

IT Infrastructure Architecture

Infrastructure Building Blocks
and Concepts

Availability Concepts

Sources of unavailability - human errors

- 80% of outages impacting mission-critical services is caused by people and process issues
- Examples:
 - Performing a test in the production environment
 - Switching off the wrong component for repair
 - Swapping a good working disk in a RAID set instead of the defective one
 - Restoring the wrong backup tape to production
 - Accidentally removing files
 - Mail folders, configuration files
 - Accidentally removing database entries
 - Drop table x instead of drop table y

Sources of unavailability - software bugs

- Because of the complexity of most software it is nearly impossible (and very costly) to create bug-free software
- Application software bugs can stop an entire system
- Operating systems are software too
 - Operating systems containing bugs can lead to corrupted file systems, network failures, or other sources of unavailability

Sources of unavailability - planned maintenance

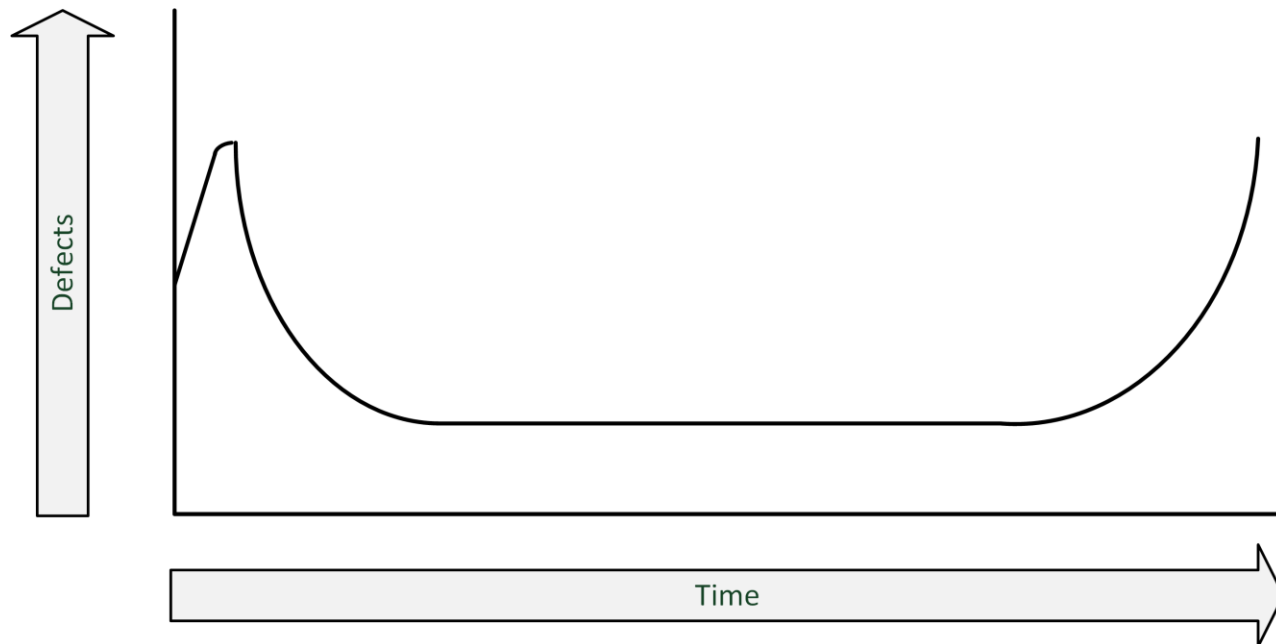
- Sometimes needed to perform systems management tasks:
 - Upgrading hardware or software
 - Implementing software changes
 - Migrating data
 - Creation of backups
- Should only be performed on parts of the infrastructure where other parts keep serving clients
- During planned maintenance the system is more vulnerable to downtime than under normal circumstances
 - A temporary SPOF could be introduced
 - Systems managers could make mistakes

Sources of unavailability - physical defects

- Everything breaks down eventually
- Mechanical parts are most likely to break first
- Examples:
 - **Fans for cooling equipment** usually break because of dust in the bearings
 - **Disk drives** contain moving parts
 - **Tapes** are very vulnerable to defects as the tape is spun on and off the reels all the time
 - **Tape drives** contain very sensitive pieces of mechanics that can break easily

Sources of unavailability - bathtub curve

- A component failure is most likely when the component is new
- When a component still works after the first month, it is likely that it will continue working without failure until the end of its life



Sources of unavailability - environmental issues

- Environmental issues can cause downtime:
 - Failing facilities
 - Power
 - Cooling
 - Disasters
 - Fire
 - Earthquakes
 - Flooding

Sources of unavailability - complexity of the infrastructure

- Adding more components to an overall system design can undermine high availability
 - Even if the extra components are implemented to achieve high availability
- Complex systems
 - Have more potential points of failure
 - Are more difficult to implement correctly
 - Are harder to manage
- Sometimes it is better to just have an extra spare system in the closet than to use complex redundant systems

Redundancy

- Redundancy is the duplication of critical components in a single system, to avoid a single point of failure (SPOF)
- Examples:
 - A single component having two power supplies; if one fails, the other takes over
 - Dual networking interfaces
 - Redundant cabling

Failover

- Failover is the (semi)automatic switch-over to a standby system or component
- Examples:
 - Windows Server failover clustering
 - VMware High Availability
 - Oracle Real Application Cluster (RAC) database

Fallback

- Fallback is the manual switchover to an identical standby computer system in a different location
- Typically used for disaster recovery
- Three basic forms of fallback solutions:
 - Hot site
 - Cold site
 - Warm site

Fallback – hot site

- A hot site is
 - A fully configured fallback datacentre
 - Fully equipped with power and cooling
 - Applications are installed on the servers
 - Data is kept up-to-date to fully mirror the production system
- Requires constant maintenance of the hardware, software, data, and applications to be sure the site accurately mirrors the state of the production site at all times

Fallback - cold site

- Is ready for equipment to be brought in during an emergency, but no computer hardware is available at the site
- Applications will need to be installed and current data fully restored from backups
- If an organization has very little budget for a fallback site, a cold site may be better than nothing

Fallback - warm site

- A computer facility readily available with power, cooling, and computers, but the applications may not be installed or configured
- A mix between a hot site and cold site
- Applications and data must be restored from backup media and tested
 - This typically takes a day

Business Continuity

- An IT disaster is defined as an irreparable problem in a datacenter, making the datacenter unusable
- Natural disasters:
 - Floods
 - Hurricanes
 - Tornadoes
 - Earthquakes
- Manmade disasters:
 - Hazardous material spills
 - Infrastructure failure
 - Bio-terrorism

Business Continuity

- In case of a disaster, the infrastructure could become unavailable, in some cases for a longer period of time
- Business Continuity Management includes:
 - IT
 - Managing business processes
 - Availability of people and work places in disaster situations
- Disaster recovery planning (DRP) contains a set of measures to take in case of a disaster, when (parts of) the IT infrastructure must be accommodated in an alternative location

RTO and RPO

- RTO and RPO are objectives in case of a disaster
- Recovery Time Objective (RTO)
 - The maximum duration of time within which a business process must be restored after a disaster, in order to avoid unacceptable consequences (like bankruptcy)

RTO and RPO

- Recovery Point Objective (RPO)
 - The point in time to which data must be recovered considering some "acceptable loss" in a disaster situation
- RTO and RPO are individual objectives
 - They are not related

