumer (Agency) shall be responsible for providing all backup requirements to comply with Data Availability requirements.

DA-55 If the consumer data can be “re-constructed” (as opposed to backup) and restored within consumer recovery objectives (RTO and RPO), an EA Exception shall be submitted describing the proposed design of the alternate data recovery/restore solution.

Design/Architecture

DA-19 All COV SaaS and storage providers shall provide for and maintain at least three (3) distinct data copies of all captured data. (Includes live production data set, plus two backup copies).

DA-20 COV on premise storage backup and restore services shall utilize data replication (or other viable technology solution) of COV production data to achieve required RPOs that are less than 24 hours.

DA-21 All COV SaaS and storage providers shall apply the COV 4-2-1-1 Backup Rule to data tagged Sensitive as to Availability or Integrity.

- First copy is production data
- Second copy must be offsite (additional copies can also be offsite)
- Third copy must be in a COV Data Vault (cyber protected)
- Fourth copy must use a distinct media backup technology (vendor backup solution) from other three copies

DA-22 COV on premise storage backup and restore services shall provide the capability to recover virtual servers (VMs) efficiently during complex events (failures that require large-scale recovery).

Example: snapshot storage infrastructure could enable Instant Access to allow for scaled-up recovery

Availability/Performance

DA-23 All COV on premise storage backups (including data safeguarding) shall be captured at times to accommodate high service demands and shall utilize techniques such as backup intervals and network and CPU throttling to ensure backup capture does not adversely affect performance of COV IT services (network, storage performance, etc.).

- Production server average CPU utilization during backups shall not exceed 90%
- Network WAN agency business use bandwidth will be maintained at 65% or greater of WAN capacity during backups
- Production storage shall never exceed 90% of maximum performance of the storage technology

DA-24 COV on premise storage services shall perform continuous automated testing for data corruption. Any corruption shall be logged and the customer shall be notified when the solution does not automatically correct the corruption.

DA-25 All failed backups shall be re-run until successful completion of the backups. All data backup failures shall be fully remediated within 24 hours.

DA-26 Any backup failures not remediated within 24 hours shall be escalated via incident ticket to Severity 2.

DA-27 COV on premise storage backup and restore performance testing shall be performed using different database sizes. This includes: Small < 500GB, Medium = 500GB to 1TB, and Large > 1TB+ databases. Testing reports shall be submitted to VITA annually and they shall include:

- Actual time needed to complete each database size restore test.

DA-28 Offsite copies of backups shall be brought up-to-date to the state/content of current primary backup within 24 hours of primary backup completion.

Capacity

DA-29 COV on premise storage backup and restore technologies and services shall be sized such that the restoration of data will occur within the defined RPO/RTO in accordance with service level agreements (SLAs).

DA-30 All data backups must be retained for a minimum of 6 months.

DA-31 Any data backups requiring over 6 months retention shall be cataloged and stored on separate archival storage.

Note: this separate archival storage is expected to be at a lower rate

Continuity

DA-32 COV on premise storage backup and restore services shall demonstrate that the backup, recovery and restore processes achieve Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) by annual testing. Testing results shall be submitted to VITA for review.

DA-33 COV on premise storage backup and restore services shall support shared storage RPOs of 4 hours; 24 hours, 72 hours, 15 days and 30 days. The default RPO for shared storage services is 24 hours.

Note: See Appendix III for the RPO distributions that support those RPOs

DA-34 For organizations using shared storage the lowest specified RTO shall be used. If an RTO is not specified, the target shall be 24 hours.

DA-35 Restoration from backup shall be completed within 24 hours of request.

DA-36 All COV captured backup data types shall have a corresponding recovery and restore plan.

Note: Captured backup data types are VMs, files/file shares, and databases

DA-37 Recovery and restore plans for all backup and restore technologies/solutions shall be tested monthly and reported to VITA. Testing results shall include:

- Target system(s) for applicable recovery and restore plans and measured time for the restore of the system(s).

DA-38 Additional sites used for offsite (or secondary site) data replication and backup shall follow the geodiversity requirement of 400 or more miles from the primary site in Virginia.

Note: See Appendix II for geodiversity requirement rationale

DA-39 All data assets tagged with “Sensitive as to Availability or Integrity” in Archer shall be protected by a COV Data Vault.

- Backups of “Sensitive as to Availability or Integrity” data assets shall be catalogued in the backup management system.
- Restores of “Sensitive as to Availability or Integrity” data assets shall be given priority during a recovery and restore event.

Example: Sheltered Harbor certified safeguarding solution.

Integration/Interoperability

DA-40 All backup and restore services shall include APIs, web services, or other integration methods that allow agencies to integrate their processes with data backup and restore services.

DA-58 All SaaS and suppliers of IT solutions shall provide unstructured production COV data to the customer in an industry-standard format on request. (e.g., .doc, .xls, .pdf, logs, and flat files).

Technology

DA-41 COV on premise storage backup and restore services shall leverage space saving technology such as deduplication and compression. Metrics shall be supplied to VITA annually on the effectiveness of the space saving methodologies.

Note: use of these technologies should reduce COV storage costs

DA-42 Data corruption risk shall be addressed by backup technologies that utilize a different media than what is used for the production data.

Note: sole reliance on storage replication of production data does not meet this requirement

DA-43 The COV Data Vault shall be a safeguarding solution that is certified by Sheltered Harbor or meet Data Vault requirements in this document.

DA-44 The COV Data Vault solution shall identify ransomware residing in existing backups using the latest virus definitions.

Note: if the virus definitions are updated then rescans may be required.

Security

DA-45 All backup and restore services/devices shall provide security access controls.

DA-46 The COV Data Vault solution shall include the capability for immutable backups.

DA-47 The COV Data Vault solution shall reside outside of the COV production network, with no external IP link to its vault.

DA-48 The COV Data Vault solution shall encrypt all data managed by the solution.

DA-49 The COV Data Vault solution shall manage data through distinct access accounts.

DA-50 The COV Data Vault solution data shall be physically air-gapped.

DA-51 The COV Data Vault solution shall leverage automated software policies used to encrypt and synchronize data through replication, breaking the link between each "air gap".

DA-52 The COV Data Vault solution shall manage vault replication as well as continuous security and integrity analysis of idle data.

DA-53 The COV Data Vault solution shall leverage AI for automated early detection of ransomware.

Data Risk Classification

Sensitive as to Availability
If any outage or data loss of the system introduce significant cost and/or significant impact to the agency mission, then Availability risk is present. If COV data is found to have dependent services and/or data that could magnify the duration of an outage, then Availability risk is present. If a service is found to have a lack of architectural transparency and/or data access, Availability risk is present. Availability risk is measured by the existence dependencies, a lack of end-to-end architecture artifacts, and the size of estimated loss (cost) or damage to the Commonwealth, Agencies, or Individuals. -Reference that this has been identified in Archer

DA-56 All COV data classified as "Sensitive as to Availability" shall apply the **COV 4-2-1-1 backup rule**.

Sensitive as to Integrity Risk
If unauthorized modification or destruction of managed data is determined to have a serious adverse effect on organizational operations, organizational assets, or individuals, then Integrity Risk is present and mitigation is required. –Reference that this has been identified in Archer

DA-57 All COV data classified as "Sensitive as to Integrity" shall apply the **COV 4-2-1-1 backup rule**.

Appendices

Appendix I: Considerations

Business Availability Considerations

Mission Criticality

Mission critical IT systems and applications provide essential IT functions and access to data whose unavailability will have an immediate and significant detrimental effect on COV and Agencies if the system fails or is interrupted. A system or application may be designated mission critical if it meets one or more of the following conditions:

1. Risk to human and research-animal life or safety
2. Significant impact on the Commonwealth IT
3. Significant legal, regulatory or financial costs
4. Serious impediment to VITA in carrying out its critical business functions within the first 48 hours following an event (48 hour Recovery Time Objective – RTO)
5. Loss of access to data with defined availability requirements

Critical Business Functions

Critical operational and/or business support functions that cannot be interrupted or unavailable for more than a mandated or predetermined time frame without significantly jeopardizing COV operations.

Compliance Regulations

Consider what data is valuable to remain in compliance with various agency regulations and requirements.

Confidentiality

Consider what data is public, proprietary or confidential.

Recovery Time Objective (RTO): The duration of time within which a business process must be restored and a stated service level achieved following a disruption in order to avoid unacceptable consequences associated with a break in service.

Recovery Point Objective (RPO)*:* The maximum tolerable period in which data might be lost from an IT system or service due to a major incident. RTO and RPO timeframes for each criticality level are listed in Table 1 below.

Data Restore Testing Considerations (Databases)

Database restore testing

Test database restores from local disk and data backup storage devices (or cloud storage).

Validating restores where possible

Validating backups using the “restore validate database” command will do everything except actually restore the database. This is the best method to determine if the backup is good and usable before being in a situation in which it becomes critical.

Refreshing nonproduction databases from production backups
Periodically build non-production or “staging” databases from production backups using appropriate backup/restore utility commands as a restore practice. *(Note: security requirements may vary, check with security to determine appropriate methods for building environments with production data)*

Performing monthly restore testing
Explain the process through a narrative, preserve logs and take screenshots of each step taken to recovery and restore backups.

Actual restores
During actual restores, the DBA should back up the database before doing the restore. Depending on the type of loss and backups available, the DBA must decide on whether to go for complete (point-in-time) or incomplete recovery. Incomplete recovery can be time- based, cancel-based or change-based.

Strategy to recover from database corruption—For Oracle databases, the DBA can turn on block checking using appropriate parameters to detect the presence of corrupt blocks in the database. This has a slight performance overhead but will allow early detection of corrupt blocks caused by underlying disk, storage system or input/output (I/O) system problems. By default, RMAN also checks for corrupt blocks during backup. In later versions of Oracle, RMAN can be used to repair corrupted blocks in the database.

Data Restore Testing Considerations (Documents)

Example: Restore individual or multiple files or folders to the original location or an alternate location.

Procedure

1. Log in to the Avamar Administrator console.
2. Click Backup and Restore.
3. Select the Restore tab, and then select clients.
4. Select the client that contains the file(s) or folder(s) to be restored.
5. Select the date from which the files are to be restored from.
6. Select the backup number from which the files are to be restored from.
7. Expand the file structure and navigate to the files to be restored.
8. Check the box next to the files to be restored.
9. Click Action and select Restore Now.
10. To restore the files to the original location, leave Restore Destination Client as the default. To restore the files to an alternate location, click Browse.
11. (NOTE)The alternate location system must have the Avamar client installed, and be registered to this Avamar system.
12. Under Restore Destination Choices select the appropriate destination for the restore and click OK.

Data Restore Testing Considerations (Virtual Machines)

Example: Perform a file level restore from an Avamar image backup using the Avamar Administrator Console.

Procedure

1. Launch and log in to the Avamar Administrator console.
2. Click Backup and Restore launch.
3. Select the Restore tab.
4. Expand vCenter server client and click Virtual Machines.
5. Select the client that contains the files that you want to restore.
6. Select the date to use when restoring the files.
7. Select the backup number to use when restoring the files.
8. Under Contents of Backup named…, select Browse for Granular Restore.
9. At the Proxy Selection Prompt, click OK.
10. Expand the file structure and locate the files that you want to restore.
11. Check the box next to the files that you want to restore.
12. Click Action and select Restore Now.
    a. To restore files to the original location, leave Restore Destination Choices as the default.
    b. To restore files to an alternate location, select Restore everything to a different location.
13. Expand the VMware vCenter Server client, click Virtual Machines, and select the VM to restore.
14. At the prompt, type the administrator username and password and click Log on.
15. Under Browse for Folders or Directories, expand the directory or folders to which to restore to and click OK. Alternate location must be a client for image backups.
16. Next to Absolute Destination, click Browse.
17. Click More Options, select Restore Access Control List (ACL) if required, and click OK.

Data Safeguarding Considerations (Ransomware)

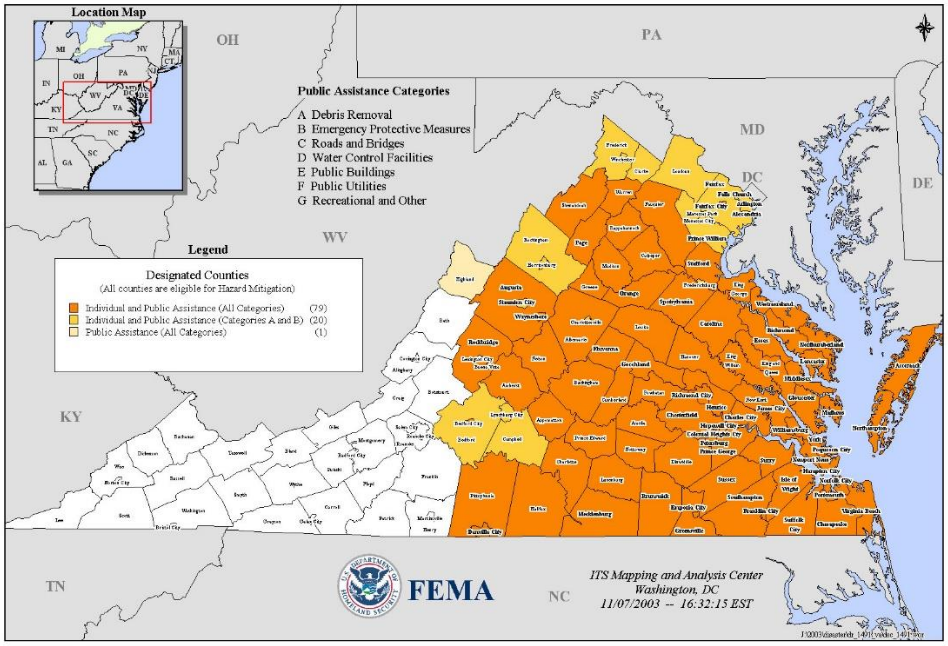
- Take a network segmentation (micro segmentation) and subnetwork isolation approach to limit the scope of an attack.
- Ensure that data backups are performed daily, ideally according to the **COV 4-2-1-1** rule.
- Make sure that a copy of this data is encrypted and stored in an external cyber-recovery vault disconnected from the business network.
- Regularly perform system and data recovery process test exercises monthly.
- Continue to advance and update your security practices based on trend analyses and the evolving nature of attacks.
- Make regular backups of all your sensitive data and systems and store them offline, as ransomware can also encrypt backup files if it can reach them.
- Maintain a complete and current inventory of all your servers, workstations, access points, cyber security devices and other business equipment, including their network addresses, so you can quickly find the source of an attack and isolate it.
- Implement the **COV 4-2-1-1** protection strategy so that you have flexibility options as to how far back in time (e.g., recovery points) you can recover from.
- Protect your data protection environment; back up your backup or data protection tools, metadata and settings.
- Supplement your DR strategy with an isolated recovery capability using a cloud-based data protection platform to secure a copy of critical data offsite and disconnect it from the network

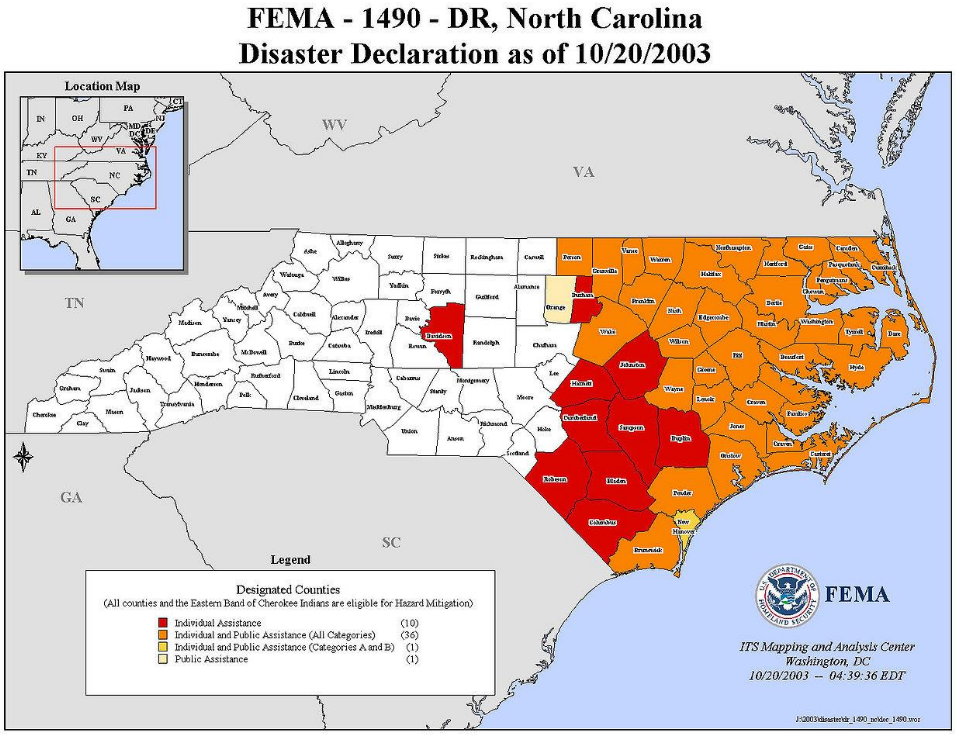
Appendix II: COV Geodiversity Requirement Rationale

COV data must meet a higher level of required availability. This requires the data centers that contain COV data to be protected from external and internal threats. These threats include: natural disasters, theft, terrorism, and other events that could cause damage or loss to COV.

Rationale for distance between primary and secondary locations of commonwealth data
Criteria: worst storm in last 50 years. The worst hurricane (going back to 1954) to impact central Virginia was Isobel in 2003 with 73 mph winds recorded in Richmond. In Virginia, Isobel "knocked out power to over 2 million households. Its combination of strong winds and heavy rain killed 36 people and caused an estimated $1.85 billion in damage."
These are FEMA maps showing Virginia and North Carolina counties where a Federal Disaster was declared:

FEMA - 1491 - DR, Virginia
Disaster Declaration as of 11/07/2003


Location Map
MI
NY
MA
CT
OH
PA
NJ
IN
WV
DE
KY
VA
TN
NC
SC
AL
GA
OH
PA
MD
DC
DE
WV
KY
TN
NC
Public Assistance Categories
A Debris Removal
B Emergency Protective Measures
C Roads and Bridges
D Water Control Facilities
E Public Buildings
F Public Utilities
G Recreational and Other
Legend
Designated Counties
(All counties are eligible for Hazard Mitigation)
Individual and Public Assistance (All Categories) (79)
Individual and Public Assistance (Categories A and B) (20)
Public Assistance (All Categories) (1)
FEMA
ITS Mapping and Analysis Center
Washington, DC
11/07/2003 -- 16:32:15 EST



FEMA - 1490 - DR, North Carolina
Disaster Declaration as of 10/20/2003

Location Map
WV
VA
TN
GA
SC
Legend
Designated Counties
(All counties and the Eastern Band of Cherokee Indians are eligible for Hazard Mitigation)
🟥 Individual Assistance (10)
🟧 Individual and Public Assistance (All Categories) (36)
🟨 Individual and Public Assistance (Categories A and B) (1)
□ Public Assistance (1)
FEMA
ITS Mapping and Analysis Center
Washington, DC
10/20/2003 -- 04:39:36 EDT
J:\2003\disaster\dr_1490_nc\dec_1490.wor

Looking at the “Individuals and Public Assistance (All Categories) county. The maximum distance of impact was from Warren County (Belleview), Virginia to Brunswick County (Calabash), North Carolina (400 miles).

Based on primary location of commonwealth data within Virginia and the impact of the worse hurricane in the past 50 years***, the secondary (off-site) backup data location must be 400 miles or more from the primary data center location in Virginia***.

Appendix III: Agency Application RPO Distributions

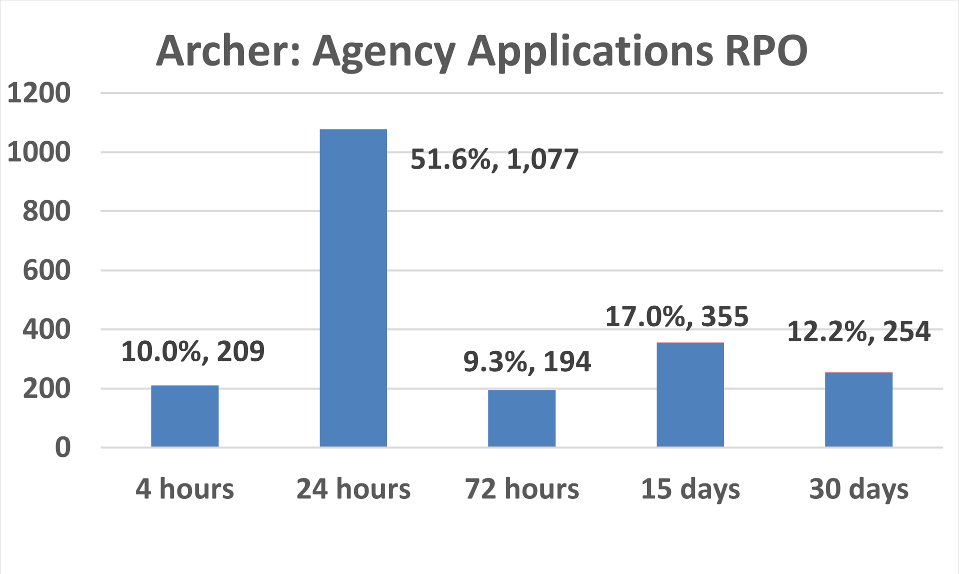
Archer contains the RPOs for 2,089 active applications.
The distribution of those RPOs is as follows:

RPO: hours	Counts	Cumulative Percent	Percent
0.25	4	0.2%	0.2%
0.5	50	2.6%	2.4%
1	47	4.8%	2.2%
2	4	5.0%	0.2%
4	104	10.0%	5.0%
5	6	10.3%	0.3%
6	10	10.8%	0.5%
8	98	15.5%	4.7%
12	22	16.5%	1.1%
15	1	16.6%	0.0%
20	2	16.7%	0.1%
24	938	61.6%	44.9%
30	2	61.7%	0.1%
36	1	61.7%	0.0%
40	1	61.8%	0.0%
48	59	64.6%	2.8%
52	1	64.6%	0.0%
60	5	64.9%	0.2%
72	125	70.8%	6.0%
96	12	71.4%	0.6%
120	72	74.9%	3.4%
144	26	76.1%	1.2%
168	61	79.0%	2.9%
169	11	79.6%	0.5%
216	28	80.9%	1.3%
240	10	81.4%	0.5%
288	4	81.6%	0.2%
330	1	81.6%	0.0%
336	121	87.4%	5.8%
360	9	87.8%	0.4%
480	2	87.9%	0.1%
504	1	88.0%	0.0%
720	88	92.2%	4.2%
744	3	92.3%	0.1%
1,200	1	92.4%	0.0%
1,440	4	92.6%	0.2%

RPO: hours	Counts	Cumulative Percent	Percent
2,160	2	92.7%	0.1%
2,880	2	92.8%	0.1%
3,000+	151	100.0%	7.2%

Recommendations for services that ensure that the data is available to meet those RPOs follows:

Archer: Agency Applications RPO			
RPO	Counts	Cumulative Percent	Percent
4 hours	209	10.0%	10.0%
24 hours	1,077	61.6%	51.6%
72 hours	194	70.8%	9.3%
15 days	355	87.8%	17.0%
30 days	254	100.0%	12.2%



Appendix IV: Compliant Data Availability Examples

Public cloud

Data not sensitive to availability or integrity: 3-2-1

Copy 1: Production
Copy 2: Backup copy of production (example: Azure Backup, AWS Backup, 3rd party cloud backup)
Copy 3: Off-site geo-diverse replication

Data sensitive to availability or integrity: 4-2-1-1

Copy 1: Production
Copy 2: Backup copy of production (example: Azure Backup, AWS Backup, 3rd party cloud backup)
Copy 3: Off-site geo-diverse replication
Copy 4: Must meet COV Data Vault service requirements

On-premise

Data not sensitive to availability or integrity: 3-2-1

Copy 1: Production
Copy 2: Second methodology: Backup via on-site Avamar
Copy 3: Off-site geo-diverse replication of Avamar

Data sensitive to availability or integrity: 4-2-1-1

Copy 1: Production
Copy 2: Second methodology: Backup via on-site Avamar
Copy 3: Off-site geo-diverse replication of Avamar
Copy 4: Must meet COV Data Vault service requirements

Appendix V: Data Replication and Backup Guidance

A common data/storage “durability” characteristic is data replication between storage devices and data centers creating multiple “instances” of production data. Durability plays a key role in SLA compliance, and the availability of the production data set. Data replication is a valid approach for achieving RPOs of less than 24 hours and site recovery in the event of a declared data center disaster recovery event. **For compliance with Data Availability Requirements, daily point-in-time backups are required since replicated “instances” of production data can be directly impacted by the integrity and viability of the primary production data set.** Disconnected point-in-time backups mitigate this type of risk for COV data.

Appendix VI: Definitions and Terminology

3-2-1 Backup Rule – a fundamental backup methodology that requires at least three total copies of data (**3** includes: one live production copy and two backup copies), **2** of which are local but on different mediums/devices represents two different media/methodologies, **1** represents a single off-site copy to support disaster recovery.

COV 4-2-1-1 Backup Rule – a modernized version of the 3-2-1 backup rule, that provides additional safeguards and recovery options in the event of a ransomware attack or other hardware media failure. This includes at least **4** copies of data (including production), utilizing **2** backup methodologies/media types, **1** offsite copy for DR, and **1** offline cyber resilient backup.

Artificial Intelligence (AI) – refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.

BaaS – Backup as a Service

Backup retention – the act of retaining backups over a determined period of time, to provide for optional recovery points for restoring a system or data set.

Backup Service Provider – provides backup services to an application, service or data center infrastructure that uses software to capture backups of data.

Block-Level backup – performs an in-depth file analysis and copies only the modified parts of files making it much faster than file-level backup.

Business Availability – a tag used in Archer used to identify systems that require a high level of data availability and/or uptime.

COV Data Vaulting – the commonwealth's data safeguarding architecture and methods for providing a cyber-resilient backup service. The COV Data Vault service will be secure from ransomware attacks. The backups within the vault will be either off-line, air-gapped or will be otherwise isolated from the COV network in such a way that the backup is unable to be changed (immutable).

Cyber Resilience – the ability of an organization to enable business acceleration (enterprise resiliency) by preparing for, responding to, and recovering from cyber threats. A cyber-resilient organization can adapt to known and unknown crises, threats, adversities, and challenges. Critical data is routinely monitored and tested for integrity, and recoverability.

Cyber Resilient Backup – a backup of critical data often “air-gapped” that is routinely monitored and tested for integrity, and recoverability

Cyber security – practice of defending computers, servers, mobile devices, electronic systems, networks, and data from malicious attacks. It's also known as information technology security or electronic information security.

DBMS (Database Management System) – software (associated files and executables) used to support the storage and retrieval of structured data.

Data archiving – generally refers to long-term storage of data that is no longer in regular use but can be restored if need be.

Database backups – database backups are backups of structured data created by the database system (e.g. Oracle, SQL, etc.) and written to storage. Database backups can only be recovered by the database system that created the backups.

Data replication – is the process by which data residing on a physical/virtual server(s) or cloud instance (primary instance) is continuously replicated or copied to a secondary server(s) or cloud instance (standby instance). Replicated data supports high availability, backup, and/or disaster recovery.

Data retention – the capability of customers to retain non-current data to meet business requirements. Data retention requirements depend on the content of the data, whether the contents are classified as public records, and agency policies and business requirements

Data safeguarding – the application of industry-standard safeguards against the destruction, loss, misuse, unauthorized disclosure, or alteration of the data or confidential information, and such other related safeguards that are set forth in applicable laws, a statement of work, or pursuant to court policies or procedures.

Differential backup – uploads any new and updated files after the last full backup. Each consequent differential backup compares datasets only with the last full backup.

DRaaS – Disaster Recovery as a Service

File-based copy – simple copies of files. Most widely used form of backup worldwide Simple Good for small organizations

File-Level backup – if a file has been modified, it will be sent to the backup repository to create a new version of it. This backup type is simple to perform and works well for a small dataset.

Full backup – a full self-contained copy of the data in question.

Geodiversity – short for geographic diversity — in the context of data centers refers to the distance between two or more facilities. Recommended distance is 400+ miles, distance of 100-400 miles is allowable under certain circumstances.

Incremental backup – a backup that copies only data that was changed since the previous backup (full or incremental). Every next backup will include only files that were changed since the most recent backup. Incremental backups require less storage space and network utilization. There are two types of incremental backups:

Instant Access – is the ability to boot a VM directly from the Data Domain appliance, to decrease downtime and provide efficient backup validation.

Labeling – proper labeling of all back-up media is critical to the timely restoration of data. There must be a correlation between the data, the media and the date the back-up was performed. The use of a back-up log would indicate the media set that contains the file(s) to be restored. The method used will be based primarily on the business objectives.

NAS snapshot – technology that tracks changes to files and keeps copies of changes so that an earlier version of the file can be recreated on demand.

Physical vs Virtual (OS software) – an event such as a hardware failure will require a complete system restore, starting with the OS, so there is a need to back up the database server OS initially and after any system updates or configuration changes.

Physical offline or cold backups – the database must be shut down and a copy must be made of all essential data files and other components of the database.

Physical online or hot backups – this method enables the database to be backed up while the database is up and running. The following points should be kept in mind while doing online backups:

Policy administrator (PA) – this component is responsible for establishing and/or shutting down the communication path between a subject and a resource. It would generate any authentication and authentication token or credential used by a client to access an enterprise resource. It is closely tied to the PE and relies on its decision to ultimately allow or deny a session. Some implementations may treat the PE and PA as a single service; here, it is divided into its two logical components. The PA communicates with the PEP when creating the communication path. This communication is done via the control plane.

Policy engine (PE) – this component is responsible for the ultimate decision to grant access to a resource for a given subject. The PE uses enterprise policy as well as input from external sources (e.g., CDM systems, threat intelligence services described below) as input to a trust algorithm to grant, deny, or revoke access to the resource. The PE is paired with the policy administrator component. The policy engine makes and logs the decision, and the policy administrator executes the decision.

Policy enforcement point (PEP) – this system is responsible for enabling, monitoring, and eventually terminating connections between a subject and an enterprise resource. This is a single logical component in ZTA but may be broken into two different components: the client (e.g., agent on user's laptop) and resource side (e.g., gateway component in front of resource that controls access) or a single portal component that acts as a gatekeeper for communication paths**.**

Private Cloud – is defined as computing services offered either over the Internet or a private internal network and only to select users instead of the general public. Also called an internal or corporate cloud, private cloud computing gives businesses many of the benefits of a public cloud - including self-service, scalability, and elasticity - with the additional control and customization available from dedicated resources over a computing infrastructure hosted on-premises.

Public Cloud – is defined as computing services offered by third-party providers over the public Internet, making them available to anyone who wants to use or purchase them. They may be free or sold on-demand, allowing customers to pay only per usage for the CPU cycles, storage, or bandwidth they consume.

Public Key Infrastructure (PKI) – this system is responsible for generating and logging certificates issued by the enterprise to resources, subjects, and applications. This also includes the global certificate authority ecosystem and the Federal PKI,4 which may or may not be integrated with the enterprise PKI. This could also be a PKI that is not built upon X.509

certificates.

Ransomware – malicious software designed to gain access to files and encrypt data by generating a private-public pair of keys. The data is impossible to decrypt without the private key which is retained by the attacker's server until the ransom is paid.

Retention period – the amount of time to keep the backup, in days, weeks, months, or years.

Recovery Time Objective (RTO) – The duration of time within which a business process must be restored and a stated service level achieved following a disruption in order to avoid unacceptable consequences associated with a break in service.

Recovery Point Objective (RPO) – The maximum tolerable period in which data might be lost from an IT system or service due to a major incident.

SaaS – Software as a Service

Sensitive as to Availability – application distinction in Archer that includes a calculated monetary value based on data sets and/or business processes.

Sensitive as to Integrity – data classification in Archer that pertains to the risk of unauthorized modification or destruction of managed system data and the resulting impact to business processes and/or individuals.

Server backups – server backups are backups of storage/data attached to servers. Server backups are created daily by the backup system for operational recovery

Server data – server backups are taken daily for operational recovery. They can be used to recover server OS and application components in case of a data loss event. Server backups are data agnostic and all types of data on the server are treated the same. It is possible to recover a file from a backup but not possible to delete a file from a backup. Server backups should be retained for sufficient time for operational recovery.

SRDF – Symmetrix Remote Data Facility, is a replication product which can be used to replicate the data from one array to second array. Primary use is for business continuity/disaster recovery.

Structured data (databases) – structured data is organized and stored in specific formats by the Database System. The server backup contains a copy of the source database files but they are not directly usable except by the database system and may not be transaction consistent. In addition to the server backups, databases need to be backed up using the database software. Scheduling and retention of database backups should be controlled by Database Administrators based on the type of data, agency requirements and the reason for the backup.

Synthetic full backup – a type of subsequent full backup that makes a comparison to the previously backed up data on the storage and uploads only the current changes from the backup source. Synthetic full backup helps to reduce the amount of data uploaded and accelerates a full backup creation.

Unstructured data, non-modifiable – unstructured data is stored as discrete files with no specific organization or relationships between the files. Non-modifiable data is generally images, video stream files, music stream files, etc. They are written and stored but not modified or edited after writing. Often the data is storage and managed by an application. Non-modifiable unstructured data can also include data types that would normally be modifiable but are being kept for data retention purposes, so modification is prohibited by policy.

Unstructured data, modifiable – unstructured data is stored as discrete files with no specific organization or relationships between the files. Modifiable unstructured data is files that are created and edited. For example, Word documents, Excel spreadsheets, notepad files, PowerPoint presentations etc. They are routinely updated / overwritten with new versions.

VM snapshot – a backup type that generates a point in time copy of the virtual machine (vmdk file).

Zero Trust – a cybersecurity paradigm focused on resource protection and the premise that trust is never granted implicitly but must be continually evaluated