Failover

Failover je mechanismus automatického přepnutí na záložní systém (server, síťový prvek nebo komponentu) v případě, že primární systém selže nebo je odstaven z důvodu údržby. Cílem failoveru je minimalizovat nebo zcela eliminovat výpadek služby pro koncového uživatele.

Failover je klíčovou součástí strategií **vysoké dostupnosti (HA)** a **Disaster Recovery (DR)**.

Typy Failover konfigurací

V praxi se nejčastěji setkáváme se dvěma základními modely uspořádání:

1. Active-Passive (Hot Standby)

V tomto režimu běží pouze jeden systém (A), zatímco druhý (B) je v pohotovosti.

- **Princip:** Pokud systém A přestane odpovídat, systém B okamžitě převezme jeho identitu (např. IP adresu a úložný prostor).
- **Výhoda:** Jednoduchost, žádné riziko konfliktů dat.
- **Nevýhoda:** Záložní hardware většinu času „zahálí“, přestože za něj bylo zaplaceno.

2. Active-Active (Cluster)

Oba systémy běží současně a sdílejí zátěž.

- **Princip:** Při výpadku jednoho systému druhý prostě převezme veškerý provoz.
- **Výhoda:** Efektivní využití hardwaru, vyšší celkový výkon.
- **Nevýhoda:** Vyšší komplexita konfigurace a potřeba zajistit konzistenci dat mezi uzly.

Klíčové komponenty Failoveru

Aby failover fungoval spolehlivě, vyžaduje několik technických prvků:

- **Heartbeat (Tlukot srdce):** Neustálá komunikace mezi uzly. Pokud uzel přestane vysílat signál, ostatní to vyhodnotí jako jeho selhání.
- **Quorum:** Mechanismus (často hlasování), který určuje, který uzel má právo převzít služby, aby nedošlo k situaci **Split-Brain** (kdy si oba servery myslí, že jsou primární, a začnou přepisovat stejná data).
- **Fencing (Stonith):** Metoda „zastřelení“ nebo izolace vadného uzlu, aby se zajistilo, že se nebude pokoušet o zápis na sdílené úložiště, zatímco už běží failover.
- **Virtual IP (VIP):** Sdílená IP adresa, která se při výpadku „přesune“ z jednoho serveru na druhý. Uživatel tak stále přistupuje na stejnou adresu.

Úrovně Failoveru

- **Hardwarový Failover:** Redundantní zdroje, ventilátory nebo HBA karty v serveru.
- **Aplikační Failover:** Např. databázový cluster (Microsoft SQL, PostgreSQL), kde jedna instance převezme práci druhé.
- **Cloudový Failover:** Automatický restart virtuálního stroje na jiném fyzickém hostiteli (funkce HA v Proxmoxu nebo VMware).
- **Geografický Failover:** Přepnutí provozu do jiného datacentra (jiného města/státu) při rozsáhlém výpadku infrastruktury.

Rozdíl mezi Failover a Load Balancing

Ačkoliv se tyto termíny často prolínají, mají jiný primární cíl:

- **Load Balancing:** Rozkládá zátěž pro dosažení vyššího výkonu. Často obsahuje failover jako vedlejší funkci.
- **Failover:** Zaměřuje se výhradně na **stabilitu a dostupnost**.

Metriky úspěchu

- **RTO (Recovery Time Objective):** Jak dlouho trvá, než se služba po výpadku opět rozběhne.
- **RPO (Recovery Point Objective):** Kolik dat může být ztraceno (např. data od poslední synchronizace mezi uzly).

— **Související termíny:** High Availability, Cluster, Multipathing, Load Balancing, Disaster Recovery.