

Milena MACHÁČOVÁ*

VYUŽITÍ INTELIGENTNÍCH NÁSTROJŮ PRO ANALÝZU TECHNOLOGICKÝCH DAT

APPLICATION OF INTELLIGENT TOOLS FOR TECHNOLOGICAL DATA ANALYSIS

Abstrakt

Složité technologické procesy produkují v každém okamžiku obrovské množství heterogenních dat. Schopnost jejich sběru, ukládání a efektivního využití je, kromě jiného, limitována běžnými technikami datové analýzy. Lepší výsledky přináší nová generace inteligentních analytických nástrojů založených na bázi umělé inteligence, genetických algoritmů, pokročilých statistických metod, data warehousingu a dalších. Důležitou složkou je také vizualizace výsledných analýz pro koncového uživatele.

Abstract

In every moment technological processes produce a huge quantity of the heterogeneous data. The possibility of collecting, stocking and effective utilization of them is except other limited by conventional procedures of data-analysis. The new generation of intelligent instruments based on artificial intelligence, genetic algorithm, statistical methods, data warehousing technology and next ones make possible better results. The most important is also data visualization.

Key words: data warehouse, data mining, visualization system.

Úvod

Vývoj posledních let naznačuje, že kvalitní informace z technologického procesu a jeho znalost, jsou předpokladem pro jeho účinné řízení. Ukazuje se, že je nutné vytvořit virtuální prostředí, které bude schopno poskytovat základní prostředky pro identifikaci a vyhodnocování všech faktorů, které technologický proces ovlivňují a využít výsledky analýz v rámci operátorského a dispečerského řízení. Moderní automatizované řídicí systémy produkují značné množství dat, která obsahují informace o normálních i abnormálních stavech řízeného technologického procesu, o operativních změnách v jeho řízení aj. Objem dat je obecně tak velký a datové struktury natolik složité, že k jejich analýze není možno využít běžně užívaných analytických nástrojů. Současný rozvoj informačních technologií otevírá cestu novým, zajímavým možnostem pro sběr, uchovávání a vizualizaci dat v průmyslovém prostředí. Velké objemy dat z rozdílných technologických celků, v různé formě a struktuře, si vynutily vznik specializovaných databází dovolujících uchovávat i data historická v rozličném časovém členění. Zpracování dat v těchto rozsáhlých databázích, které jsou označovány jako datové sklady, je nesmírně náročné a vyžaduje použití zvláštních technik k tomuto účelu vyvinutých.

Teoretický princip datového skladu

Pojem datový sklad není v současné době chápán zcela jednotně, ve skutečnosti se pod něj zahrnují dva zcela odlišné pohledy na jeho architekturu. První z nich představuje filosofie **R. Kimballa** [2], která říká, že

* Ing., Institut ekonomiky a systémů řízení, HGF VŠB-TU Ostrava, 17. listopadu 15,
708 33 Ostrava-Poruba, e-mail: milena.machackova@vsb.cz

datový sklad je sjednocená množina datamartů (datových tržišť). Druhý pohled, jehož autorem je **W. H. Inmon** [2], je charakterizován jediným centrálním datovým skladem.

Datamart je zpravidla menší relační databáze sloužící pro uspokojení požadavků určité skupiny uživatelů. Datamarty mohou tedy mít teoreticky různou strukturu, takovou, která bude vyhovovat nástrojům a technikám pro zpracování uživatelských dotazů. Uživatelsky zaměřené datamarty mívají svůj vlastní pojmový aparát, vlastní historii dat a taktéž perioda aktualizace dat nemusí být pro všechny stejná. Důsledkem může být nekonzistence mezi jednotlivými datamarty, redundance v datech, odlišné dimenze pro stejný typ dat atd.

Uvedený přístup se jeví vhodným pouze pro projekt menšího datového skladu. Jeho výhodou je jednodušší implementace a menší počáteční finanční náročnost. Chybí zde však celkový pohled na data, který je možno získat pouze sjednocením datamartů, což vzhledem k výše uvedenému je velmi složitý proces, který může úspěšně proběhnout dle Kimballa pouze za předpokladu tzv. všeobecně přijatých dimenzí a faktů. Technická náročnost tohoto řešení vede v současné době k příklonu k řešení, které nabízí W.H. Inmon, tedy k myšlence centrálního datového skladu, ze kterého budou jednotlivé datamarty teprve druhotně odvozeny.

Pod pojmem **centrální datový sklad** se dle W. H. Inmona [2] rozumí integrovaná, předmětově orientovaná, nepodléhající změnám, časově proměnná kolekce detailních dat.

Integrace znamená, že data, která jsou ukládána do centrálního datového skladu, pocházejí z více provozních systémů. Předmětová orientace vymezuje určitou pojmovou aparaturu pro okruh dat, která datový sklad obsahuje, jejich vzájemné vztahy a členění. Schopnost uchovávaných dat nepodléhat změnám souvisí se skutečností, že v datovém skladu nejsou zachyceny všechny změny, ke kterým během dne došlo, ale pouze konečné, tzn. k určitému definovanému okamžiku, kdy mají být data z provozních systémů do datového skladu načtena. Data jsou tak ukládána po časových snímcích a vytvářejí posloupnost měnící se v čase tj. historii dat.

Zásadní přednost tohoto řešení spočívá v naprosté konzistenci dat, obsažených v centrálním datovém skladu, hovoří se o takzvané „jediné verzi pravdy“. Snižuje se zde také počet načítacích procesů z provozních systémů a jejich správa je jednodušší. Problém také nepředstavuje vytváření dalších datamartů v případě potřeby. Nevýhodou je složitější realizace a pomalejší implementace, kde na rozdíl od předchozího přístupu, kdy jednotlivé samostatné datamarty mohou vznikat postupně, je třeba mít navržen celý koncept řešení najednou.

Budování datových skladů eventuelně datových tržišť je náročný, nikoli jednorázový, ale nepřetržitý proces. Požadavky na design datového skladu zpravidla nejsou zcela přesně formulovány, dokud není alespoň částečně zaplněn a používán. Pro návrh datového skladu se využívají heuristické metody, kdy určitá fáze vývoje závisí zcela na výsledcích dosažených ve fázi předchozí.

Návrh datového skladu začíná vytvořením datového modelu, který se zpracovává v rozlišení na tři hierarchické úrovně, což umožňuje jeho postupné iterativní zavádění.

Při zakládání datového skladu je třeba také vzít v úvahu stupeň granularity uchovávaných dat v souvislosti s jejich nárůstem v průběhu delšího časového horizontu a řešit otázky eventuelního přetečení.

Každý datový sklad má svou vlastní originální architekturu, takže k jeho vybudování se používají vždy specifické nástroje. Může se jednat o normalizovaný nebo denormalizovaný entitně-relační datový model anebo model multidimenzionální se speciálním uložením dat ve schématu „hvězda“, „sněhová vločka“ či „galaxie“.

Hvězdicové schéma se skládá z rozsáhlé centrální tabulky – tzv. tabulka faktů a z dalších malých doprovodných tabulek pro každou dimenzi. Vizuální vyjádření konceptuálního schématu připomíná hvězdu, kdy okolo centrální tabulky faktů jsou na jednotlivých paprscích navěšeny tabulky dimenzí, kdy každá dimenze je reprezentována právě jednou tabulkou.

Uspořádání sněhová vločka je určitým druhem hvězdicového schématu. Rozdíl spočívá v tom, že tabulky dimenzí jsou normalizované, což znamená, že data pro jednu dimenzi jsou rozdělena do více tabulek. Grafické vyjádření konceptuálního schématu pak připomíná sněhovou vločku. Výhodou je snížení redundance v uložených datech, avšak nutné spojení více tabulek při kladení dotazu se může nepříznivě odrazit v časovém výkonu datového skladu.

V některých případech, kdy je vyžadováno použití několika tabulek faktů současně, aby mohly být sdíleny tabulky dimenzí, se přistupuje ke konceptuálnímu schématu galaxie nebo též souhvězdí.

Plnění datového skladu daty

Naplnění datového skladu daty není jednoduchou záležitostí. Neznamená pouze extrakci dat z provozních systémů a jejich umístění do datového skladu. Každá aplikace poskytující data byla vytvářena k určitému účelu, chceme-li tedy stahovat data z vícero aplikací, je zřejmé, že tatáž data se mohou vyskytovat na různých místech systému pod různými jmény, s různými rozměry jednotek, v rozličných formátech uložení atd. Z toho plyne, že než jsou data do datového skladu uložena, musí projít náročnými procesy, jejichž cílem je jejich integrace na stejnou platformu.

Speciální důraz je kladen na tzv. metadata, která definují vazbu na zdroje, odkud data do datového skladu přitékají. Metadata jsou uložena ve speciální databázi nazývané repozitář. Popisují jednak data, která datový sklad obsahuje, jednak zdrojová data a transformace, kterými data prošla před uložením do datového skladu.

Vzhledem k pravděpodobnému většímu rozšíření, věnujme se modelu datového skladu dle W.H. Inmona.

Obecně platí, že data se do datového skladu nezadávají, ale načítají z provozních systémů a z externích zdrojů. Plnění centrálního datového skladu zajišťují tzv. primární načítací procesy, které musí umožňovat extrakci vstupních dat prakticky z libovolného formátu, jejich transformaci, načtení do výstupního formátu pro uložení v datovém skladu a vlastní uložení.

V rámci transformace dat jsou vykonávány speciální operace jako:

- **Validate** – ověření věcné správnosti extrahovaných dat.
- **Čištění** – odhalování chybných dat, jejich odstranění či změna.
- **Integrace** – zajištění konzistence dat pocházejících z různých zdrojů zejména z hlediska datových typů, formátů, apod.
- **Derivace** – vytvoření odvozených dat na základě dat vstupních.
- **Denormalizace** – snížení počtu relací mezi tabulkami databáze.
- **Sumarizace** – vytvoření souhrnů z detailních dat.

Vlastní aktualizace datového skladu může probíhat několika způsoby. První z možností je uložit vždy znova celý obsah datového skladu, což lze zrealizovat pouze u velmi malých objemů dat. Další způsob je stahovat do datového skladu pouze data, která jsou v provozních systémech časově označena. Jinou možností je ukládání pouze nových přírůstků a změněných dat.

Centrální datový sklad tedy slouží jako společný zdroj vyčištěných a ověřených dat z provozních systémů i externích zdrojů. Následně se z něho pomocí sekundárních načítacích procesů mohou plnit specializované datamarty na základě zjištěných nových přírůstků nebo změn v datech.

Nástroje pro využití dat z datového skladu

Možnost efektivně a účinně využívat data uložená v centrálním datovém skladu je podmíněna existencí speciálních analytických nástrojů. Základními předpoklady jsou schopnost pracovat nad velkými datovými soubory, možnost práce s různými typy dat, postihu vztahů v mnohorozměrných datech, schopnost použití i při ne zcela jasně definovaném problému a v neposlední řadě musí nabídnout široké spektrum analytických metod.

V současné době se pro analýzu dat z datových skladů používá tzv. **Data Mining** (dolování dat), který beze zbytku naplňuje požadavky na analytické mechanismy kladené. Techniky Data Miningu vycházejí z induktivního a strojového učení, statistických metod, neuronových sítí a nejnověji se zde také uplatňují genetické algoritmy.

Úlohy pro dolování dat je možno rozdělit do níže uvedených kategorií [1]:

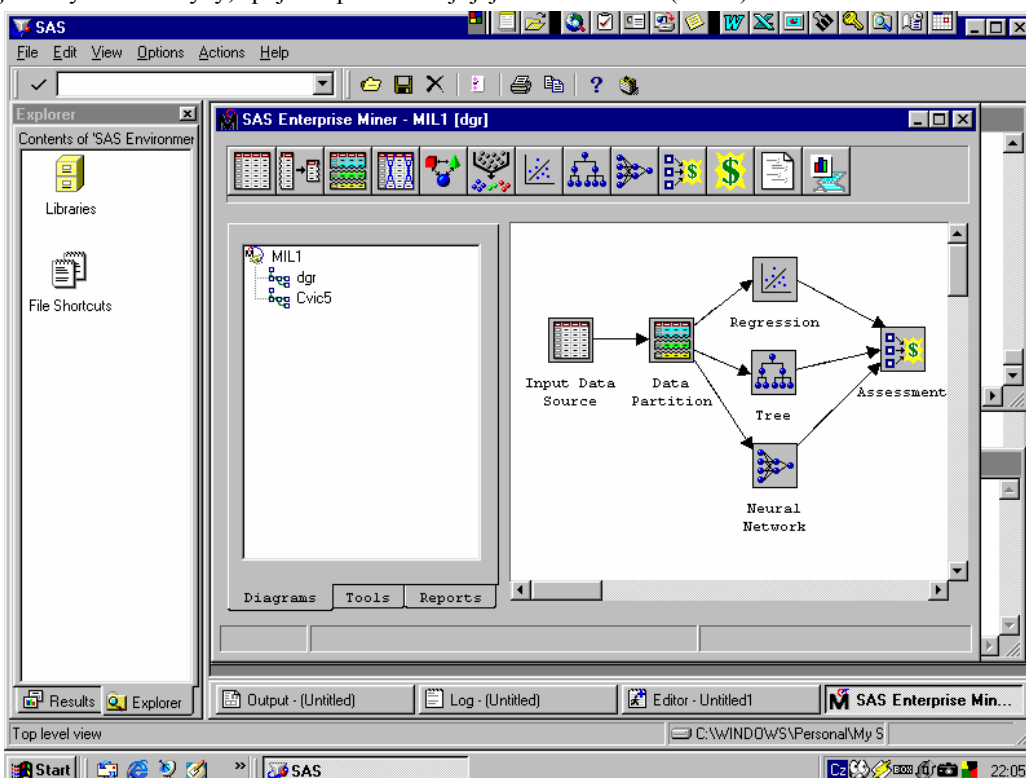
- **Klasifikace** – v procesu klasifikace se kategoriální data rozpadají do několika předem předdeklarovaných diskretních tříd. Cílem je nalezení odpovídajících vzorů atributů nebo charakteristik, které třídu identifikují. Výpočtem kategoriální proměnné pak klasifikace přiřazuje jednotlivé zkoumané případy do příslušných tříd. Získáním vzorů lze porozumět existujícím datům a předpovědět nové případy. Algoritmy klasifikace bývají často založeny na rozhodovacích stromech nebo neuronových sítích.
- **Regrese** – řada již dříve zjištěných hodnot je použita k předpovědi, jaké další hodnoty budou následovat. Nejjednodušší případy je možno řešit pomocí klasické lineární regrese, pro složitější úlohy je vhodné zvolit algoritmy neuronových sítí nebo rozhodovací stromy.

- **Shlukování** – výchozí databáze je segmentována do několika skupin tak, aby se tyto od sebe co nejvíce lišily a současně obsahovaly co nejpodobnější prvky. Každý záznam v databázi má většinou velké množství atributů, ale na rozdíl od klasifikace není jasné, podle kterých z nich bude seskupování probíhat. Vytvořené shluky mohou být vzájemně exkluzivní a vyčerpávající nebo mohou obsahovat bohatší, hierarchickou nebo lehce se překrývající reprezentaci. Shlukování bývá často úvodním krokem k data miningu. Identifikuje skupiny vztažených záznamů, které mohou být dále použity pro objevování dalších souvislostí.
- **Sumarizace** – zahrnuje metody pro hledání uceleného popisu podmnožiny dat. Důmyslnější metody dovolují odvozování sumarizačních pravidel, obsahují multivariantní vizualizační techniky a objevují funkční závislosti mezi proměnnými. Sumarizační techniky jsou často používány k interaktivní analýze dat a automatizovanému generování reportů.
- **Modelování závislosti** – spočívá v hledání modelu, který popisuje významné závislosti mezi proměnnými. Vytvořený model funguje ve dvou rovinách: strukturní úroveň modelu specifikuje (často v grafické podobě), které proměnné jsou navzájem místně závislé, kvantitativní úroveň modelu specifikuje sílu těchto závislostí pomocí zvolených numerických stupnic.
- **Detekce změn a odchylek** – soustřeďuje se na objevování nejvýznamnějších změn v datech od předchozího měření nebo na odchylky od závazné normy.
- **Predikce podle časových řad** – výchozí představa je stejná jako u regrese tj. z už existujících hodnot se předvídají hodnoty budoucí. Rozdíl spočívá v tom, že zjištěné hodnoty jsou závislé na čase. Prediktivní modely mohou používat různé časové měřítka.
- **Asociace** – jsou dány prvky, které se vyskytují společně v jednom záznamu nebo události. Nástroje pro vyhledávání asociací se snaží najít pravidla tvaru: je-li prvek A součástí nějaké události, pak platí, že z x % je součástí téže události i prvek B. Levá strana pravidla se nazývá antecedent, pravá část je konsekvent. Parametr x představuje faktor důvěryhodnosti.
- **Objevování posloupností** – je úzce spjato s asociací analýzou, avšak položky, které spolu souvisejí jsou obohacené o časový údaj. Cílem tedy není najít pouze seskupení prvků, ale naleznout také uspořádání jednotlivých událostí v čase, eventuelně délku časových intervalů mezi událostmi.

Dataminingové algoritmy jsou založené na vyhledávání tzv. vzorů v datech pomocí modelů [6]. Zkoumaná data se rozdělí na dvě eventuelně na tři skupiny (trénování, validační a testovací), na části dat je vytvořen model a nechá se učit tj. hledat vzory v datech. Naučený model pak slouží pro vyhodnocování zbylé části dat, která jsou předmětem dané úlohy.

Dataminingové modely mohou být dvojího typu, podle charakteru zadaného problému. Deskriptivní model popisuje vzory a existující vztahy v datech, které mohou ovlivnit rozhodování v aplikační sféře. Model prediktivní využívá nalezených vzorů v datech k předpovědi budoucího vývoje.

Ilustrací dataminingové úlohy může být její zpracování pomocí softwaru SAS Enterprise Miner [5]. Model je zde prezentován formou procesního diagramu (Proces Flow Diagram), kdy jednotlivé ikonky (nódy) představují určitý krok analýzy, spojnice pak definují jejich souslednost (obr. 1).



Obr. 1: Úloha Data Miningu v prostředí SAS Enterprise Miner

Typická úloha pro dolování dat se skládá z několika fází [1]:

- **Identifikace problémového okruhu pro zpracování**
Vztahuje se k požadavkům a cílům uživatele.
- **Vytvoření reprezentativního vzorku dat**
Představuje výběr množiny dat nebo podmnožiny proměnných případně vzorků dat, ve kterých bude vyhledávání uskutečněno. Předpokladem je identifikace všech datových zdrojů, které by mohly ovlivnit řešení úlohy.
- **Čištění a předzpracování dat**
Uskutečňují se operace sloužící k odstranění šumu v datech nebo k vyloučení irelevantních dat. Součástí je sbírání nezbytných informací pro modelování příčin šumu, nalezení strategie pro doplnění chybějících dat, úprava formátů dat apod.
- **Redukce a projekce dat**
Zabývá se vyhledáváním podstatných rysů v datech v závislosti na cíli úlohy, využívá redukci dimenzí nebo transformačních metod pro získání optimálního počtu uvažovaných proměnných apod.
- **Výběr kategorie úlohy pro dolování v datech**
Podle povahy problému je třeba stanovit charakter úlohy. Může se jednat o úlohu klasifikační, regresní, shlukovou analýzu apod.
- **Výběr dataminingového algoritmu**
Podle charakteru úlohy přiřazuje typ modelu, který bude použit – prediktivní, deskriptivní a současně určuje metodu pro extrakci vzorů z dat.
- **Vlastní dolování dat**
Probíhá s pomocí vybudovaného teoretického modelu z předchozí etapy.
- **Interpretace a vyhodnocení výsledků**
Vzory identifikované dolovacími mechanismy jsou interpretovány jako znalosti, které mohou být přímo včleněny do původního systému nebo mohou sloužit k podpoře rozhodování uživatele. V případě nepřesností nebo nesrovnalostí je možno se vrátit ke kterémukoliv z předchozích kroků.

Vizualizace dat z datového skladu

Nedílnou součástí konceptu datového skladu je vizualizační systém, jehož úkolem je předkládat uživateli výsledky analýz v názorné a srozumitelné podobě. Do pozadí ustupují obyčejné textové výpisy, současný vývoj grafických a multimediálních prostředí nyní dovoluje bezprostřední komunikaci uživatele s datovým skladem v mnohem komfortnějším a uživatelsky přívětivějším prostředí. Současně dochází ke zmnohonásobení použitelnosti a srozumitelnosti uchovávaných údajů.

K projektu vizualizačního systému pracujícího nad centrálním integrovaným datovým skladem se vážou určité koncepční zásady [4]:

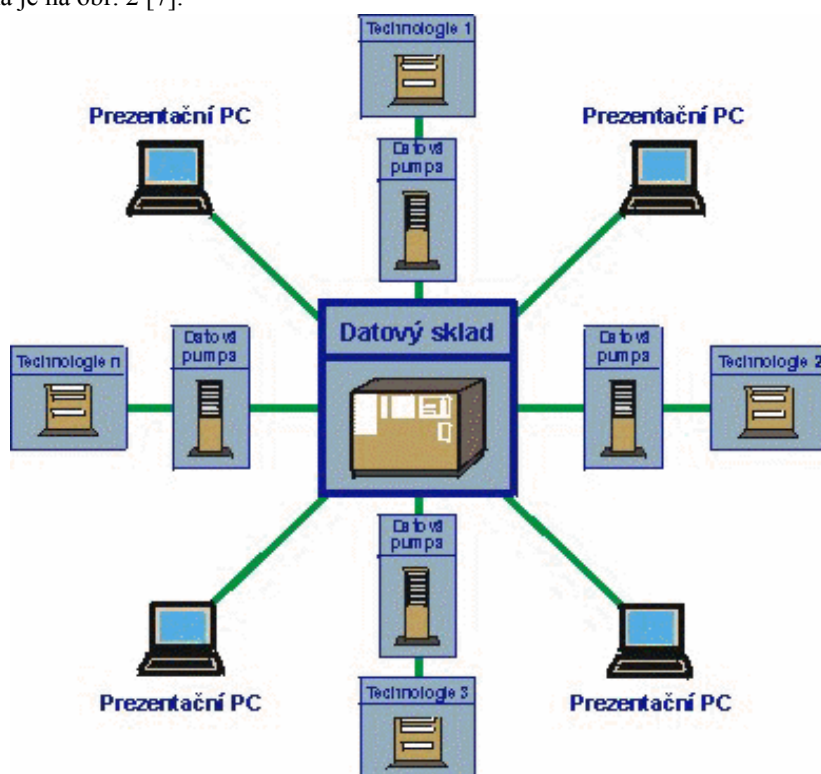
- Systém je vhodné koncipovat shora, tj. jako integrovaný řídicí a informační systém. V tomto pojetí se jednotlivé problémově orientované aplikace realizují samostatně k tomuto účelu implementovanými (resp. pořízenými) softwarovými prostředky.
- Celkové řešení, resp. funkce zaváděné z hlediska potřeb úrovní řízení nad jednotlivými dílčími oblastmi, je implementováno pomocí nástroje zvláště vyvinutého pro účely integrování informace z heterogenních zdrojů dat, pro jejich vizualizaci, zpracování a archivaci.
- V celém životním cyklu řídicího a informačního systému má mimořádný vliv jeho architektura a to jak architektura "vertikální" (členění od úrovní řízení a měření, PLC, dispečerská úroveň čili vizualizace technologie, úroveň operativního řízení čili vizualizace všech dat), tak architektura "horizontální" (dekompozice do aplikačně a logicky uzavřených samostatných subsystémů a jejich distribuovaná organizace na jednotlivých vertikálních úrovních).
- Mimořádnou roli v návrhu architektury hraje koncept tzv. zapouzdření. Umožňuje izolovat prvky subsystému, které je nutné ochránit před nežádoucími externími zásahy a současně i definovat rozhraní

představující nástroje dovolené manipulace se systémem. Má to pozitivní vliv na bezpečnost i udržitelnost systému a snižuje to případné problémy při hledání závad.

- Je důležité dimenzovat systém s přiměřenými rezervami. Lze předpokládat, že po uvedení systému do chodu budou následovat požadavky na jeho rozvoj resp. úpravu podle požadavků uživatele na základě zkušeností s jeho používáním. Tyto požadavky mohou mít pro systém často větší význam, než některé původně navržené funkce a vlastnosti. Systém musí umožnit jejich přirozenou realizaci.
- Možnost konfigurovatelnosti pohledů na stav systému individuálním uživatelem umožňuje jeho efektivní a individualizovaně pojímané využití.
- Pro trvalý rozvoj systému je důležitá i otevřenost spodních úrovní řešení, při dodržení pravidel o bezpečnosti interakce a přístupu ke zde monitorovaným údajům.
- Otevřenost rychle se rozvíjejícím novým technologiím umožňuje systému přetrvat delší časové období bez nutnosti větších změn v koncepci jeho rozvoje a využívání.

Ideový návrh vizualizačního systému

Integrovaný informační systém pro sběr, analýzu a vizualizaci dat z více heterogenních zdrojů je postaven na myšlence datového skladu a internetových technologií. Jeho úkolem je zpracovávat data z libovolného počtu rozdílných technologických celků, které jsou umístěny na různých místech v rámci podniku. Ideové schéma je na obr. 2 [7].



Obr. 2: Ideové schéma vizualizačního systému

Hierarchicky se jedná o nadřazenou vrstvu softwaru, která se nepodílí na přímém technickém řízení technologií, ale na vyhodnocování vztahů mezi technologiemi. Základním principem je **centralizace dat získávaných z různých zdrojů, jejich zpracování na jednom místě a distribuování v požadované formě uživateli**. Zdroje dat mohou být uloženy na různých datových serverech nebo technologických (velinových) vizualizačních počítačích, které archivují data příslušné technologie nebo provozu a jsou přístupné počítačovou sítí.

Transport zdrojových dat zajišťují speciální počítače nazývané **datové pumpy**. Obecně mohou technologické systémy produkovat různé objemy dat se specifickými formáty uložení a taktéž disponovat různým rozhraním. Úkolem datové pumpy je zajišťovat spojení s provozními počítači. Data se zde načítají, verifikují a archivují pro případ nečekané poruchy. Datová pumpa mění formát a strukturu přichozích dat na

formát vyžadovaný datovým skladem a následně takto upravená data do datového skladu transportuje počítačovou sítí. Protože datová pumpa celý proces opakuje, lze použít představu pumpování dat z provozu na centrální počítač.

Datový sklad je zde počítačový server, na kterém je uložena speciálně koncipovaná databáze spolu s aplikačními programy pro její chod. Vstupní modul datového skladu přijímá data zasílaná datovou pumpou a řadí je do databáze. Přijaté údaje jsou však rovněž zpracovány podle požadovaného předpisu, který může být jedinečný pro každý vstupní údaj. Toto zpracování se nazývá agregací údaje a je popsáno agregační funkcí. Agregační funkcí může být např. hodinový průměr údaje, směnový průměr, nejvyšší hodnota za hodinu, výsledek z výpočtu podle vzorce apod. Agregační funkce může být pro každý vstup jiná. Údaje v datovém skladu jsou posléze archivovány. Případné změny v zadání jsou umožněny několika editačními okny. Na serveru datového skladu je rovněž spuštěn program www server. Ten umožňuje vlastní komunikaci s uživatelem pomocí HTML dokumentů.

Pro samotného uživatele je nejdůležitější vizualizační nebo také prezentační část systému, se kterou přichází do styku [3]. Komunikace uživatele se systémem je jednoduchá. Je založena na internetovském prohlížeči, pomocí kterého si uživatel načte z datového skladu úvodní obrazovku. Podobně jako na Internetu i zde mu dokument nabízí různé volby a možnosti. Uživatelé jsou tak k dispozici data z různých podnikových provozů, které jsou do systému připojeny. Většinou jsou tato data už předem předzpracovaná, agregovaná, což je činí pro uživatele zajímavějšími z hlediska jejich vypovídací schopnosti. K jejich zobrazení slouží jednotlivé nástroje jako např. tabulky, grafy, trendové závislosti, technologický snímek apod. Důležitou vlastností je, že **uživatel má k dispozici údaje z různých provozů současně**, což mu umožňuje je vzájemně porovnávat nebo hledat mezi nimi neznámé vazby, jež díky oddělenému chodu provozních počítačů nebyly zřetelné.

Data pocházející z různých informačních zdrojů a prezentovaná uživateli mohou mít multimediální charakter. Výstupy z aplikací jsou představovány ve formě grafiky, zvuku, nejrůznějších animací, videa apod. Multimedia otevírají nové možnosti v přijímání informací. Ztělesňují nový typ rozhraní aplikace umožňující rychlou a srozumitelnou komunikaci uživatele s počítačem.

Důležitou součástí vizualizačního systému je subsystém registrující poruchy, závady a mimořádné stavy. Umožňuje provázání událostí s trendy. Trendy zobrazují analogové veličiny příslušné technologické události a zároveň k nim přiřazený digitální vstup provázaný časovým ohodnocením. Lze tedy identifikovat typ události, průběh veličin i stav vykonané akce v jistém časovém okamžiku.

Specializovaný software, na bázi technik pro dolování dat, dokáže prohledávat shromážděné údaje a hledat typické kombinace poruch a mimořádných stavů. Tímto vyhledáváním typických **vzorů událostí** pomáhá provoznímu dispečerovi při odstraňování poruch a závad a také i k jejich určité predikci. Systém současně umožňuje vedení dokumentace o přijatých opatřeních k odstranění poruch, závad, časový postup apod.

Systém může mít výstup jednoduchých zpráv na pager, mobil nebo e-mail a taktéž může disponovat hlasovým výstupem - hlasovými výstrahami z počítače. Předpokladem těchto funkcí je existence specializované databáze pro výstrahy, akustická hlášení a podpůrné systémy, která je uložena na samostatném počítači.

Sběr dat se řeší prostřednictvím datových pump, jejichž výstupy proudí do části databáze určené pro tento subsystém. Pro odstraňování poruch slouží sada nástrojů, které dispečerovi pomohou ve hledání příčiny poruchy a v nalezení opatření k odstranění poruchy.

Ke každému typu poruchy se asociuje zvolená množina vstupů, jejichž hodnoty mohou být s konkrétní poruchou ve vztahu příčina - následek. V případě potřeby je možno soubor poruchových hlášení zobrazit na obrazovce. Kliknutím myši na zvolené chybové hlášení je možno vyvolat automatické zobrazení průběhu asociovaných vstupů v časovém intervalu odpovídajícím době výskytu poruchy a následně interaktivně listovat zobrazeným seznamem v čase vpřed i vzad.

Pokud již v databázi budou uloženy údaje o stejném typu poruchy vzhledem k té, která právě vyvstala, je možné informaci o předchozí poruše vyvolat ve zvláštním okně, aby se daly porovnat detailnější podrobnosti. Každá porucha by měla být doprovázena návodem, popisujícím jaká opatření je možno použít pro její odstranění. Opatření mohou mít obecný charakter nebo budou využívat již existujících informací uložených v databázi týkajících se řešení případů konkrétních, již dříve se vyskytnuvších poruch.

Závěr

Mnohorozměrná, složitě strukturovaná data, snímaná z technologických procesů, jsou bohatým zdrojem informací, které je možno využít na podporu efektivního řízení těchto technologií. Zkvalitnění rozhodovacího

procesu přináší využívání nových, inteligentních analytických nástrojů, které jsou schopny pracovat nad velkoobjemovými soubory dat, umožňují práci s různými datovými typy, ale především se snaží postihnout skryté závislosti mezi zkoumanými daty. Maximální účinnost těchto nástrojů je podporována vznikem specializovaných databází, datových skladů. Výsledky analýz jsou uživateli předkládány v názorné, srozumitelné podobě prostřednictvím vizualizačního systému, který pracuje nad datovým skladem.

Literatura

- [1] Fayyad, M.U.: Piatetsky-Shapiro G., Smyth P., Uthurusamy R.: *Advances in Knowledge Discovery and Data Mining, USA: AAAI Press, 1996.*
- [2] Inmon, W.H.: *Building the Data Warehouse, Third Edition, USA, Wiley Computer Publishing, 2002.*
- [3] Štefánek, J.: *Datový sklad a vizualizace dat technologických procesů.* In *Informační technologie pro praxi. Ostrava: TANGER s.r.o. Ostrava, 1999, 7 s.*
- [4] Provazník, P.: *Dolování dat – vizualizace dat jako klíčový prvek data warehousingu. IT systém, 2000, roč.2, č. 11, s. 54 – 55.*
- [5] *Cook book Data Mining Using Enterprise Miner software. A Case Study Approach published by company SAS Institute , 2000.*
- [6] *Cook book Getting Started with Enterprise Miner Software. Version 4.0 published by company SAS Institute, 2000.*
- [7] <http://www.mediumsoft.cz>

Summary

By the recent years development it has been outlined that a highly quality information following from technological process and its knowledge represent a prerequisite for efficient management. It is obvious it is necessary create a virtual environment which would be capable to provide basic means for identification and evaluation of all factors which could affect technological process and to make use of analytical results within framework of operator and dispatcher control and managements activities. The existing development of information technologies has opened the way for new interesting possibilities of collecting, saving and visualization of data in an industrial environment. Huge amounts of data from various technological units and of various forms and structures have provoked rise of specialized databases called data warehouse which has enabled to save historical data with various time schedule.

The meaning of the term data warehouse is not understood unambiguously in fact it includes two quite different views on its architecture. The first of them is represented by philosophy of R. Kimball saying that data warehouse is a unified entity of datamarts. According to the second view of W. H. Inmon it is featured only as one central data warehouse.

According to W. H. Inmon meaning of central Data warehouse it is an integrated, subject oriented, no subject changes time variable collection of detailed data. The integration means that data which are saved by central data warehouse can come from more operational systems. The subject orientation defines a certain term apparatus for the range of data which are saved in data warehouse their mutual relations and breakdown. The capability of data warehouse not to be subject to changes is connected with the fact that not all changes which have occurred during a day are caught in data warehouse, but only the final ones, i.e. at a certain defined moment when such data are to be entered from operation systems into data warehouse. Thus data are stored in time snapshots and a sequence changing in time, i.e. data history, is created by them.

Each data warehouse has got its own architecture so that always specific tools are applied for its building up. This is a matter of normalized or non-normalized entity-relationship data model or multidimensional model with special data storage scheme called “star”, “snow flake”, “galaxy”.

Data does not feed into data warehouse but they enter from operational systems or external sources. Data warehouse filling ensure so-called primary entering processes.

The possibility to utilize effectively data that are saved in data warehouse is conditioned by existence of special analytical tools. The basic prerequisite is their capability to work on huge amount of data, to work on

various types of data, capability application in case not completely defined problem and last not least a wide spectrum of analytical methods.

Data Mining is the most important technique for data analyse based on inductive and machine learning, statistic methods, neural networks, genetic algorithms and so on. This is a process on searching for patterns in data through models. There are two model types according to nature of the problem given. Descriptive model describes patterns and existing relationship in data, prediction type of model make use of patterns found in data for prediction of future development.

An integral part of data warehouse is visualization system. Its task is to submit end user the results of analyses with well arranged and understandable form.

Recenzent: Prof. Ing. Jaroslav Balátě, DrSc., Univerzita Tomáše Bati ve Zlíně