STUDENT HANDOUT · WEEK 10

Protecting Data, Resilience, and Recovery

SY0-701 objectives 3.3, 3.4. Notes, exercises, review questions, and the lab.

Name	
Date	

THIS WEEK YOU WILL

1. Identify data types, from regulated data to trade secrets, and what makes each sensitive.
2. Apply the six data classification labels to an organization's data.
3. Match a protection method to each data state: at rest, in transit, and in use.
4. Compare encryption, hashing, masking, tokenization, and obfuscation and pick the right one.
5. Distinguish load balancing from clustering, and hot, warm, and cold sites.
6. Plan backups that meet a stated RPO and RTO, and choose how to test them.
7. Explain how capacity planning, generators, and UPS support continuity of operations.

KEY IDEAS

Know what kind of data you hold

Data type	What it is	Clinic example
Regulated	Governed by law or binding rules	Patient records under HIPAA
Trade secret	Valuable because it stays secret	Proprietary scheduling method
Intellectual property	Creations protected by copyright or patent	Patient education videos it produced
Legal information	Contracts, case files, legal advice	Malpractice case files
Financial information	Payment card, bank, and billing data	Card numbers under PCI DSS

Data type	What it is	Clinic example
Human, non-human readable	Text people read vs codes and binary	Visit notes vs a barcode or DB file

Labels tell people how to handle data

Public	Approved for anyone; no harm if disclosed. (Office hours, website text)
Private	About individuals; personal and often regulated. (Staff home addresses)
Sensitive	Harm if disclosed; needs handling rules. (PII, PHI)
Confidential	Approved people only; disclosure causes damage. (Payer contracts)
Restricted	Tightest access; named people only. (Encryption keys, legal holds)
Critical	Needed to operate; loss stops the business. (EHR database)

States, sovereignty, and geolocation

Consideration	Meaning	Typical control
Data at rest	Stored on disks, databases, and backups	Full disk or database encryption
Data in transit	Moving across a network	TLS, IPSec VPN
Data in use	In memory and being processed	Access control, secure enclave
Data sovereignty	Subject to the laws where it is stored	Choose cloud regions by law
Geolocation	Where a user, device, or data is	Block sign-ins from unexpected places

Change the data or limit who reaches it

Transform the data	Restrict access
Encryption: reversible with the right key	Permission restrictions: least privilege on data
Hashing: one way; proves integrity	Segmentation: sensitive data in its own zone
Masking: hide part, such as all but the last 4	Geographic restrictions: block by location
Tokenization: swap for a token; a vault maps it	
Obfuscation: make data or code hard to read	

Load balancing and clustering differ

Load balancing	Clustering

Spreads requests across independent servers	Several servers act as one, sharing state
Health checks remove failed servers	Active/passive: a standby takes over
Scales out to handle more users	Active/active: every node shares the work
Fits stateless web front ends	Fits databases and file services

Hot, warm, or cold: pay for speed

Site	What is ready	Recovery time and cost
Hot site	Hardware, current data, network, all live	Minutes to hours; highest cost
Warm site	Hardware and network; data must be loaded	Hours to days; moderate cost
Cold site	Space, power, and cooling only	Days to weeks; lowest cost
Geographic dispersion	Sites far enough apart for separate disasters	Survives regional events

Plan for more than one way to fail

Platform diversity	Different vendors or systems so one flaw cannot stop all. (Two firewall vendors in series)
Multi-cloud systems	Spread services across providers to survive an outage. (Backups in a second cloud)
Continuity (COOP)	Keep essential functions running while systems recover. (Paper check-in during an outage)
Capacity planning	Enough people, technology, and infrastructure for demand. (Cross-train a second admin)

Test the plan before a disaster does

Tabletop exercise	Talk through a scenario; no systems are touched. (Ransomware at the clinic)
Simulation	Act out an event with realistic injects and roles. (Mock outage drill with the staff)
Parallel processing	Run the backup system alongside the primary; compare. (Recovery site processes a day's work)
Fail over	Actually switch production to the backup system. (Move EHR to the standby server)

Backups meet the RPO; restores meet the RTO

Choice	What it means	Watch for
Onsite and offsite	Local copy for speed, remote for disaster	Keep one copy offline or immutable
Frequency	How often a backup runs	Sets the best RPO you can meet
Encryption	Protects backup media and cloud copies	Lost keys mean lost backups
Snapshots	Point-in-time image of a VM or volume	Same storage; not a full backup
Replication, journaling	Copy changes live; log every write	Replication copies corruption too
Generator and UPS	UPS bridges seconds; generator runs hours	Test both under real load

EXERCISE 1: LABEL IT, STATE IT, PROTECT IT

10 MINUTES · PAIRS

For each Lakeside item, give its data type, classification, current state, and the single best protection method.

#	Prompt	Your answer
1	Patient visit notes stored in the EHR database	
2	Card numbers the front desk takes for copays	
3	A lab result sent to a hospital's system	
4	The clinic's hours and phone number	
5	A support vendor viewing a patient chart on screen	
6	Patient data copied for a developer to test new software	
7	Backups kept with a cloud provider in another country	

REVIEW QUESTIONS

1. What is the difference between tokenization and encryption?

2. Name the three data states and one control for each.

3. Of hot, warm, and cold sites, which costs least and takes longest to recover?

4. Why is a VM snapshot not enough on its own as a backup strategy?

5. Backups run every 6 hours and a full restore takes 3 hours. What RPO and RTO can you meet?

KEY TAKEAWAYS

- **Know the data first.** Type and classification decide how hard you protect each data set.
- **Protect every state.** Encrypt at rest and in transit; limit and watch data in use.
- **Buy recovery speed knowingly.** Hot, warm, and cold sites trade cost against the RTO you can meet.
- **Untested backups are hopes.** Time a real restore and compare it with the RPO and RTO.

LAB 10: BACKUP AND RESTORE TEST AGAINST AN RPO AND RTO

60 MINUTES

You need the Linux VM from Lab 1, VirtualBox or VMware, and about 10 GB of free disk space. Work only on lab VMs; step 5 deletes data on purpose.

Steps

1. Set targets for a mock Lakeside records folder: an RPO of 1 hour and an RTO of 30 minutes. Write both at the top of your report, with the time you start.
2. Create test data and record its hashes: mkdir -p ~/records && for i in $(seq 1 200); do head -c 50K /dev/urandom > ~/records/visit_$i.dat; done, then sha256sum ~/records/* > ~/records.sha256.
3. Take an encrypted file backup to a separate location that stands in for offsite (ideally a second virtual disk mounted at /backup): sudo mkdir -p /backup && sudo chown $USER /backup, then tar -czf - -C ~ records | gpg -c -o /backup/records-$(date +%H%M).tar.gz.gpg. Note the time and the passphrase location.
4. In the hypervisor, take a VM snapshot named before-incident and note the time.
5. Simulate the incident: add 10 new files to ~/records, note the time, then run rm -rf ~/records. Start a stopwatch.
6. Restore from the file backup: gpg -d /backup/records-*.tar.gz.gpg | tar -xzf - -C ~, then verify with sha256sum -c ~/records.sha256. Stop the stopwatch when every hash reports OK. That is your measured recovery time.
7. Measure the data loss: list which files are missing (the 10 created after the backup) and how much time passed between the backup and the incident. Compare that with the 1-hour RPO.
8. Repeat the incident, then restore the before-incident snapshot in the hypervisor instead and time it. Compare the two methods for speed and for what each one recovered.
9. Write a backup schedule for the records folder that meets both targets: frequency, onsite and offsite copies, encryption and key storage, and how often to run this restore test.

Hand in

A one to two page report with the targets, timestamps, measured recovery times for both methods, the sha256sum -c result, the measured data loss, and your proposed schedule.

Questions to answer in your report

1. Did the file restore meet the 30-minute RTO? What took the most time, and what would make it faster?
2. How much data did you lose, and what backup frequency would keep the loss inside the 1-hour RPO?
3. Why is the VM snapshot not a substitute for the offsite file backup, even though it restored faster?

4. What would happen to this restore if the gpg passphrase were lost, and where should Lakeside keep its backup keys?
5. What does the hash check prove that a successful tar extraction does not?