Compaq NonStop Himalaya Server

Kolik devítek je NonStop?

Umíte si představit, že byste si ani nepamatovali, kdy jste naposledy bootovali operační systém? Nebo že existují administrátoři, kteří bootují jednou ročně jen proto, aby nezapomněli, jak se to dělá? Že ne? Opusťme tedy svět padajících Woken a nahlédněme do říše "mnoha devítek".

Těmi mnoha devítkami jsou myšlena procenta dostupnosti systému, přesněji řečeno počet devítek v dostupnosti v procentech vyjádřené desetinným číslem. A systémy, o které půjde, jsou servery Tandem (dnes Compaq) NonStop Himalaya Server – systémy projektované k dosahování absolutně nejvyšší dostupnosti ve světě IT již po 25 let. Toto seznámení bude možná také užitečné v tom, že zaměří vaši pozornost trochu jinam než jen do módní oblasti PDA, palmtopů, WAP telefonů a osobních počítačů v náramkových hodinkách – bez spolehlivých a výkonných serverů v pozadí by tato tak trochu bondovská zařízení ztratila značný díl své atraktivnosti.

Možná že už jste slyšeli o dvou-, tří- či pětidevítkové spolehlivosti (99,999% dostupnost, tj. asi 5 minut výpadku při nepřetržitém ročním provozu). Většinou se tím však myslí dostupnost samotného hardwaru, serveru. Problém je však trochu širší – k bližšímu vysvětlení musím trochu odbočit k základům teorie spolehlivosti. Zařadíme-li tři prvky tak, že funkce každého bude závislá na výstupu předchozího, tedy sériově, budou se jejich spolehlivosti násobit. To znamená, že bude-li mít každý z nich 90% spolehlivost, bude výsledná spolehlivost systému jen 72,9 %. Naopak budou-li prvky zapojeny paralelně, vedle sebe, celková spolehlivost se zvýší, v daném případě na 99,9 %. To jsou teoretické případy, ale vysvětlují například to, proč jsou systémy s mnoha autonomně pracujícími procesory nebo s redundantními ("nadbytečnými") prvky mnohem spolehlivější než jednoprocesorové, ale i to, že u systému, jehož hardware má pět devítek, ale na němž pracuje software s nějakou ne absolutní spolehlivostí a je napojena také ne úplně spolehlivá síť, můžeme zjistit celkovou spolehlivost, u níž se už ani o devítkách nedá hovořit. A vysoká celková spolehlivost systému je to, co uživatele jedině zajímá – cíl, pro nějž jsou projektovány systémy Himalaya, u nichž je vysoce spolehlivý hardware s prvky paralelizmu, redundance a samokontroly kombinován s podobně koncipovaným systémovým i aplikačním softwarem a síťovými prvky.

K čemu a jak devítky

Píšete-li doma na počítači v "texťáku", spolehlivost vás asi moc nebolí. Když selže systém, na němž se počítá vaše výplata, to už vás naštve. Pokud však půjde o počítač obhospodařující peníze mnoha lidí, zajišťující ve velkém měřítku telekomunikace či e-commerce apod., bude nejen míra naštvání neúnosná, ale odpovědnému subjektu to přinese i značné finanční ztráty a také ztrátu renomé, která jej v důsledcích přijde ještě mnohem dráž. Není proto náhodou, že systémy NonStop Himalaya používá všech 35 největších světových telekomunikačních společností a že zabezpečují 90 % operací na světových burzách. Jsou na nich zpracovávány téměř všechny hovory s předčíslím 800 a více než 65 % hovorů s předčíslím 911, vznikajících v USA, 80 % aplikací pro péči o zákazníka a účtování bezdrátových hovorů v Evropě a Asii, 80 % platebních transakcí, 66 % transakcí s kreditními kartami a mnoho dalších tzv. kritických aplikací.

Tyto údaje potvrzují výjimečné postavení systémů NonStop Himalaya v tomto segmentu trhu, který vyžaduje tzv. dostupnost 7 x 24, tedy sedm dní v týdnu po celých 24 hodin. Možná také proto se u systémů Himalaya neuvádějí počty devítek, protože jejich spolehlivost i pověst, potvrzené jmény renomovaných zákazníků (kromě zmíněných telekomunikačních gigantů je mezi nimi např. NYSE – newyorská burza, Nasdaq – světové středisko trhu s cennými papíry, MeritaNordbanken – první "internetová" banka v Evropě a další), nepotřebují marketingová čísla. Ukazují také význam, jaký vysoká dostupnost systémů v této oblasti má. Může se zdát, že jde o úzký a vysoce specializovaný trh. Jmenované oblasti však patří k dynamicky velmi rychle se rozvíjejícím oborům, takže je zřejmé, že podíl vysoce spolehlivých systémů bude v podmínkách boje o získání konkurenční výhody stále větší.

Základní jednotkou hardwaru systémů Himalaya je procesorový pár (v současnosti procesory MIPS R12000, za tři až čtyři roky mají být nahrazeny procesory Alpha), v němž se paralelně provádějí shodné operace a výsledky se stále vzájemně porovnávají. Tato jednotka má vlastní paměť, cache paměť a zdvojené propojení na router a pracuje na ní vlastní kopie distribuovaného operačního systému. Jednotky jsou vzájemně propojeny routery až do počtu 16 jednotek v jednom uzlu, mezi nimiž zprostředkovávají zasílání zpráv. Jednotlivé uzly mohou být dále propojeny LAN nebo WAN sítí a vytvořit tak jediný (single-image) systém s teoreticky až 4080 procesory.

Jednotky v uzlu jsou propojeny tzv. ServerNet propojením s vysokou kapacitou. Je zprostředkováno routery, které u nejnovějšího typu serveru podporují 12 nezávislých cest propojení, z nichž každá má kapacitou 125 MB/s, tedy úhrnně 1,5 GB/s kapacity pro duplexní přenos dat. Pro výrazné snížení latence přenosu dat (300 ns v jednom routeru) užívá NonStop Himalaya tzv. wormhole routed přenos paketů. Znamená to, že pro další přenos paketu dat není nutné čekat, až se načte celý paket, ale lze jej do místa

určení zasílat už po přečtení jeho hlavičky v době, kdy se zbytek paketu teprve přijímá. Prostřednictvím routerů jsou procesorové jednotky spojeny také s I/O (vstupně/výstupními) porty. Přidáváním routerů se průchodnost systému lineárně zvyšuje.

Ke stálé kontrole funkce zajišťované procesorovými páry přistupuje ECC paměť, která ošetřuje korekci jednobitových chyb. Vznikne-li vícebitová chyba, operační systém odstaví paměť včetně procesoru až do výměny vadné paměti a přenese jeho zatížení do jiného procesoru. Podobně jsou monitorována a nahrazována zřízením náhradní cesty pro data také všechna I/O zařízení a komunikační adaptéry.

Integrita přenosu dat je zajištěna CRC kontrolou paketů. Disky jsou zásadně zrcadlené, vadný disk může být za provozu vyměněn a rovněž automaticky za provozu proběhne reintegrace opraveného disku. Speciální diagnostický a údržbový subsystém TSM monitoruje a testuje činnosti systému, lokalizuje poruchy, testuje, provádí analýzy a restartuje komponenty systému, vše během chodu serveru. Automaticky je také monitorována činnost napájení, bateriový fault-tolerant záložní systém udrží server v chodu až do 30 sekund, čímž umožní korektní odstavení systému, a také udržuje obsah paměti po dobu až jedné hodiny.

Software a vývoj aplikací

Základem softwarového vybavení serverů Himalaya je distribuovaný operační systém založený na zasílání zpráv, podpora transakčního zpracování NonStop Tuxedo, paralelní aplikační server a vlastní paralelní relační distribuovaný databázový systém NonStop SQL/MP, optimalizovaný pro architekturu Himalaya i pro transakční a OLTP zpracování. Progresivními složkami programového vybavení jsou také NonStop Distributed Object Manager/MP, Server for Java, umožňující snadný vývoj aplikací standardními vývojovými prostředky, a ISG Navigator, podporující jednotný přístup ke všem zdrojům podnikových dat prostřednictvím internetu.

Vývoj aplikací může probíhat ve standardním C/S (klient/server) prostředí, tedy i na PC. Předpokládá se vytvoření samostatných softwarových serverů pro jednotlivé činnosti. Jejich práci koordinuje důležitá komponenta aplikačního serveru nazvaná Pathway, která řídí startování příslušných serverů podle požadavků klientů, tedy podle zatížení systému. Koncepce systému umožňuje startování zcela nových serverů bez přerušení běhu systému, takže je možné za běhu spouštět nové aplikace, tedy rozšiřovat a upgradovat systém bez přerušení chodu a ihned využívat možnosti nově zařazených aplikací.

Aplikace pro NonStop Himalaya Server si zaslouží samostatný článek, protože v rámci tohoto již není pro jejich odpovídající popis dostatek prostoru. Zatím mohu pouze předeslat, že hlavními aplikačními oblastmi jsou telekomunikace, bankovnictví a finance, internet a e-aplikace a také řízení výroby, zejména se zaměřením na tzv. just in time model v řízení dodávek. Úplně novou a fascinující oblastí je koncepce ZLE (Zero Latency Enterprise), řízení procesů na základě analýzy rozsáhlých dat s prakticky nulovým zpožděním, která byla na systému Himalaya prvně implementována v oblasti telekomunikací – ale o tom více až příště.

Josef Chládek