

Planning for Disaster

DISASTER PLANNING · RPO & RTO · BCP · AWS DR BUILDING BLOCKS · FOUR DR PATTERNS · WELL-ARCHITECTED

STUDY & REVIEW

01 MUST KNOW

if you remember five things, remember these

- 01 RPO vs. RTO:** **RPO** (recovery point objective) is the maximum acceptable *data loss*, measured in time — you look **back** to the last recoverable data. **RTO** (recovery time objective) is the maximum acceptable *downtime* a business process can remain out of commission — you look **forward** to restoration. Both are set from the financial impact of an outage.
- 02 Four patterns, one spectrum:** **backup & restore** (RPO/RTO in hours, cheapest) → **pilot light** (10s of minutes) → **warm standby** (minutes) → **multi-site** (near real-time, costliest). Cost rises and RTO falls left to right; pick the pattern the business RPO/RTO and budget justify.
- 03 Everything fails — work backward:** Werner Vogels: “*Everything fails, all the time.*” Assume small, large, and global failures **will** happen; design to avoid disasters **and** to recover from them. Always start from the business need. The DR plan is a **subset** of the business continuity plan (BCP).
- 04 DR spans more than one Region:** Scope DR across five service categories — **Storage** (S3), **Compute** (EC2), **Database** (RDS), **Networking & Content Delivery** (VPC), and **Deployment orchestration** (CloudFormation). RPO and RTO guide backup and restore across every one, and shape the production architecture.
- 05 Replication is not a backup:** Replicas (S3 CRR, RDS read replicas) copy the source — **including corruption** — so they give no protection against data corruption. Also **version or snapshot** data for point-in-time recovery, and **test the failover path frequently** (Game Days). The only recovery that works is the path you test.

02 KEY TERMS

TERM	DEFINITION
RPO (recovery point objective)	Maximum acceptable amount of data loss after an unplanned data-loss incident, expressed as an amount of time
RTO (recovery time objective)	Maximum acceptable amount of time after a disaster that a business process can remain out of commission
BCP (business continuity plan)	System of prevention and recovery from threats; includes business impact analysis, risk assessment, the DR plan, and evaluated RPO/RTO
Fault tolerance	Redundancy that lets a system withstand failure of one or more components and keep functioning (often at reduced capacity)
Disaster recovery (DR)	Set of policies and procedures for recovering vital technology infrastructure and systems after any disaster
S3 Cross-Region Replication (CRR)	Automatic asynchronous copy of S3 objects from a source bucket to a destination bucket in another Region; needs an IAM role and retains object metadata
S3 Replication Time Control (S3 RTC)	Replicates 99.99% (4 nines) of new S3 objects within 15 minutes, backed by an SLA
EBS snapshot	Incremental point-in-time backup of an EBS volume stored in S3; can be copied to another Region and automated by Data Lifecycle Manager
Golden AMI	An AMI preconfigured with all applications and services needed to perform its designated function
Pilot light	Pattern where a minimal core (typically the database) is always running; the rest is provisioned around it during recovery
Warm standby	Pattern where a scaled-down but fully functional copy runs 24/7 and scales up on failover

TERM	DEFINITION
Multi-site (active/active)	Pattern where a full-capacity duplicate runs simultaneously in a second Region/site, sharing production traffic via weighted routing

03 THE FOUR DR PATTERNS AT A GLANCE

cost rises · RTO falls, top to bottom

PATTERN	HOW IT WORKS	RPO / RTO	COST
Backup & restore	Data backed up (e.g., to S3) and restored when needed; nothing extra runs — simplest, slowest	Hours	Lowest
Pilot light	Only the critical core (e.g., the database) always runs; provision the rest around it on recovery	10s of minutes	Low
Warm standby	A scaled-down but fully functional stack runs 24/7; scale it up on failover	Minutes	Higher
Multi-site	A full-capacity active/active duplicate runs in a second Region at the same time as the primary	Near real-time	Highest

04 RPO VS. RTO

two different measures — do not swap them

RPO — HOW MUCH DATA <i>look BACK to the last good copy</i> — Maximum acceptable data loss, measured in time — Disaster 8 h after the last backup ⇒ RPO of 8 hours — Drives backup frequency (RPO 8 h ⇒ back up at least every 8 h) — 800 records max loss ÷ 100 records/hour = RPO of 8 hours	RTO — HOW MUCH DOWNTIME <i>look FORWARD to restoration</i> — Maximum acceptable time out of commission after a disaster — Time to restore applications and recover data — Disaster at 10 p.m., RTO 1 h ⇒ restored by 11 p.m. — Drives the recovery strategy (pattern) you choose
---	---

05 AWS DR BUILDING BLOCKS BY SERVICE CATEGORY

DR plans span more than one Region

CATEGORY	EXAMPLE	DR BUILDING BLOCKS
Storage	Amazon S3	S3 CRR & S3 RTC, EBS snapshots, EFS/FSx replication via DataSync, S3 Glacier
Compute	Amazon EC2	EC2 automatic recovery, custom AMIs and golden AMIs
Database	Amazon RDS	Cross-Region snapshots, read replicas, Multi-AZ; DynamoDB PITR & global tables
Networking & Content Delivery	Amazon VPC	Route 53 failover, ELB, Amazon VPN, AWS Direct Connect, EventBridge global endpoints
Deployment orchestration	AWS CloudFormation	Redeploy environments to a new Region/VPC; Elastic Beanstalk, OpsWorks

06 RECOVERING FROM A DISASTER — THE COMMON THREAD

detection → restoration → failover

01 Detect the disaster (Route 53 / CloudWatch health checks) → **02 Restore or provision** infrastructure in the recovery Region → **03 Restore or promote** the data → **04 Fail traffic over** — adjust Route 53 DNS records → **05 Scale** the recovery site to the full production load

- ✗ “RPO and RTO measure the same thing.” — No. **RPO** is how much *data* you can lose (time back to the last good copy); **RTO** is how much *time* the process can be *down* (time forward to restoration).
- ✗ “Replication replaces backups.” — A replica copies the source **including corruption**; S3 CRR and RDS read replicas give no protection against data corruption. Also **version or snapshot** for point-in-time recovery.
- ✗ “Backup and restore recovers fastest because it is simplest.” — Backwards. It is the **cheapest** but has the **highest (slowest) RTO** — restoring from backups takes a long time.
- ✗ “Multi-site is always the best choice.” — It has the least downtime but **costs the most** (a full running duplicate). For a 7-minute RPO that must handle most traffic immediately *cost-effectively*, the answer is **warm standby**, not multi-site.
- ✗ “Pilot light keeps the whole environment running, just small.” — No. In **pilot light** only the *core* (e.g., the database) always runs; **warm standby** is the pattern where *all* components run 24/7 (scaled down).
- ✗ “You can snapshot an EC2 instance store volume.” — You **cannot**. Copy the data to a new EBS volume first (and reconsider why it lives on instance store, which is temporary storage for buffers, caches, and scratch data).
- ✗ “On failover, resize instances bigger (vertical scaling).” — **Horizontal** scaling (adding more EC2 instances behind the load balancer) is **preferred** over vertical scaling (larger instance types).
- ✗ “A DR plan by itself guarantees business continuity.” — The DR plan is only a **subset of the BCP**; aggressive DR targets are pointless if external business elements (e.g., shipping goods after an earthquake) cannot function.

08 SELF-QUIZ

answers are upside-down below

Q1 RPO is the maximum acceptable amount of what, expressed in what unit?

Q2 RTO is the maximum acceptable amount of what after a disaster?

Q3 If 800 records is the acceptable loss and no more than 100 records are created per hour, what RPO does that give?

Q4 List the four common AWS DR patterns from lowest cost to highest.

Q5 Which DR pattern is the most cost-effective, and what is its trade-off?

Q6 Which DR pattern keeps only the critical core (such as the database) always running?

Q7 Which DR pattern runs a full-capacity active/active duplicate and can have the least downtime?

Q8 Which S3 feature replicates 99.99% of new objects within 15 minutes, backed by an SLA?

Q9 Name the five AWS service categories a DR plan spans.

Q10 Sample exam: business-critical apps, max data loss 7 minutes, must handle the majority of traffic immediately then scale to full — which pattern is the most cost-effective fit?

ANSWER KEY (rotate the page)

Q1 data loss; expressed as an amount of time **Q2** time a business process can remain out of commission (downtime) **Q3** 8 hours (800 ÷ 100) **Q4** backup and restore, pilot light, warm standby, multi-site **Q5** backup and restore; the highest (slowest) RTO **Q6** pilot light **Q7** multi-site **Q8** S3 Replication Time Control (S3 RTC) **Q9** Storage, Compute, Database, Networking & Content Delivery, and Deployment orchestration (Management & Governance) **Q10** warm standby