Jak se vyznat v událostech

Michal Džmuráň

Komplexní zpracování událostí v reálném čase CEP (Complex Event Processing) je relativně nová softwarová technologie, která aplikacím umožňuje monitorovat více toků byznys událostí najednou, analyzovat je z hlediska výkonnostních ukazatelů podle předem daných pravidel a reagovat na základě zjištěných příležitostí a hrozeb v reálném čase.

Tradiční podnikové ERP systémy příchozí data o jednotlivých byznys událostech zpravidla pouze registrují a evidují, ukládají snímek stavu organizace v daném časovém okamžiku. Tento snímek je vyhodnocován až s časovým zpožděním. Reakce a podnikatelská rozhodnutí následují až relativně dlouho poté, co k událostem došlo – obvykle na základě analýzy zpracované nástroji pro business intelligence nebo reporting formou controllingového procesu nebo přípravy informací pro management. Takový přístup je dnes již možno považovat za zastaralý a neodpovídající nárokům na moderní řízení obchodních procesů.

Naproti tomu zpracování událostí v reálném čase probíhá prakticky okamžitě, nejvýše se zpožděním několika málo sekund od chvíle, kdy k nim dojde. Na událost nebo kombinaci událostí je tak možné reagovat prakticky ihned. Příkladem je obchodování na kapitálovém trhu. Pokud dojde k zajímavé kombinaci pohybů cen předem vytipovaných komodit, je třeba reagovat ihned, nikoli až druhý den.

Ani pro běžné integrační nástroje není žádný problém reagovat na řádově desítky příchozích událostí za vteřinu. Pokud však jejich počet stoupne o tři nebo čtyři řády, potřebujeme specializované nástroje, které jsou schopné takové množství událostí zpracovat. Hlavní výhoda nástrojů pro CEP tkví v tom, že za vteřinu dokáží zpracovat řádově desetitisíce událostí z více různých zdrojů (obr. 1).

Obr. 1: Princip komplexního zpracování událostí

CEP se uplatní všude tam, kde potřebujeme být schopni zareagovat na náhodné nebo nepředvídatelné jevy. Nebude to dlouho trvat a zákazník v hypermarketu si naloží košík stovkami položek s RFID visačkami, projede bránou, která sejme veškeré zboží v košíku a téměř okamžitě dostane kompletní účtenku. Podobně ve

skladu, kam kamion vjede čtecí bránou, se celý jeho obsah prostřednictvím RFID visaček okamžitě sejme a řidič bezprostředně dostane pokyn, kam má pokračovat, zatímco skladníkům systém poskytne přesné pokyny, kam mají zboží z kamiónu uskladnit.

CEP se začíná využívat i v dalších odvětvích:

- Ve finančních službách se komplexní zpracování událostí osvědčilo jako základ jedněch z nejpropracovanějších systémů algoritmického obchodování na světě, správy rizik a souladu s legislativou. Finančním institucím umožňuje co nejlépe využít příležitostí k obchodování a generovat výrazné zisky.
- Díky schopnosti CEP monitorovat přenos dat v síti, analyzovat jej a reagovat na něj mohou telekomunikační společnosti realizovat infrastrukturu, která jim umožňuje zajistit odpovídající kvalitu služeb QoS jejich bezdrátových, širokopásmových a IP sítí 2,5 a 3. generace.
- Provozovatelé utilit využívají schopnost CEP shromažďovat a analyzovat data ze sítí pro přenos energie k získávání aktuálních informací v reálném čase o současné a budoucí poptávce po energii a o kapacitě přenosových sítí potřebné pro splnění těchto požadavků.
- V těžbě ropy a zemního plynu začínají společnosti využívat systémy komplexního zpracování událostí SCADA (Supervisory Control and Data Acquisition) ke zvýšení efektivity provozu potrubí a získávání informací v reálném čase o únicích médií a dalších varovných situacích ve výrobních a distribučních kapacitách.
- V armádě mohou nástroje pro komplexní zpracování událostí shromažďovat data přenášená satelity z ručně i automaticky řízených zdrojů a poskytovat pozemním jednotkám aktuální obrázek o tom, co se děje „za nejbližším kopcem".

Srdcem CEP je korelátor

Hlavním prvkem celého systému je event engine, neboli stroj pro zpracování událostí (jinak nazývaný korelátor). Tento proces leží v srdci celého systému a dokáže konzumovat a zpracovávat desítky tisíc událostí za vteřinu (viz obr. 2). Do korelátoru se prostřednictvím speciálních vývojových nástrojů zadávají specifikace událostních struktur, které chceme sledovat, a definice scénářů, podle kterých mají systémy na události (případně na kombinace těchto událostí) reagovat.

Obr. 2: Schéma komplexního zpracování událostí pomocí korelátoru

Součástí samotné práce korelátoru je algoritmus pro filtrování událostí. Vzhledem k tomu, že na pozadí pracují hardwarové technologie, může docházet k přeslechům, ke generování redundantních či nevýznamných událostí atd. Proto je v první fázi nutné tyto nevýznamné události odfiltrovat a pracovat jen se základními (atomickými) událostmi.

Události navíc téměř vždy přicházejí v nějakém kontextu a mohou mezi nimi být různé souvislosti. Sledováním vztahů různého typu mezi atomickými událostmi (jejich posloupnost, vlastní náplň, obsah apod.) můžeme generovat komplexní události (byznys události), které jsou z hlediska podnikových procesů významné a na které chceme nějakým způsobem reagovat. Průjezd kamionu je jedna atomická událost, zjištění nákladu a počtu kusů je druhá událost, určení místa pro zaskladnění je třetí událost. Z těchto atomických událostí se následně vygeneruje byznys událost – množství na skladu se zvětšilo o určitý počet kusů daných komodit.

Architekturu stroje pro zpracování událostí lze škálovat implementací masivně paralelní architektury zpracování, kde jsou toky událostí zpracovávány ve vláknech, a doplnit o optimalizátor, jenž rozhoduje, která událostní pravidla ze stovek či tisíců aktivních pravidel by měla být na daná data aplikována. Velký výkon stroje je nezbytný, pokud CEP systém přijímá událostí z automatizovaných zdrojů, kdy může jít o desetitisíce událostí za vteřinu. V takových případech stroj pro zpracování událostí obvykle využívá pipelining podobný jako u počítačového hardwaru. Zpracovává pak velké soubory toků událostí, aniž by docházelo ke zpomalení rychlosti zachytávání dat a funguje stejně jako CPU, který pro dosažení vyšší propustnosti využívá paralelní zpracování instrukcí.

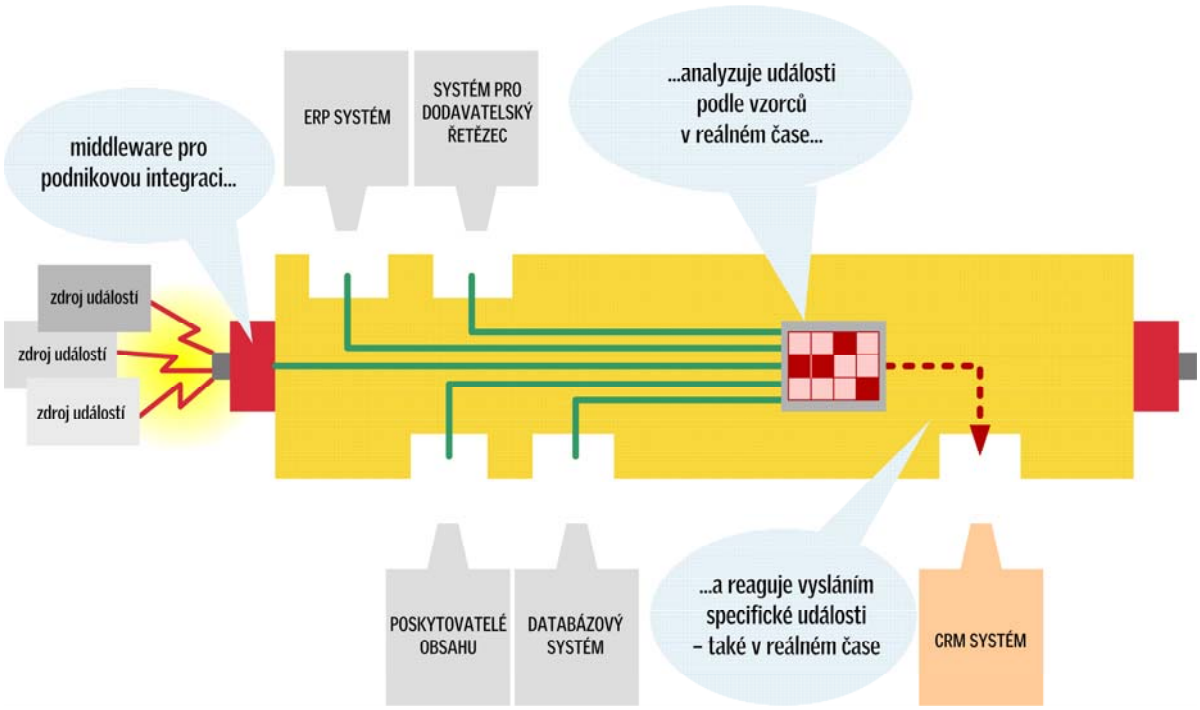
Tento poměrně obecný modul, který dokáže zpracovávat události různých typů, je třeba implementovat do systému přizpůsobeného potřebám konkrétního vertikálního segmentu. Dociluje se toho prostřednictvím adaptérů, tj. nástrojů, které určitým způsobem zprostředkovávají propojení a vytvářejí rozhraní mezi korelátorem (obecným strojem na zpracování událostí) a konkrétními daty přicházejícími z daného oboru (finanční data, skladová data atd.).

V průběhu zpracování událostí systém CEP často spolupracuje s různými back-office systémy (například klasickými ERP systémy), odkud načítá nebo kam zapisuje informace. Proto se systémy pro zpracování událostí neobejdou bez integračních nástrojů, například sběrnice podnikových služeb apod. Výkon takových back-office systémů je pro požadavky okamžité reakce na vzniklé události zpravidla nedostatečný. Tento problém je možné řešit pomocí tzv. Cache Forward Architecture. Tato architektura umožňuje umístění kritických dat v operační paměti s jejich automatickým obnovením v případě, že se v databázi změní. Tak je zajištěna aktuálnost dat z ERP systému dosažitelných rychlostí odpovídající přístupu do paměti.

Využití ESB

Podniková sběrnice služeb ESB (Enterprise Service Bus) se stala de facto integrační kostrou moderních servisně orientovaných architektur SOA. Propojením a zprostředkováním (mediací) služeb se ESB podílí na vytváření nových toků událostí s výraznou hodnotou pro podnik.

CEP umožňuje architektuře SOA řízené službami monitorovat a analyzovat více různých toků událostí procházejících sběrnicí ESB, identifikovat v nich určité struktury (například po A následuje B a poté C) a bezprostředně po rozpoznání určité hrozby nebo příležitosti vyvolat příslušnou akci. Místo aby se data průběžně shromažďovala a následně vyhodnocovala až na konci pracovního dne, mohou se systémy vybavené ESB infrastrukturou doplněnou o CEP rozhodovat tehdy, kdy je to skutečně potřeba – v reálném čase.



Obr. 3: Komplexní zpracování informací využívající podnikovou sběrnici služeb

Komplexní jazyky pro zpracování událostí

Stroj pro zpracování událostí realizuje instrukce, které dostává prostřednictvím jazyka pro zpracování událostí EPL (event processing language). EPL definuje stručnou syntaxi, jejíž pomocí hledá a reaguje na určité struktury, které se objevují v různých příchozích tocích událostí. EPL musí být navržen tak, aby zpracovával dotazy po částech – například jestliže A a B jsou pravda, pak pokud se C objeví v N minutách, vykonej určitou akci.

Jednoduchost použití mnoha CEP nástrojů kontrastuje se složitostí činností, které vykonávají. Například grafický nástroj pro modelování EPL umožňuje definovat vztahy a pravidla přetahováním myší, ale platforma pro zpracování událostí, na níž je nástroj založen, obsahuje propracovanou logiku pro zpracování dat z toků událostí prostřednictvím pravidel zavedených přes GUI. Tento přístup ke komplexnímu zpracování událostí se značně liší od strojů založených na tradičních pravidlech, které jsou navrženy tak, aby tato pravidla uplatňovaly na statickém souboru dat. Do systému pro zpracování událostí však vstupují toky dynamických dat a jejich způsob zpracování se výrazně liší.

Vizualizace CEP a automatizace reakcí

Vizualizace CEP funguje podobně jako palubní deska automobilu, jejíž kontrolky nepřetržitě a v reálném čase informují řidiče o tom, co se děje. Řídicí panel CEP poskytuje podobný přehled o aktivitách založených na sledovaných událostech v reálném čase, přičemž zobrazují aktivní pohled na realizované strategie a výsledná byznys rozhodnutí navrhovaná nebo učiněná CEP strojem.

Příkladem aplikace, která zobrazuje reprezentaci dat z událostí, je monitoring podnikových aktivit BAM (Business Activity Monitoring). Mnoho CEP systémů však vůbec řídicí panely nemá – logika zpracování událostí vyjádřená pomocí EPL a

vykonávaná CEP strojem často spouští akce ve fyzickém světě automaticky. Například aplikace pro správu sítě může odhalit útok typu DoS, okamžitě se rozhodnout odstavit daný směrovač v reálném čase a zapsat tuto akci do databáze údajů sloužících pro audit a odstranění poruch.

Databáze událostí a přístup k datům

Platforma CEP může obsahovat databázi událostí s dotazovacím jazykem EQL (Event Query Language), jehož prostřednictvím se lze nepřetržité dotazovat na data organizovaná podle času a typu událostí. Například vzájemný fond může udržovat časově utříděný seznam akcií v portfoliovém „koši", nepřetržitě porovnávat ceny jednotlivých akcií v koši oproti stále se měnícímu indexu a upravovat složení portfolia podle výše rizika spojeného s těmito pozicemi.

Podobně systém pro správu sítě, který nepřetržitě monitoruje její výkonnost, může využít komplexního zpracování událostí k úpravám přidělování šířky pásma tak, aby byly splněny požadavky na kvalitu služeb. Může také nepřetržitě zjišťovat, zda počet určitých aktivit přesahuje daný práh četnosti v určitých uzlech, což může znamenat počínající útoky DoS. Relační databáze se s takovými úkoly vyrovnávají mnohem hůře, protože jsou optimalizovány pro dotazy nad statickými daty a vytvářejí vztahy mezi daty na základě jejich hodnot, nikoli podle času a příčiny.

Naopak data, jež nejsou založena na událostech, mohou tvořit kritické vstupy do EPL logiky, která pak musí fungovat jak nad daty z událostí, tak z tabulek. Například RFID aplikace budou často vyžadovat přístup k produkčním datům získaným z ERP systému nebo systému pro jejich správu. Podobně mohou obchodní aplikace vykonávat různé obchodní strategie podle existujícího klientského portfolia nebo rizikového profilu – což jsou data, která pocházejí z back-office systémů.

Architektura Progress Apama

Architektonický návrh platformy Progress Apama pro komplexní zpracování událostí byl vytvořen koncem 90. let při výzkumu požadavků distribuovaných systémů prováděném na univerzitě v anglické Cambridgi. Tento výzkum ukázal, že zpracování událostí vyžaduje nový architektonický přístup, který se zásadně liší od klasických architektur pro zpracování dat.

Progress Apama je komerční realizace výsledků tohoto výzkumu. Architektura Apama je vybudována kolem modulárního škálovatelného návrhu, jehož základní funkcionalita by se dala shrnout následovně:

- monitoruje příchozí události z infrastruktury pro přenos zpráv nebo přenášené jako výstupní data z finančních či jiných trhů,
- analyzuje tyto události v paměti – a to buď samostatně nebo ve spojení s jinými událostmi, jejichž atributy a dočasné pořadí reprezentují určité vzorce,
- spouští odchozí události, které reprezentují akci učiněnou jako odezvu na výsledky analýzy.

Aby Apama mohla tyto operace provádět, obrací paradigma tradičních systémů. Místo obvyklého modelu „ulož-indexuj-dotazuj" Apama využívá korelátor pracující v reálném čase, v němž jsou předem definované vzorce (logický ekvivalent databázového dotazu) předem zavedeny do systému jako „scénáře" událostí. Tyto scénáře se spouští v závislosti na tom, zda jsou v tocích událostí detekovány předem specifikované události. Pokud události v datových tocích odpovídají podmínkám specifikovaným ve scénářích, dojde k vykonání specifických souborů akcí.

Obr. 4: Jednotlivé prvky architektury Progress Apama

Architektura Apamy vyniká svou schopností vykonávat až tisíce jednotlivých scénářů zároveň, přičemž každý má svou vlastní logiku monitorování toků událostí, je schopen vyhledávat vzorce událostí a po jejich nalezení spouštět předem definované akce. Architektura Apamy přitom funguje nikoli jako engine pro zpracování událostí, ale jako obsáhlá platforma pro zpracování událostí. Jako taková obsahuje:

- vývojářské nástroje, které pokrývají potřeby různých typů uživatelů,
- možnosti flexibilní integrace s obousměrnou konektivitou s mnoha rozmanitými zdroji toků událostí,
- graficky bohaté řídicí panely (dashboards) pro uživatelské sledování aktivit spojených se zpracováním událostí,
- nástroje pro testování a analýzu chování aplikací s využitím skutečných datových toků.

Zatímco tradiční architektury mohou reagovat na události až určitý čas poté, co tyto události proběhnou, architektura Apama řízená událostmi reaguje na rychle se vyskytující události jakéhokoli druhu v reálném čase. Využívá přitom platformu, která kombinuje flexibilitu, výkonnost a interoperabilitu.

Hlavní devízou Apamy je kombinace vytříbených analytických schopností a výkonnosti. Jde o mnohem více než pouhý engine pro zpracování událostí. Se svými propracovanými vývojovými nástroji, flexibilním testovacím prostředím, rozsáhlým integračním rámcem a graficky bohatými řídicími panely je Apama obsáhlá platforma pro komplexní zpracování událostí určená pro vývoj a provoz řešení řízených událostmi v reálném čase.