Data Center Availability

OPERATIONAL AVAILABILITY · THE 9s OF UPTIME · MTBF & MTTR · AVAILABILITY VS. RELIABILITY · MAINTAINABILITY · THREATS

STUDY & REVIEW

01 MUST KNOW

if you remember five things, remember these

01 **Availability = readiness to serve: availability** is the readiness of a data center to be accessible for use. **Operational availability** means the data center can perform *all* its expected functions; it is set by the total uptime the facility must support without interference, and the goal is to find the **maximum yearly permissible downtime** and the operational level that goes with it.

02 **SLAs speak in nines:** a **service-level agreement (SLA)** is provided whenever data center space or service is hired; it contractually sets availability and reliability levels. SLAs are usually stated in the **9s of availability** — more 9s means better availability. **Five nines = 99.999% ≈ 5.256 minutes** of downtime per year. Gartner cites an estimated loss of about **$300,000 per hour** of downtime.

03 **Availability ≠ reliability: availability** is the *percentage of time* equipment is in an operational state and is tied to **fault-tolerant** systems that stay up during component failures. **Reliability** is the ability (probability) of equipment to perform its intended function — its durability or quality. Availability is the more **accurate and meaningful** measure of data center performance.

04 **Availability = f(reliability, maintainability): maintainability** is the ease and speed of repair after a failure. With one held constant, raising reliability *or* maintainability raises availability. Server availability is calculated as **MTBF ÷ (MTBF + MTTR)**, where **MTBF** is mean time between failures and **MTTR** is mean time to repair.

05 **Threats cause downtime:** common threats are **power failure**, **human error**, **data center-level failure** (natural disasters), **DDoS attacks** on DNS servers, web applications, and SSL traffic, and **brute force / fragile authentication**. Defenses include out-of-band verification (for example, SMS), centralized authentication, and **application delivery controllers (ADCs)**.

02 KEY TERMS

TERM	DEFINITION
Operational availability	The ability of a data center to perform all expected functions; set by total uptime it must support without interference, and by maximum yearly permissible downtime
Availability	The readiness of a data center to be accessible for use; the percentage of time equipment is in an operational state
Reliability	The time taken for equipment to achieve its intended functionality — the probability a component does its job over a given period; related to durability and quality
Maintainability	The ease and speed of performing repairs to bring equipment back up after a failure; high = fast, easy repair, low = complex, slow repair
Service-level agreement (SLA)	Contract set when data center space or service is hired; fixes agreed availability and reliability levels and holds the data center accountable, usually stated in the 9s of availability
9s of availability	Terminology for required uptime by the number of 9s in the percentage (three 9s = 99.9%, five 9s = 99.999%); more 9s = better availability, less downtime
Uptime	The time a data center is up and running, aiming for 99.999% or higher availability across a year; measured annually as a near-perfect percentage
Downtime	Time a data center is failing or unavailable; scheduled (preplanned maintenance, upgrades, testing) or unscheduled (unpredictable failures or delays)

TERM	DEFINITION
MTBF	Mean time between failures — uptime between consecutive failures of repairable equipment; total service duration ÷ number of failures
MTTR	Mean time to repair — the average time a device takes to recover from a failure; total maintenance time ÷ number of repairs
Five 9s	99.999% availability; considered optimal because it yields very low downtime (about 0.86 second per day, ≈ 5.256 minutes per year)
Application delivery controller (ADC)	Network component placed between the firewall and application servers that blocks threats, inspects encrypted traffic, and prevents unauthorized access

03 THE 9S OF AVAILABILITY

annual SLA downtime — more 9s = less downtime

UPTIME	NUMBER OF 9S	SLA DOWNTIME PER YEAR
99.9%	Three	525.6 minutes
99.99%	Four	52.56 minutes
99.999%	Five (near perfect)	5.256 minutes
99.9999%	Six	0.525 minute
99.99999%	Seven	0.0525 minute

04 AVAILABILITY VS. RELIABILITY

frequently confused — keep them apart

AVAILABILITY

percentage of time in an operational state

- Related to **fault-tolerant** systems
- System stays operational even during component failures
- The more **accurate and meaningful** measure of performance
- UPS fails on a bad inverter, but a backup UPS takes over — availability is preserved

RELIABILITY

does the equipment do its job?

- Related to **durability and quality** of equipment
- Probability a component performs its job over a given period
- Expressed as success ÷ failure over that period
- A UPS with a common inverter failure has **low reliability**

05 AVAILABILITY FORMULAS BY SCENARIO

UT = uptime · DT = downtime

SCENARIO	FORMULA
Server availability (hardware check)	MTBF ÷ (MTBF + MTTR)
Prescheduled maintenance downtime	(UT + prescheduled maint.) ÷ (DT + UT + prescheduled maint.)
Around-the-clock (24/7) availability	UT ÷ (UT + DT)
Reliability of an element	e ^ (–t ÷ MTBF), where t = time in hours

06 FROM FAILURES TO AN AVAILABILITY FIGURE

how the numbers connect

01 A component **fails** → 02 **MTTR** = mean time to repair it → 03 **MTBF** = mean time between failures → 04 **Availability = MTBF ÷ (MTBF + MTTR)** → 05 Stated as the **9s** in the SLA

07 EXAM TRAPS

where the knowledge check gets you

- × “Availability and reliability mean the same thing.” — No. **Availability** is the percentage of time equipment is operational (backed by fault tolerance); **reliability** is whether a component performs its intended function (durability and quality). A system built from unreliable parts can still be highly available if backups take over.
- × “More 9s of availability means more allowed downtime.” — Backwards. More 9s = higher uptime percentage = **less** downtime. Five nines (99.999%) permits only about **5.256 minutes of downtime per year**.
- × “A higher operational-availability level number allows more maintenance time.” — Backwards. **Level 0** has the most yearly maintenance hours (more than 400, least operational); **Level 4** has **zero** hours (most operational).
- × “Reliability is the best measure of data center performance.” — **Availability** is the most critical, accurate, and meaningful measure, because it accounts for fault tolerance; reliability alone ignores backup components.
- × “Improving maintainability lowers availability.” — Opposite. With reliability constant, **higher maintainability** (faster, easier repair) raises availability; lower maintainability lowers it. Availability is a function of reliability *and* maintainability.
- × “All downtime is unplanned failure.” — Downtime is **scheduled** (preplanned maintenance, upgrades, internal testing) or **unscheduled** (unpredictable failures or delays). Scheduled downtime is expected and can even appear in an availability formula.
- × “DDoS attacks only hit web servers.” — They also target **DNS servers** and **SSL/encrypted traffic**; reflection attacks on DNS can cause significant downtime and even power outages across the data center.
- × “A single strong password is enough.” — Single-factor, password-based authentication invites password guessing, stolen identities, and automated brute force. Link passwords with **out-of-band verification** (for example, SMS) and handle authentication **centrally**; ADCs add further protection.

08 SELF-QUIZ

answers are upside-down below

Q1 What single word describes the readiness of a data center to be accessible for use?

Q2 In an SLA, what key term is used to express the required level of uptime?

Q3 “Five nines” equals what uptime percentage, and about how much downtime per year?

Q4 According to the Gartner survey cited in the module, roughly how much does an organization lose per hour of downtime?

Q5 Downtime is divided into which two categories?

Q6 Write the formula for a server's availability using MTBF and MTTR.

Q7 Availability measures the percentage of time equipment is operational — what does reliability measure?

Q8 Availability is described as a function of which two other properties?

Q9 In the levels of operational availability, which level allows zero yearly hours for prescheduled maintenance?

Q10 Name three common threats to data center availability.

ANSWER KEY (rotate the page)

Q1 availability **Q2** the 9s of availability (the nines) **Q3** 99.999%; about 5.256 minutes per year **Q4** about $300,000 per hour **Q5** scheduled and unscheduled downtime **Q6** Availability = MTBF ÷ (MTBF + MTTR) **Q7** whether/how well equipment performs its intended function over time (its durability and quality; probability of success) **Q8** reliability and maintainability **Q9** Level 4 (0 yearly hours) **Q10** any three of: power failure, human error, data center-level failure (natural disasters), DDoS attacks (DNS / web apps / SSL), brute force / fragile authentication