ONE PAGE SUMMARY

IT & Software › System Design › Availability

Introduction to Availability

Mastering Availability: The Key to Reliable Systems

DEFINITION

> Availability refers to the degree to which a system is operational and accessible when needed. High availability ensures minimal downtime and consistent service for users.

KEY CONCEPTS

Uptime

Uptime is the percentage of time a system is operational and available for use.

Downtime

Downtime is the period when a system is not operational, affecting user access.

Redundancy

Redundancy involves having backup components to maintain availability during failures.

Failover

Failover is the process of switching to a backup system automatically when the primary fails.

Load Balancing

Load balancing distributes workloads across multiple systems to enhance availability and performance.

EXAMPLES

→ A web service that guarantees 99.9% uptime.

→ Using multiple servers to ensure service remains available during maintenance.

→ Implementing automatic failover systems for critical applications.

MEMORY TIPS

★ Remember 'Uptime' as 'Up and running'.

★ Think of 'Redundancy' as 'redundant backups'.

★ Use 'Failover' as 'fail and switch'.

COMMON MISTAKES

✗ Confusing uptime with performance metrics.

✗ Neglecting to plan for redundancy.

✗ Underestimating the impact of downtime.

QUICK RECAP

Availability is crucial for system reliability, focusing on uptime and minimizing downtime. Key strategies include redundancy, failover, and load balancing to ensure continuous access for users.