



ZÁPADOČESKÁ
UNIVERZITA
V PLZNI

KIV/TSI - Samostatná práce

Cloud computing

Stručný úvod

David Brůha (A09N0062K)
d.bruha@1art.cz

Obsah

Úvod.....	3
Definice pojmu.....	3
Neomezené zdroje.....	3
Prvotní investice.....	5
Pružné využívání zdrojů.....	5
Cloud computing v praxi.....	5
Amazon Web Services (AWS).....	6
Microsoft Azure.....	6
Google App Engine.....	7
Závěr.....	7

Úvod

Cloud computing (někdy též zmiňovaný jako SaaS – Software as a Service) je pojem, který se v oblasti moderního webhostingu objevuje čím dál tím častěji. Mnohé společnosti se jím zaklínají (přízně si, že je to pojem, který na zákazníky dělá dojem) a to i v případě, že jejich řešení nemá s cloud computingem téměř nic společného. Jsou zase naopak projekty, které možnosti cloud computingu využívají již mnoho let, ale nijak toto nezdůrazňují.

Co to tedy cloud computing doopravdy je, jak bychom ho mohli definovat a jaké nám přináší možnosti? Pokusím se ve své práci rozsáhlou problematiku cloud computingu alespoň obecně shrnout a naznačit, které oblasti mě zaujaly a jaké jsou mé názory na některé dnes dostupné služby. Popíši také základní možnosti Google App Engine – aplikačního frameworku, který je za určitých podmínek možno využívat i zdarma.

Definice pojmu

Při pročítání článků na internetu jsem narazil na mnoho různých definic co to cloud computing je, kde začíná a jaký je jeho potenciál. Bohužel se domnívám, že mnoho autorů ne zcela pochopilo danou problematiku a vytvářejí tak prostředí zmatenosti. I článek na wikipedii (http://en.wikipedia.org/wiki/Cloud_computing) je podle mého názoru minimálně nedostatečný – odhlédneme-li tedy od toho, že se na wikipedii podobných článků dá najít více.

Dovolil bych si souhlasit s autory článku „A view of cloud computing“¹ a cloud computing definovat na základě tří jednoduchých vlastností:

1. Z úhlu pohledu uživatele cloudu jsou k dispozici v podstatě neomezené zdroje.
2. Abych tyto zdroje využíval, není nutné nejprve provést rozsáhlou investici.
3. Zdroje mohou využívat tak, jak je potřebuji. Neplatím za nevyužitý čas nebo nevyužitou kapacitu výpočetního výkonu.

Podívejme se na tyto vlastnosti blíže.

Neomezené zdroje

Pro začínající internetový projekt je obvyklé, že jeho nároky jsou velmi malé. Jeho návštěvnost je nízká a nezatěžuje výrazným způsobem servery, ze kterých je poskytován (ačkoliv ve své praxi jsem již několikrát narazil na výjimky a musel jsem ocenit například „extrémně dobře“ napsané elektronické obchody schopné přetížit server se čtyřjádrovým procesorem).

Nicméně cílem většiny internetových projektů je růst, lákat stále více návštěvníků a pokud možno tak i zvyšovat výdělky společností, které tyto projekty provozují. Pro mě, jako autora a provozovatele podobného projektu je důležité, aby jeho prvotní náklady byly co možná nejnížší a abych měl jistotu, že ve chvíli, kdy projekt začne být úspěšný a jeho obliba poroste, nenarazím na kapacitní nebo výkonnostní hranice serveru nebo linky a neodradím tak návštěvníky nízkou rychlostí nebo přímo nedostupností. Je bohužel internetovou realitou, že získat nové pravidelné návštěvníky (zákazníky) není nijak snadné. Ztratit je, to je naopak velmi jednoduché.

Představme si nyní hypotetickou situaci – jsem programátor, který přišel s revolučním webovým projektem, který chci realizovat. Sice doufám, že bude velmi úspěšný, ale dopředu to nedokážu zaručit. Umístím ho proto na běžný webhosting a začnu ho provozovat a propagovat. Postupně přichází více a více návštěvníků (ideálně samozřejmě zákazníků) a

1 Článek je on-line dostupný na adrese <http://portal.acm.org/citation.cfm?id=1721672>

projekt stále více a více zatěžuje přidělené zdroje ať už to je počet dotazů do databáze nebo výkon procesoru webového server. Může také samozřejmě jít o objem dat, která se po internetu přenáší.

V takové chvíli jsem zkontaktován poskytovateli svého běžného webhostingu a upozorněn, že je můj projekt příliš náročný na provoz ve zvolené hostingové kategorii a pokud se jedná o slušného poskytovatele, je mi nabídnuto navýšení této kategorie. Pokud jsem poskytovatele zvolil špatně, je můj projekt bez milosti odstaven a má smlouva vypovězena na základě nějaké drobné klauzule v provozních podmínkách, tuto variantu zde ale nebudu uvažovat (ačkoliv není zejména v českých podmínkách bohužel nijak vzácná).

Rozhodnu se tedy zvýšit hostingovou kategorii a projekt roste dál. Poskytovatel ho přesunul na virtuální server, posléze na dedikovaný server a pokud mám štěstí a poskytovatel je schopný, vše proběhlo bez problémů a bez výpadku. Projekt se mezitím snažím optimalizovat, využívám kešování, upravil jsem SQL dotazy a přesto zátěž dále roste. Výkon jednoho serveru již nestačí a poskytovatel pro projekt připraví systém s rozkladem zátěže (load balancing + replikace databáze).

Protože je projekt určen pro návštěvníky z celého světa, roste i objem dat přenesených do zahraničí a jejich cena se stává významnou částí pravidelných plateb. To však nevadí, projekt vydělává a já přidávám nové vlastnosti, které opět lákají nové návštěvníky. Jejich objem se zvyšuje skokově a je třeba přidávat další a další servery, které budou běh projektu zajišťovat, datové toky vzrůstají tak, že je již není schopen softwarový systém rozumným způsobem rozkládat a návštěvníci si možná začínají stěžovat na občasnou pomalost nebo dokonce nedostupnost projektu. Někteří odejdou a pravděpodobně se již nevrátí (pokud se nejedná o opravdu jedinečný projekt, který nemá svou obdobu, to je ale jen velmi málo pravděpodobné).

Pokud byl nárůst postupný, škálování zdrojů se dalo dobře plánovat a poskytovatel služeb je schopný, pořídil pro mě již v tuto chvíli hardwarové zařízení pro rozklad zátěže v ceně několika milionů korun, nakoupil do zásoby servery, které je možné ve špičkách zapnout a zajistit tak dostatek zdrojů. Posílil také konektivitu tak, aby projekt nenarazil na omezenou šířku pásma.

Pokus je ale nárůst skokový, může dojít k situaci, kdy je potřeba výkon doplnit, ale poskytovatel nemá potřebné stroje k dispozici a není je možné obratem dokoupit. Čekat v podobné situaci týden na dodání strojů od výrobce (a to v ideálním případě, řada specializovaných zařízení má dodací lhůty i několik desítek týdnů) může projekt v podstatě zabít. Objektivně ale není možné nic jiného dělat, dodací lhůty jsou neúprosné, limitují můj růst a návštěvníci hledají alternativní projekty. Je zřejmé, že toto není ideální situace, ale jiné řešení není. Nebo je?

Představme si, že jsem od začátku umístil projekt k poskytovateli, který umožňuje zdroje dynamicky škálovat a již má k dispozici velmi rozsáhlou infrastrukturu. Jako tvůrce projektu nemusím vědět jaké stroje běh projektu zajišťují ani kde jsou umístěny. Získávám pouze prostor, výpočetní výkon a samozřejmě síťovou konektivitu a vše roste s mým projektem podle jeho potřeb. Poskytovatelova technologie dokonce můj projekt rozloží geograficky tak, aby byl co nejbližší svým návštěvníkům, sníží tak latence a zvýší rychlost odezvy po celém světě.

Taková technologie bude zároveň škálovat a rozkládat automaticky bez zásahu administrátorů a nebude tak závislá na rychlosti reakce lidského faktoru. Odstraní se tak úzká hrdla běžných webhostingových služeb a návštěvníci nebudou narážet na problémy, které výše popsané řešení přináší.

Mnou využívaná služba je tedy něco jako black box – nevím kde je projekt fyzicky umístěn, co zajišťuje jeho provoz ani jak je zajištěno jeho škálování. Vím pouze, že jeho funkčnost je bezproblémová a že si nemusím dělat starosti s dalším nárůstem návštěvníků. Naopak o něj samozřejmě usiluji a problémy s provozem neřeším. Nemusím se bát ani nasazení nových skvělých vlastností, které přinesou skokový nárůst zájmu o projekt a to i

v řádu hodin. Abych parafrázoval jeden poněkud překroucený výrok z české politiky – zdroje prostě jsou.

Toto je první kritérium, které definuje službu cloud computingu.

Prvotní investice

Vraťme se opět na začátek našeho hypotetického příkladu a uvažujme situaci, kdy předpokládám, že mnou vytvářený projekt bude velmi úspěšný. Odhlédněme nyní od toho, že v obrovskou úspěšnost projektu věří téměř každý, kdo se snaží realizovat svou vizi, předpokládejme, že jsem si vytvořil business plán, který s rapidním nárůstem návštěvnosti počítá a který je postaven na doufejme reálném základě.

Abych projektu zajistil dostatek zdrojů na provoz, již od začátku nakoupím dostatek serverů, zařízení pro rozklad zátěže a vše nechám zprovoznit. V první fázi si projekt hledá zákazníky, většina kapacity strojů leží ladem a perspektiva jejich využití je i přes dobrý business plán nejistá a ve velké většině případů musím počítat s dlouhou dobou finanční návratnosti velmi rozsáhlých investic.

Uvažme zde opět možnost, že svůj projekt budu od začátku provozovat na zmíněném black boxu. Ten má přesně určeno, jaké platby budu provádět za jak rozsáhlé využití výpočetních prostředků a já tak na začátku platím pouze velmi nízké částky za provoz. Samozřejmě, jakmile bude projekt a jeho úspěšnost růst, budu platit více a více, ale projekt mi v tu dobu také bude více a více vydělávat.

Tento black box tedy splňuje i druhý z požadavků na cloud computing – nenutí mě k vysoké prvotní investici do infrastruktury a umožňuje mi tak zaměřit se cele na vývoj projektu a na jeho vylepšování a budování uživatelské základny.

Pružné využívání zdrojů

Buďme nicméně realisté. Je jen velmi málo projektů, které pouze rostou a mají stále větší a větší nároky na zdroje. I ty nejúspěšnější projekty budou mít období, kdy návštěvnost klesá. Představme si například situaci, kdy v projektu nasadím úžasnou novou funkčnost. Ta logicky návštěvníky zaujme a přichází jich více a více. Po nějaké době, ale jejich zájem poleví, tato funkčnost se „okouká“ a nebudou ji tak často vyhledávat. To znamená, že objem zdrojů, které projekt potřebuje, se v tuto dobu sníží, proč bych tedy za ně měl platit? Není to zbytečné?

Pokud mnou zmiňovaný black box podporuje dynamické škálování zdrojů a plateb za ně, nemusím tyto úvahy řešit. V době nižšího zatížení spotřebovávám méně zdrojů a bude mi tak vyúčtována nižší částka. Sice méně vydělám, ale také méně utratím.

Tato filozofie je třetím ze základních kamenů cloud computingu.

Cloud computing v praxi

Definovali jsme si, jaké vlastnosti by měla mít služba, která splňuje parametry cloudu. Taková služba nevyžaduje vysokou prvotní investici, finančně se přizpůsobuje aktuálně spotřebovávaným zdrojům a tyto zdroje pokud mono žádným způsobem neomezuje. To je samozřejmě ideální stav, ale chceme se k němu přiblížit co možná nejlépe. Nebudu zde uvažovat speciální projekty, které by dokázaly vyčerpat prostředky kteréhokoliv aktuálně poskytovaného komerčního cloudu (příkladem by mohl být například facebook, který by rychle přerostl možnosti zdrojů podobných služeb), a budu uvažovat běžnější projekty, které také potřebují dynamicky škálovat svůj výkon a jsou na rozumné úrovni náročnosti.

Pokud tedy uvažujeme projekt výše zmíněného typu, je v současné době možné zvolit z celé řady alternativ. Asi nejznámější cloud computing služby jsou Amazon Web Services², Microsoft Azure³ a Google App Engine⁴. Je samozřejmě možné i uvažovat o vytvoření

2 <http://aws.amazon.com/>

3 <http://www.microsoft.com/cze/azure/>

4 <http://code.google.com/intl/cs/App Engine/>

vlastního cloudu postaveného například na open source projektu Eucalyptus⁵, tuto variantu ale nebudu dále rozvádět, ačkoliv je velmi zajímavá je podle mého názoru mimo oblast zaměření této práce.

Všechny tři zmíněné služby mají svá specifika, která je třeba vzít v úvahu již při prvotní analýze a návrhu projektu. Podrobný popis kterékoliv z nich by rozsahem pravděpodobně připomínal spíše diplomovou práci, popíšu zde tedy alespoň hlavní rysy těchto tří služeb hodnocené z mého subjektivního úhlu pohledu.

Amazon Web Services (AWS)

Tato služba je podle mého názoru jedna z nejvariabilnějších. Zákazníkovi poskytuje tzv. instance, které mohou jsou v podstatě virtuální operační systémy určené pro provoz specifických aplikací. Je na výběr více typů operačních systémů, takže nejsme omezeni např. na jistý specifický operační systém z dílny firmy z Redmontu.

AWS umožňuje provozovat těchto instancí více a mezi nimi provádět rozklad zátěže. Jednotlivé instance je možné přehledně monitorovat a plánovat tak další škálování. Nové instance je možné rezervovat, takže budou připraveny ve chvíli, kdy je budeme potřebovat.

Jako databázový backend je možné využít buďto nerelační Amazon SimpleDB nebo relační MySQL. Ve druhém čtvrtletí roku 2011 je plánováno zprovoznění databáze Oracle 11g.

Pokud potřebujeme ukládat data, využijeme Amazon Simple Storage Service (S3), která poskytuje „neomezený“ diskový prostor pro libovolná data.

Pro programátora je důležité, že mu AWS umožňuje využívat celou řadu programovacích jazyků – zmínil bych především PHP, Javu, .NET nebo i Python nebo Ruby. Není tak nutné se učit nový jazyk a pouze si nastudovat příslušný SDK.

AWS zároveň poskytuje další možnosti – je například možné vytvořit virtuální privátní síť a tu propojit s interní firemní infrastrukturou. Je tak možné vytvářet nejen běžné internetové aplikace, ale i z veřejného internetu nedostupné intranetové aplikace.

V současné době AWS poskytuje čtyři umístění, ve kterých uchovává data. Jedná se o tzv. US Standard, Severní Kalifornii, Irsko a Singapur. Mezi těmito umístěními je možné data přesouvat nebo replikovat pomocí specializovaných nástrojů AWS, tyto nástroje (Import/Export) by měly být údajně výrazně efektivnější, než kdyby se data běžným způsobem kopírovala z místa na místo.

Všechny tyto služby jsou rozděleny na balíčky, které jsou potom zpoplatněny. Balíček je možné chápat např. jako objem dat přenesených za minutu, velikost databáze, počet redundantních uložení dat atp.

Microsoft Azure

Microsoft Azure, respektive Windows Azure je cloud služba poskytovaná datacentry společnosti Microsoft. Využívá známého prostředí Windows a SQL serveru (zde pro cloud upraveného a pojmenovaného SQL Azure) a umožňuje dle definice cloudu škálovat výkon a prostředky pro aplikace tak, jak zrovna potřebují.

Komunikaci mezi nahranými aplikacemi a případně i mezi aplikací a interním systémem zákazníka zajišťuje technologie Azure Platform AppFabric.

Platforma Azure je primárně určena pro provoz .NET aplikací. Je možné použít i programovací jazyky PHP nebo i Java, ale pro ty není platforma optimalizována. Rozsah jejich podpory je ale výrazně větší než na začátku.

Pro řadu vývojářů může představovat nezanedbatelnou výhodu relační databáze SQL Azure, která je stále ještě funkčně bohatší než MySQL.

5 <http://open.eucalyptus.com/>

Windows Azure umožňuje zakoupení služby buďto v modelu Pay As You Go (tj. zákazník zaplatí pouze za reálně využitou službu) nebo v modelu předplatného na šest měsíců, kdy je možné výrazně ušetřit u aplikací, jejichž nároky nenarůstají skokově.

Google App Engine

Služba společnosti Google využívá rozsáhlou infrastrukturu tohoto vyhledávacího giganta po celém světě. Umožňuje umístit a sdílet aplikace naprogramované buďto v Javě nebo v Pythonu a využívající Google API.

Nespornou výhodou je možnost propojení s ostatními aplikacemi, které jsou na tomto cloudu provozovány – a to včetně služeb typu Google Calendar, GMail, Google Docs a dalších poskytnutých jak společností Google, tak i třetími stranami.

Aplikace běží uzavřeny v sandboxu, který definuje jejich bezpečnostní kontext, izoluje je (kromě funkcí, které se rozhodneme publikovat navenek) a umožňuje jim abstrahovat jak od operačního systému a hardware, na kterém jsou provozovány, tak i od lokace, ve které jsou umístěny.

I zde je samozřejmě poskytován prostor pro ukládání dat, v tomto případě se jedná o transakční databázi postavenou na enginu BigTable.

Jako obrovské pozitivum vnímám to, že kdokoliv může do cloudu nahrát svou vlastní aplikaci zcela zdarma, pokud je menší než 500MB a nemá více než pět milionů zobrazení stránky za měsíc. Google zároveň poskytuje Google App Engine zdarma pro nevýdělečné školní a výukové projekty.

Závěr

Cloud computing je bezesporu jednou z extrémně zajímavých alternativ k běžnému webhostingu. Není ani zdaleka samospasitelný, náklady na úpravu některých stávajících aplikací pro provoz na cloudu nebo i nasazování nových pod mnohdy velmi svazujícím API není vždy možné. Pro mnoho společností může být klíčovou otázkou přesun dat na blíže nespecifikované servery v cloudu a v některých diskusích na internetu i zmiňovaná nejistota, jestli je data možné z cloudu po nahrání zase odstranit.

Pokud ale akceptujeme modely funkčnosti cloud služeb tak, jak jsou jednotlivými poskytovateli dodávány, přináší nám to obrovský potenciál pro tvorbu, nasazení a dynamické škálování provozu aplikací. Služby tohoto typu se určitě budou dál a dál rozvíjet a budou poskytovat stále nové a nové funkčnosti. Navíc je pravděpodobné, že minimálně některé z nich budou provozovat svou funkčnost s aplikacemi provozovanými na nich zákazníci a umožní tak stále rozmanitější služby na své infrastruktuře.

Ze mnou zmiňovaných tří služeb považuji za nejvariabilnější Amazon Web Services. Microsoft Azure je podstatně svázanější a Google App Engine je podle mého názoru zdaleka nejstriktnější. Na druhou stranu je zde možné využívat velké množství již naprogramovaných nástrojů a propojit je se svou aplikací.

V budoucnosti bych rád vyzkoušel i tvorbu a testovací provoz vlastních cloud služeb postavených například na zmiňovaném open source projektu Eucalyptus. V podobných službách bude pravděpodobně definována budoucnost internetu, jisté náznaky se dají najít již dnes v tlaku na virtualizaci a provoz virtuálních strojů.