Vysoká dostupnost dat (nejen) v cloudu

Ve snaze o zajištění vysoké dostupnosti dat ve vlastním datovém centru se typicky zaměřujeme na dvě základní úrovně.

JAN CIPRA

První z nich je ochrana proti fyzické chybě, druhou je pak ochrana proti chybě logické.

Fyzickou chybou může být například výpadek jedné komponenty či kompletní ztráta celého datového centra. Ochranou proti nedostupnosti způsobené těmito chybami jsou pak plně redundantní HW zařízení, serverová virtualizace zajišťující dostupnost systémů v případě výpadku části host serverů a také řešení zajišťující distribuci dat na více lokalit a jejich dostupnost i v případě výpadku jednoho kompletního datového centra. Tyto technologie typicky fungují transparentně a bez dopadu na uživatele.
Naproti tomu logická chyba, kterou může být například chyba v softwaru, lidská chyba či úmysl vedoucí ke smazání či poškození dat a také v poslední době hodně sklňovaný ransomware, je mnohem zákeřnější. Zejména z těchto důvodů:

- Na poškození nebo ztrátu dat se nepřijde okamžitě a typicky problém objeví uživatelé, protože má negativní dopad na jejich práci.
- Oprava vyžaduje lidskou intervenci a může jít o netriviální operaci.
- Technologie zabezpečení proti fyzické chybě nám ve většině případů nepomohou.
- Minimalizaci dopadu logické chyby na dostupnost dat a potažmo na uživatele zajišťujeme technologiemi, které po danou dobu a granularitu udržují historická data. Ta jsou v případě potřeby použita pro obnovu těch poškozených. Čím vyšší časová granularita ukládaných dat, tím kratší historii typicky dané řešení uchovává. Konkrétně jde o následující technologie:
- Snapshoty na primárních či sekundárních diskových systémech
- Kontinuální zálohování či replikace, ideálně s jistou formou žurnálu pro umožnění point-in-time obnovy (obnova dat do téměř libovolného bodu v čase v rámci uchovávaného časového intervalu)
- Tradiční zálohování formou ukládání kopií primárních dat (cílem jsou disky, pásky či cloudová úložiště)

Správnou kombinaci technologií je pak ideální volit na základě požadavku na dostupnost v rámci business continuity a disaster recovery plánování. Ideální je využít zkušeného partnera, který se problematikou dlouhodobě zabývá.

Jak je to ale se zajištěním dostupnosti dat v případě cloudové infrastruktury, konkrétně v případě public cloudu? V takovém prostředí typicky nemáte možnost ovlivnit použitou HW infrastrukturu, takže ochranu proti fyzické chybě vám poskytuje cloudový poskytovatel a je na vás se seznámit s SLA jednotlivých poskytovaných služeb a možnostmi, jak případně dostupnost zvýšit. Ke službám většinou existuje několik variant s definovanou úrovní SLA, a je tak na zvážení uživatele, jaká varianta vyhovuj jeho požadavkům.

Ale jak je to s chybou logickou? Jsou data v cloudu chráněná i proti těmto rizikům? Odpovědí na tyto otázky je tzv. shared responsibility model, což je rozdělení odpovědnosti za zabezpečení dat mezi zákazníka a poskytovatele služeb.

Níže uvedená tabulka shrnuje to, jak je shared responsibility model definován cloudovými poskytovateli – viz například *https://aws.amazon.com/compliance/shared-responsibility-model/* a *https://docs.microsoft.com/cs-cz/azure/security/fundamentals/shared-responsibility*.

Z tabulky je evidentní, že zabezpečení dat je věc, která je čistě na zodpovědnosti zákazníka. Zabezpečením dat zde chápejte nejen zajištění bezpečnosti dat oproti jejich kompromitaci, ale také zabezpečení dat proti logické chybě popsané v úvodu tohoto článku. Je tedy velmi důležité si uvědomit, že ani v cloudu se o zabezpečení vašich dat nepostará nikdo za vás! Je potřeba se o data starat stejně jako u vás v datovém centru.

Často se bohužel setkáváme s přístupem bezmezné a ničím nepodložené důvěry v cloudového poskytovatele, že vše je již vyřešeno. Tento přístup vede k nemilým překvapením v situaci, kdy dojde k nějakému výpadku nebo poškození dat, a zákazník je seznámen se strohou realitou poskytované služby. Proto je potřeba se s realitou poskytované služby obeznámit již na začátku a v rámci kompletního řešení neopomenout i na zajištění požadované dostupnosti dat.

Způsobů a možností, jak toho dosáhnout, je několik a vesměs odpovídají tomu, jak dostupnost můžeme zabezpečit i v on-premise infrastruktuře.

- Využití nativních služeb poskytovaných pro danou službu.
  - Snapshoty virtuálních serverů a jejich ukládání na cloudovou storage
  - Nativní prostředky zálohování a retence dat pro služby, jako jsou například databáze
  - Verzování a zamykání dat proti přepisu a smazání u objektových úložišť
  - Obnovitelné položky a archivace pro Office 365
- Nástroje třetích stran
  - Zálohování (tradiční i kontinuální) dat uložených v cloudových instancích a SaaS službách (cílem může být objektové úložiště nebo úložiště u jiného cloud providera či on-premise)

Stejně jako v on-premise infrastruktuře je potřeba se zamyslet nad vhodnou kombinací výše uvedených nástrojů k dosažení požadované dostupnosti. Ideální je opět konzultace s partnerem s dlouholetými prokazatelnými zkušenostmi v této problematice.

Autor je konzultant ve společnosti GAPP System.

Shared Responsibility Model

On-Premises	IaaS	PaaS	SaaS
Data	Data	Data	Data
Aplikace	Aplikace	Aplikace	Aplikace
Operační systém	Operační systém	Operační systém	Operační systém
Hypervizor	Hypervizor	Hypervizor	Hypervizor
HW infrastruktura	HW infrastruktura	HW infrastruktura	HW infrastruktura

☐ Zodpovědnost zákazníka ■ Zodpovědnost poskytovatele