Masarykova univerzita

Fakulta informatiky

Manažerské informační systémy a jejich úloha v řízení podniku

Diplomová práce

Brno, leden 2008 Bc. Michal Kašík

Poděkování

Na tomto místě bych rád poděkoval panu prof. RNDr. Jiřímu Hřebíčkovi, CSc., vedoucímu mé diplomové práce, za uvedení do problematiky, jeho ochotu a cenné náměty využité při tvorbě této práce. Rád bych také velice poděkoval mému otci Václavu Kašíkovi za informace poskytnuté k problematice stavebnictví a stavebních firem. Dále bych chtěl poděkovat za pomoc panu Zdeňku Kratochvílovi z Jihlavské stavební s.r.o. za formulaci požadavků a hodnocení navrženého MIS.

Shrnutí

Nasazení manažerského informačního systému je vždy strategickým rozhodnutím. Jedním z klíčových ukazatelů pro vhodný výběr produktu je také jeho návratnost. Použitím open source řešení lze v případě malých společností tuto návratnost očekávat v kratším čase, neboť náklady na pořízení vlastního produktu jsou v tomto případě podstatné. Problematikou open source manažerských systémů v oblasti Business Intelligence se zabývá tato práce.

Klíčová slova

Manažerský informační systém, MIS, Business Intelligence, BI, ETL, datový sklad, Data Warehouse, OLAP, reporting, dolovaní dat, data mining.

Prohlášení

Prohlašuji, že tato práce je mým původním autorským dílem, které jsem vypracoval(a) samostatně. Všechny zdroje prameny a literaturu, které jsem při vypracování používal(a) nebo z nich čerpal(a), v práci řádně cituji s uvedením úplného odkazu na příslušný zdroj.

Obsah

1. Úvod

V současnosti je stále více kladen důraz na efektivní řízení společnosti. Společnosti jsou vedle zefektivnění výroby, snížení nákladů či zvýšení kvality produktů nuceny také analyzovat nové možnosti nebo získávat významná data nejen z vlastních, ale také z externích systémů. Manažerské informační systémy se tak stále více stávají nepostradatelnou součástí podnikových systémů a jejich úloha stále roste.

Pořízení open source produktů do společnosti snižuje počáteční náklady a stává se tak vhodným řešením především pro malé a střední společnosti, které mohou zvýšit rychlost návratnosti investic do těchto systémů – ROI (Return On Investment). Investicím se však samozřejmě společnost nevyhne, neboť je nutné započíst náklady na implementaci i při řešení vlastními silami, avšak takto je mohou výrazně snížit. Pro větší společnosti však cena open source systémů nehraje roli, protože vedle výše nákladů na analýzu a implementaci se výhoda open source vytrácí. Použití open source právě v těchto společnostech se tak omezuje spíše na rozsáhlejší možnosti nastavení a úprav, kde proprietární systémy toto významně nepodporují.

V současnosti lze použít open source ve společnosti od operačního systému (např. Linux), přes kancelářské programy (OpenOffice.org) až po ERP systémy jako například OpenBravo. V případě manažerských informačních systémů (MIS) je situace jiná, v open source se vyskytují pouze ty, které pokrývají problematiku

Business Intelligence. Open Souce Business Intelligence (OSBI) nástroje pokrývají především oblast transformace (ETL), datových skladů (DW), analýz (OLAP), dolování dat a reportování. Lze se tak setkat buď s jednotlivými nástroji nebo s kompletními balíky těchto nástrojů.

Druhá kapitola uvede problematiku informačních systémů, MIS a zařazení problematiky business intelligence do podnikových informačních systémů. V této kapitole je dále uvedena architektura MIS, podnikových informačních systémů a komponent Business Intelligence.

Ve třetí kapitole je představena problematika open source, způsoby vývoje open source software a základní koncepty licencování těchto softwarových produktů.

Čtvrtá až osmá kapitola se tak bude zabývat jednotlivými komponentami MIS, kde na začátku kapitoly je uvedena problematika, řešená těmito komponentami, v závěru každé kapitoly jsou uvedeny vybrané open source produkty vhodné k řešení dané problematiky, tabulkové srovnání těchto produktů je uvedeno v příloze 1. Výběr komponent do tohoto srovnání byl závislý hlavně na existující alespoň základní dokumentaci v anglickém popř. českém jazyce a možnost podpory alespoň pomocí diskuzních fór. Díky tomuto omezení tak nebylo do srovnání zařazeno několik zajímavých produktů, které však pro implementaci z těchto důvodů jsou nepoužitelné.

Devátá kapitola se zabývá celými balíky komponent pro řešení problematiky BI, které obsahují vlastní server a dále většinou převzaté komponenty z předchozích kapitol. Výběrem z těchto balíků pak bude realizován MIS pro vybranou společnost. Jelikož nemusí právě všechny komponenty v těchto balících vyhovovat našim potřebám, lze, je-li to možné, je nahradit některou z komponent uvedených v předcházejících kapitolách.

V desáté kapitole je popsána vybraná společnost, ve které bude navrhovaný systém implementován. Dále je zde popis realizace, včetně požadavků na systém a architektury. Takto vytvořený systém se pak nachází na přiloženém CD.

V závěru je zhodnocena realizace pomocí zvoleného produktu a připomínky k samotnému vývoj produktu.

Příloha 1 obsahuje srovnání vybraných aplikací formou grafu a hodnocení, příloha 2 obsahuje fyzický model zdrojové databáze a schémata konfiguračních souborů

2 Informační systémy

Informační systémy (IS) jsou strukturované komplexy technik, nástrojů a zdrojů umožňující ukládání, zpracování a prezentaci dat. [19] Informačním systémem tak nemusí být nutně systém podporovaný nebo řešený softwarovými nástroji. Řešení systému v podnicích úplně bez užití softwarových nástrojů je v současnosti téměř nemožné, neboť jsou kladeny vysoké nároky evidenci účetních záznamů, dokumentaci, plnění norem (např. ISO 9001, 14000 apod.), personalistiky a dalších v závislosti na oboru podnikání jednotlivých společností. Především v prostředí malých organizací se však lze setkat jen s částečnou integrací softwarových nástrojů do informačního systému podniku, kde tato skutečnost je důsledkem buď obav z nového IS nebo neefektivností vynaložených nákladů na realizaci této části systému. V prostředí středních a velkých firem je situace jiná, neboť návratnost investovaných nákladů zde může být kratší, obzvláště je-li doprovázena změnou podnikových procesů (BPR).

2.1 Informační systémy v podniku

Primárním cílem podniků je generování zisku. Výroba produktů, poskytování služeb, snižování nákladů, propagace, optimalizace procesů apod. Úlohou informačních sytémů v podnicích je tak podpora všech těchto procesů. Lze se setkat s následujícími primárními (označované též OLTP, transakční nebo „legacy“) systémy, které zajišťují provoz společnosti. Následující definice jsou

převzaté z ČSSI[1]:

- Systémy na podporu vztahů se zákazníky – Customer Relationship Management (CRM), systém maximalizující spokojenost a loajalitu zákazníků k zajištění profitu společnosti.
- Systémy pro plánování podnikových zdrojů – Enterprise Resource Planning (ERP), systém podporující řízení a koordinaci všech disposibilních podnikových zdrojů a aktivit s cílem zajištění potřeb trhu a vlastních potřeb podniku. ERP systémy pokrývají všechny základní oblasti podnikového řízení: prodej, nákup, sklady, finanční účetnictví, controlling, majetek, lidské zdroje, práce a mzdy, technickou přípravu výroby, plánování výroby a podporují operativní řízení včetně dílenského řízení výroby.
- Systémy pro řízení dodavatelského řetězce (SCM), řízení všech procesů v rámci dodavatelského řetězce počínaje zajištěním surovin pro první článek řetězce přes zhotovení produktu a konče dodávkou konečnému spotřebiteli posledním článkem řetězce. Tyto procesy se integrují na bázi informačních a komunikačních technologií a zahrnují činnosti uvnitř i vně podniku.

OLTP systémy jsou optimalizovány pro velké objemy transakcí zpracovávaných v reálném čase. Obsahují velké objemy provozních dat, ale pro potřeby dalšího zpracování a analýzy těchto dat nejsou vhodná. Mezi další části podnikových systémů se řadí například systémy pro řízení podnikové dokumentace (DMS), systémy pro řízení managementu jakosti (QMS), systém enviromentálního

1 Česká společnost pro systémovou integraci, http://www.cssi.cz

managementu (EMS), ale i informační systémy dodavatelů, odběratelů, státní správy a další.

Pro potřeby efektivního řízení jsou do společností také implementovány informační systémy pro podporu podnikového managementu. Úkolem těchto systémů je nabízet transformovaná data z ostatních systémů nejen v podniku pro efektivní řízení společnosti. Umožňují provádět analýzy výroby, provozu, prodeje, připravovat podklady (reporty) či objevovat dříve neznámé souvislosti. Jsou souhrně označovány jako manažerské informační systémy a stojí v hiearchii odnikových systému na úplném vrcholu – schéma 1.

schéma 1 – MIS v podniku

2.2 Manažerké informační systémy

2.2.1 Historie MIS

Historie manažerských informačních systému začala již v říjnu 1958, kdy vychází v IBM Journal článek „A Business Intelligence System“ od H. P. Luhna. V té době bylo zpacování dat velmi limitováno a jednalo se tak spíše o koncept distribuce dat. Další významný posun nastal až v sedmdesátých letech minulého století, kdy společnost Lockheed použila interaktivní aplikaci pro manažery MIDS – Management Information and Decision Support. V osmdesátých letech začaly

vznikat první významné práce k tomuto typu aplikací jako například článek „CEO Goes On-Line“ od autorů John Rockarta a Michael Treacyho vydaný v roce 1982 v Harvard Business Review. V druhé polovině osmdesátých let se začínají objevovat první komerční EIS (Executive Information System) produkty založené na multidimensionálním zpracování a uložení dat. Od začátku devadesátých let se tyto produkty začínají objevovat i na českém trhu. Ve stejném období se také začínají objevovat řešení založená na datových skladech (DW – Data Warehouse) a datových tržištích (DMA – Data Mart), za kterými stojí Bill Inmon a Ralph Kimball. Díky tomuto kroku se zpracovávají velké objemy dat, což umožňuje nasazení nástrojů dolování dat (DMI – Data Mining), využívající matematických a statistických metod.

Manažerský informační systém můžeme definovat jako sadu postupů, procesů a technologií na základě kterých lze ze všech dostupných zdrojů poskytnout informace potřebné pro efektivní řízení organizace a to ve formě pochopitelné člověkem.

K pojmu Manažerský informační systém lze přistupovat ze dvou úhlů pohledu. Prvním z nich je chápání MIS jako soubor technologií, postupů a procesů řešící oblast podpory pro rozhodování veškerého managementu společnosti nebo jako k systému řešící problematiku pro danou část managementu. MIS se tak mohou lišit dle jejich zaměření. Pro nižší a střední management jsou to systémy pro podporu rozhodování (DSS), umožňující provádění analýz a tvorby reportů pro jednotlivá oddělení, jehož uživatelé se starají o efektivní, včasné a kvalitní realizace objednávek výrobku a služeb pro zákazníka. Pro vrcholový management, který určuje strategii podniku, jsou to systémy EIS, které integrují nejdůležitější datové zdroje. Nabízí práci s daty z interních a externích zdrojů. Umožňují

pracovat s daty v agregované formě, poskytují on-line analýzy trendů, drill-down, drill-up, slicing a dicing. Jsou ovladatelné myší v grafickém uživatelském rozhraní, využívají manažerského pultu (dashboard), pro zobrazení klíčových ukazatelů na jedné obrazovce. V současnosti se postupně techniky a nástroje z EIS systémů přesouvají i pro střední a nižší management. Další částí, kterou manažerské informační systémy pokrývají, je oblast Expertních systémů (ES).

2.2.2 Architektura

Zdrojem dat pro MIS jsou podnikové informační systémy, systémy dodavatelů a odběratelů, státní správy a další informační systémy jako například systémy, zajišťující vývoj měnového kurzu či předpověď počasí. K datům těchto systémům lze přistupovat přímo, nebo je uložit do datového skladu. Ten pak slouží jako zdroj dat například pro prováděné analýzy, reporty nebo monitoring. Schéma architektury MIS je znázorněno na schématu 2.

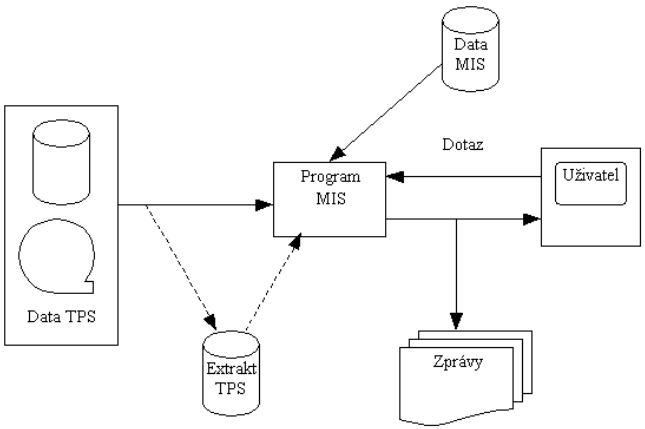


Schéma 2 – Architektura MIS převzato z [26]

2.3 Business Intelligence

Stále častěji se lze namísto pojmu Manažerský informační systém setkat s anglickým slovním spojením Business Intelligence (BI). Oblast, kterou BI pokrývá, není sice přesně vymezena, ale její primární určení je pro podporu tvorby obchodní strategie a marketingu. Příkladem sporných oblastí občas zahrnovaných do BI zde může být Competitive Intelligence (analýza konkurence a konkurenčního prostředí), expertních systémů nebo DSS, které však mohou být chápány i jako samostatné celky. Pro Business Intelligence je obecně přijata tato definice: *Business intelligence (BI) je sada postupů, procesů a technologií, jejímž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Představuje komplex aplikací, které podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na specifických, tzv. OLAP (On-Line Analytical Processing) technologiích a jejich modifikacích.*[18]

Business Intelligence lze spíše charakterizovat jako problematiku v řízení podniku, která je řešena manažerské informační systémy, což může být příčinou, proč tyto dva termíny jsou často zaměňovány. Důvodem bývá také skutečnost, že mezi prvními implementovanými systémemy s manažerskými funkcemi do společnosti jsou právě ty, které řeší problematiku spadající i pod BI, především pak OLAP nástroje. Tyto systémy se ostatně také vyskytují v drtivé většině společností, které využívají manažerské informační systémy. Zařazení problematiky BI v rámci podnikových systémů je zobrazeno na schématu 3.

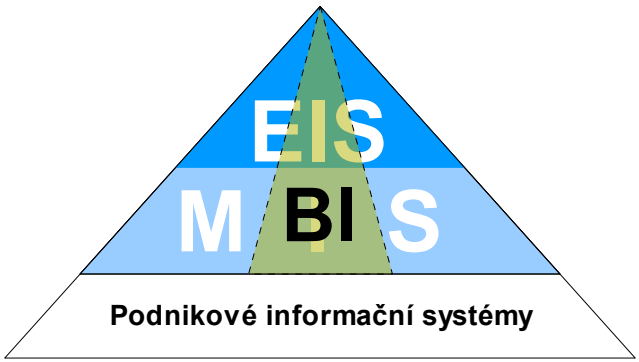


Schéma 3 – oblast BI v podnikových systémech

Mezi běžné komponenty řešící problematiku BI a MIS se řadí následující nástroje podle [4]:

- Transformační nástroje (ETL)
- Integrační nástroje (EAI)
- Datové sklady (DWH)
- Datové tržiště (DMA)
- Dočasné úložiště dat (DSA)
- Operativní úložiště dat (ODS)
- OLAP nástroje
- Reportovací nástroje
- Manažerské aplikace (EIS)
- Dolování dat (DMI)
- Nástroje pro zajištění kvality dat
- Nástroje pro správu metadat
- Produkční a zdrojové systémy
- ostatní

Architektura komponent MIS a BI je zobrazena na schématu 4.

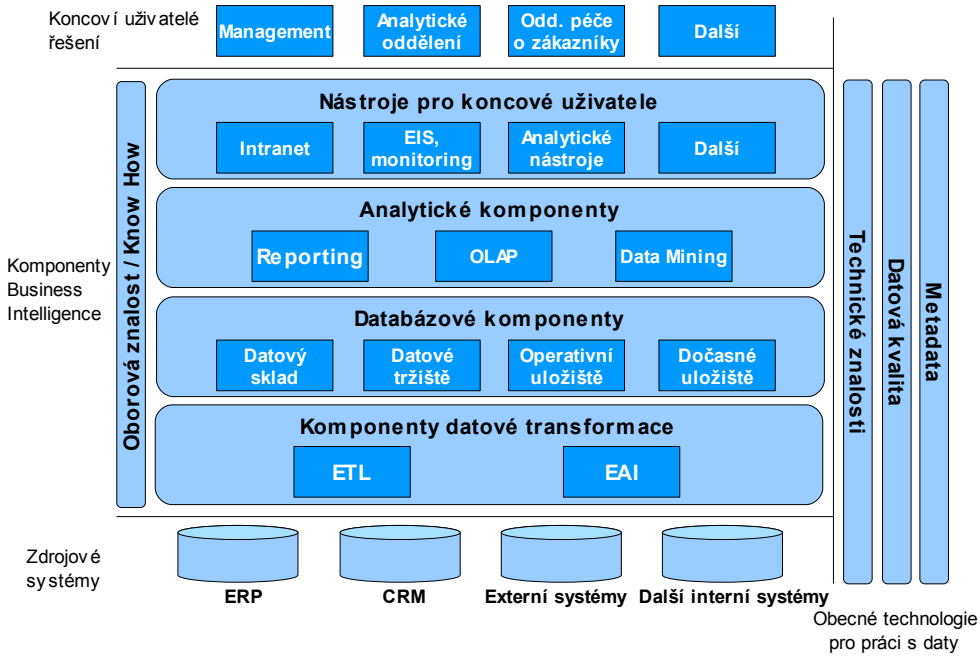


Schéma 4 – Komponenty BI a MIS, převzato z [4]

2.3.1 Transformační komponenty

Transformační vrstva slouží k transformaci dat z ostatních systémů do datových skladů. Obsahuje nástroje ETL (podle architektury označované i jako ELT) a EAI.

ETL nástroje

Extract, transform, load nástroje jsou blíže popsány v kapitole 4 a slouží k nahrání dat do datových struktur datového skladu. Výběr dat pro extrakci a kvalita transformace (úprava do požadované formy a vyčištění) jsou klíčovými

prvky kvality a úspěchu celého manažerského informačního systému. Zanedbání této části tak vede často ke špatným či zavádějícím údajům. Jsou spouštěny dávkově a mohou pracovat v denních, týdeních a měsíčních intervalech.

EAI nástroje

Enterprise Aplication Integration (EAI) nástroje integrují primární podnikové systémy a tím redukují celkový počet vzájemných rozhraní a na rozdíl od ETL/ELT nástrojů doručují data do datových skladů v reálném čase. Rozdíl mezi tzv. Approach (bez EAI platformy) a architektury využívající EAI je zobrazen na schématu 5.

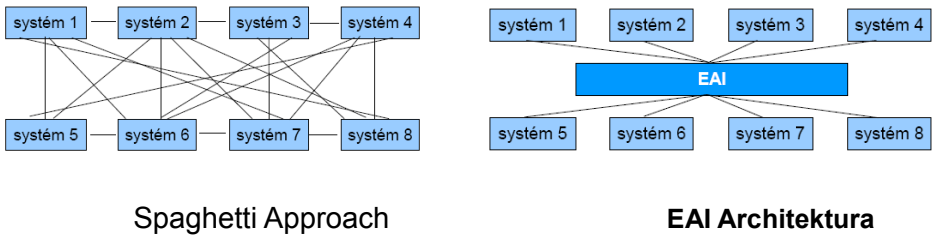


Schéma 5 – Srovnání Spaghetti Approach a EAI

Podle [4] se dělí do dvou kategorií podle úrovně, na které principiálně pracují:

- na úrovni datové integrace – integrace a distribuce dat
- na úrovni aplikační integrace – vedle integrace a distribuce dat jsou především určeny pro sdílení určitých funkcí informačních systému.

2.3.2 Datové repozitory

V této části systému (blíže popsané v kapitole **5**) se vyskytují komponenty pro práci s daty uložených v databázi určené pro MIS. Základem jsou datové sklady (DW) s vybudovanými datovými tržišti (DMA) a případně použitým dočasným úložištěm dat (DSA) anebo operativním úložištěm dat ODS. DSA ODS jsou nepovinné součásti a uplatňují se především u rozsáhlejších systémů.

Datové sklady obsahují data z produkčních systémů a to včetně historie. Existují dva základní přístupy k budování datového skladu. První podle W.H. (Billa) Inmona, kde se budují závislá datová tržiště, ke kterým pak přistupují všichni uživatelé. Druhý přístup podle R. Kimballa, kde se budují jednotlivá datová tržiště pro každý okruh uživatelů. Datová tržiště obsahují potřebná data pro jednotlivá oddělení společnosti v agregované formě.

Důležitou součástí jsou také metadata. Lze si je představit jako data o vlastních datech. Obsahují informace, kdo, jak, kdy, kde... data pořídil, ale také obsahují definice dat, informace o vzhtahu dat mezi sebou, apod.

2.3.3 Analytické komponenty

Mezi analytické komponenty se řadí OLAP (OnLine Analytical Processing), příprava sestav (Reporting) a dolování dat (DMI - Data minnig). Jsou to nástroje, se kterými pracuje přímo uživatel, proto jsou zde mimo jejich funkčnosti také důležité jejich intuitivní ovládání.

OLAP

OLAP nástroje[2] slouží k zobrazení dat z datového skladu při použití OLAP

2 Kapitola 6 OLAP nástroje

kostky. Na každou dimenzi kostky, nemusí být nutně třídimenzionální, se přiřadí množina sledovaných hodnot, jako je čas, pobočky nebo produkty. Výhodou takového použití jsou operace rozpadu hodnot (drill-down) jejich sloučení (drill-up) nebo uzamknutí jednotlivých dimenzí (Slicing a Dicing).

OLAP je oproti OLTP přímo navrhován pro provádění analýz a lze se setkat se třemi základními typy OLAP. MOLAP (Multidimenzionální OLAP) – data jsou uložena v multidimenzionálním formátu, ROLAP (Relační OLAP) – data jsou uložena v relační databázy a HOLAP (Hybridní OLAP) – část dat je uložena jako MOLAP a část jako ROLAP.

Reportovací nástroje

Reportovací nástroje slouží k přípravě a tisku sestav, kdy se můžeme setkat se standardními reporty, prováděnými v určitých časových intervalech a ad-hoc reporty, generovanými podle aktuálních potřeb managementu.

Dolování dat

Dolování dat je další analytickou komponentou sloužící k objevování dříve neznámých skutečností. Nejčastěji se používá k analýze nákupního košíku, analýza uvěrových rizik a pojistných podvodů a riziku přechodu zákazníka ke konkurenci. Převážně používanými technikami jsou techniky statistické, lze se však setkat i s generickými algoritmy nebo neuronovými sítěmi.

2.3.4 Nástroje pro uživatele

Pro koncové uživatele jsou určeny nástroje umožňující pomocí grafického rozhraní přistupovat k nejrůznějším částem manažerského informačního systému. Nejčastějším řešením je použití podnikového intranetu, zpřístupnění EIS,

monitoringu a analytických nástrojů oprávněným zaměstnancům pomocí klientských aplikací, nejčastěji s použitím internetového prohlížeče[3] a využít tak výhod tenkého klienta. Toto řešení tak umožňuje centralizovanou správu celého systému. Pro sjednocení grafického rozhraní různých komponent se používá tzv. portletů, tedy nástrojů, které se mohou integrovat do více systémů tak, že pro své grafické zobrazení (uživatelské rozhraní) použijí předem definované zobrazení serveru, na kterém se právě zobrazují (např. standard JSR-168).

Dashboard

Dashboard, do češtiny překládaný jako Manažerský pult, lze označit jako výchozí místo manažera, který zobrazuje předdefinované sestavy a přehledy aktuálních dat o chodu společnosti. Umožňuje tak sledovat průběh výroby, prodeje, dodání zakázky, spokojenost zákazníků a podobně. Lze nastavit i některá kritéria (KPI, metriky) a v případě překročení nebo přiblížení k hranici některého z nich tak varovat o této skutečnosti uživatele. Uživateli se tak zobrazují pouze důležité skutečnosti a není zahlcen velkým objemem informací, které ho mohou odvádět od podstatných probémů. Dashboard je zpravidla realizován pomocí interaktivního rozhraní s použitím technologií JSP, Flash nebo AJAX, které nabízí rychlý přístup k problémovým oblastem. Dashboard je tam prvním nástrojem, se kterým se uživatel ve většině Business Intelligence systémů setkává při každodenní práci.

Vedle výše zmíněných komponent celou architekturou prostupují také technologie pro správu metadat, datové kvality a technické znalosti. Součástí je také oborová znalost (know how), podstatná pro úspěšnou implementaci a vlastní fungování manažerského informačního systému.

3 Microsoft Internet Explorer, Mozilla Firefox, Safari apod.

2.4 Kvalita MIS

Pro určení kvality a užitnosti MIS lze použít 12 pravidel a 18 znaků formulovaných v roce 1993 podle E. F. Codd & Associate, nebo kratší hodnocení FASMI[4] (Fast Analysis of Shared Multidimensional Information) o pěti bodech.

- Fast: Rychlost dpovědi systému na anaytický dotaz má být do 5 sekund, u jednodušších dotazů pak do jedné sekundy. V případě složitých dotazů do 20 sekund. V případě překročení 30 sekund lze očekávat, že uživatel toto vyhodnotí jako chybu systému.
- Analysis: Manažerský systém by měl nabízet možnost snadné tvorby analýz a reportů na základě potřeb manažerů bez nutnosti doprogramování či nastavení dané oblasti. Také by neměl vyžadovat znalosti statistických či metod k nim vedoucích. Výstupy těchto analýz musí být snadno pochopitelné.
- Shared: Systém má nabízet sdílení dat ve společnosti. Každý, kdo má v systému náležité oprávnění, má mít možnost sdílet s ostatními ve společnosti své analýzy.
- Multidimensional: Manažerské systémy musí nabízet pohled na data z více úhlů pohledů (dimenzí). Multidimenzionalita je základem všech OLAP nástrojů.
- Information: Výstupem manažerských systémů jsou kvalitní, relevantní a správné informace.

4 http://www.olapreport.com/fasmi.htm

3 Open source

Open source, do češtiny překládáný také jako svobodný software, je softwarový produkt (dílo) dodávaný včetně zdrojového kódu. Svoboda v tomto případě představuje volnost v modifikaci, distribuci a užívání. Přiložený zdrojový kód a překlad svobodný software nebo označení zdarma může zavádět k tomu, že lze s takovým produktem libovolně nakládat – což samozřejmě není pravda.
Ke takovému produktu existují autorká práva a právě na autorovi záleží, pod jakou licencí chce svůj produkt distribuovat.

Open source tak nabízí uživateli hlavně volnost v modifikaci programu, ale klade důraz na to, aby tato modifikace byla zdokumentovaná a modifikovaný zdrojový kód byl opět přístupný. Popis modifikace je důležitý k určení toho, co, kdo, kde, jak modifikoval, aby bylo jasné autorství a existovala možnost odhalení případné chyby produktu, která by mohla být neprávem připisovaná původním autorům. K takovýmto programům zpravidla nebývá žádná záruka. Licence open source produktů se vyskytují v několika vlastních verzích. Rozdíly spočívají zpravidla v reakci na nově se vyskytnutý problém. Například GPL v.2 z roku 1991 je až v současnosti nahrazována třetí verzí. Zde se vyskytuje několik licencí:

GNU GPL

General Public License (dále jen GPL) je nejčastější licencí. Umožňuje program modifikovat, volně spouštět a případně i distribuovat za poplatek. Modifikace musí být vždy volně dostupná a plně zdokumentovaná k určení autora.

GNU LGPL

Lesser General Public License (dále jen GPL) je licence vhodná především pro softwarové knihovny funkcí. Produkt opatřený touto licencí lze volně zařadit do vlastního produktu a dále za určitých podmínek distribuovat jako vlastní komerční produkt. Změna licence z GPL na LGPL je zpravidla zpoplatněna, autoři tak získávají z tohoto převodu finanční prostředky a mohou tak generovat z těchto produktů také zisk.

BSD

Berkeley Software Distribution je nejvolnější licence v open source. Produkt lze zařadit do komerčního produktu a to i bez nutnosti zveřejnit zdrojový kód. Musí pouze obsahovat zmínku o autorech a zřeknutí se odpovědnosti. Lze také měnit licenci na GPL – opačně to samozřejmě nejde.

MPL

Mozilla Public License je volnější než GPL, lze ji přiřadit ke komerčnímu produktu. Pro splnění podmínek je nutné zařazenou část dalé šířit pod MPL.

Lze se setkat také s licencemi vlastními, zpravidla vycházejí GPL a upravují jen části dle potřeby autorů či charakteru produktu. Licence musí být vždy distribuovaná společně s produktem. Licence také ošetřují používání patentů, které jsou v rosporu s filozofií open source. Na software či jeho části tak nelze získat patent a tak de facto zamezit dalšímu šíření, používaní nebo modifikaci produktu.

3.1 Typy licencí pro software

U softwarových produktů se lze setkat s několika dalšími typy licencí. Jsou to především:

- Public domain: Software, na který se nevztahují žádná autorská práva a s ním nakládat dle vlastního uvážení. Doba pro uplynutí majetkových práv je 70 let od úmrtí autora.
- Copyleft software: Označuje produkty svobodného softwaru, nemající žádné omezení k užívání nebo k distribuci. Nedovoluje však přidávat další omezení při modifikaci či distribuci.
- Freeware: oproti shareware ho lze využívat po neomezenou dobu a v plné funkčnosti. Šířit tento produkt je opět možné volně, ale bez úplaty. Příkladem může být Sun Microsystems Java nebo Microsoft SQL Server Express.
- Shareware: Jedná se o produkty s plnou či omezenou funkčností, který lze volně bezplatně šířit. Jedná se o běžnou propagaci produktů, plnou verzi lze získat za licenční poplatek. Do té doby produkt pracuje s omezením některých funkcí nebo po určitý časový úsek.
- Komerční (proprietání) software: Běžně distribuovaný komerční software bez zdrojového kódu a bez možnosti modifikace. Jeho změny nebo šíření jsou bez povolení přímo zakázány. Jako příklad zde lze uvést Microsoft Windows.

3.2 Open source BI

Manažerské informační systémy se staly podstatnou součástí podnikových systémů. Tuto skutečnost si uvědomují přední výrobci podnikového software, proto

lze najít nabídku těchto systémů u každého z nich. Jedná se o společnosti vyvýjející databázové systémy, nabízející podnikové systémy nebo nabízející ekonomický software. Pojetí MIS se tak společnost od společnosti liší v nabídce komponent nebo podporovaným zaměřením. Především se však jedná o komponenty spadající i do Business Intelligence. Prim na trhu komerčních BI hrají společnosti Microsoft, Hyperion, Cognos, MicroStrategy, SAP nebo Oracle.

Výrobci open source řešení BI (označované také jako OSBI) jako je Pentaho (Pentaho BI Suite), JasperSoft (JasperSoft BI Suite), Engineering Ingegneria Informatica (SpagoBI), Insight Strategy (The Bee Project) nebo Loyalty Matrix (OpenI) také reagují na tento trend, avšak s odlišným ekonomickým modelem. Své produkty nabízejí jako open source, bez jakékoli záruky a zákaznické podpory. Jedná se většinou o soubory ETL nástrojů, databázových strojů, OLAP, reportovacích a dataminingových nástrojů nabízených jinými výrobci. Nabízejí zdrojové kódy, návody instalace, až na vyjímky i kvalitní základní podporu celku a pro všechny jednotlivé komponenty. Pomocí těchto produktů lze realizovat projekty menší a střední organizace, při realizaci velkých projektů je pak pořizovací cena MIS případně BI zanedbatelná.

Výrobci OSBI i jednotlivých komponent přistupují s různými modely financování. Vývoj těchto produktů zpravidla probíhá nejčastěji pěti způsoby:

- Vedle vlastního komerčního produktu společnosti nabízejí také open source verzi (např. Apatar, MySQL), komerční verze je pak buď obsáhlejší anebo je doplněna lepší podporou formou telefonu, emailu, školení apod.
- Společnost nabízí k vlastnímu open source produktu pouze placenou

podporu, záruku nebo školení (Talend).

- Vývoj je financován dalšími společnostmi (Apache Software Foundation)
- Produkt je vyvíjen jednotlivci popřípadě skupinkami vývojářů. Takové produkty však většinou nejsou nikterak kvalitní, jsou však mnohdy zdrojem nových řešení, které se později přesouvají do jedné z předchozích forem finacování.
- Projekty vznikají na akademické půdě (WEKA).

U společností se tak zpravidla jedná o jistou marketingovou strategii, kdy se snaží získat zákazníka, který chce na počátku ušetřit investice do systému. Ve chvíli, kdy zákazníkovi stávající systém pomalu přestává stačit z důvodu nedostatečného výkonu, chybějící funkce čí nutnosti větší podpory výrobcem, s velkou pravděpodobností se může obrátit na dodavatele s požadavkem na již komerční produkt. Společnosti sponzorující open source (např. Google, IBM nebo HP) tak mohou naproti tomu získávat nové technologie, postupy a případně i nové talenty. Sponzorují tak nejčastěji technologie „z oboru“ nebo technologie, které s vlastními produkty přímo spolupracují.

Dokumentace, podpora a funkčnost jsou kritéria sledovaná v představení a hodnocení MIS produktů v následujících kapitolách. Open source manažerské informační systémy a jejich jednotlivé komponenty splňují tato kritéria velmi rozmanitě. Nejlepší produkty, jako například produkty Pentaho, JasperSoft, MySQL nebo Talend jsou již vyspělé a mohou tak konkurovat komerčním produktům, což dokazuje například MySQL s miliony instalací ve světě. Již

samozřejmě neplatí, že open source produkty nejsou vhodné pro nasazení do řízení v podniku, avšak stále nejsou na úrovni produktů předních výrobců komerčně nabízených systémů. Tyto rozdíly bývají vidět již při zkoumání uživatelského prostředí, které však díky otevřenosti lze modifikovat a v současnosti tento nedostatek i dohánějí. Podstatnějším nedostatkem je zde absence některých funkcí, čímž výrobci komerčních produktů stále drží výhodu.

Na druhé straně stojí produkty, které mají se splněním těchto kritérií problémy. Jedná se zpravidla o nedostatečnou dokumentaci, absenci alespoň základní podpory či velmi omezenou funkčnost. Použitím těchto komponent v podnikovém potažmo komerčním prostředí je velmi rizikové a nelze ho doporučit.

Programovacím jazykem, ve kterém se nejčastěji programují open source MIS popř. BI, je Java. Výhodou použití tohoto jazyka je možnost nasazení na více operačních systémech popřípadě platformách díky použití frameworku Java Runtime Environment (JRE). Nejčastěji se jako operační systém používá Linux, Unix, Windows případně i MacOS. MIS/BI založené na platfrmě JRE lze provozovat i na dalších systémech jako HP-UX nebo Sun Solaris, ale náklady na pořízení těchto platforem přenášejí nasazení spíše do segmentu větších společností, kde výhody open source nejsou tak patrné, proto je podpora těchto operačních systémů zpravidla zahrunuta do komerčních verzí postavených na OSBI.

4 Datová transformace

Datová transformace (ETL – Extract, Transform, Load, popř. ELT – Extract, Load, Transform) slouží pro transformaci dat z transakčních systémů společnosti do datových skladů (označované také jako datové repository). Datová extrakce by měla být zajištěna samostatným produkčním systémem, protože tím lze omezit zatížení transakčních systémů v době zátěže a extrahovat až v době, kdy zatížení nebude tak vysoké. Datová transformace se skládá ze tří částí: extrakce, transformace a importu.

4.1 Extrakce (Extract)

Slouží k získávání potřebných dat pro datové sklady. Jako zdroj dat mohou sloužit transakční systémy, ERP, SCM, CRM (Microsoft Dynamix, SAP, Oracle), relační nebo jiné typy databází (nejčastěji přístup pomocí ODBC, JDBC), flat-file soubory, soubory XLS, emaily a korespondence obecně, ale také logovací soubory internetových stránek (například pro zjištění obliby zákazníků určitých produktů či služeb) apod. Extrakce mění data do formátu pro proces transformace. Výběr vhodných zdrojů pro extrakci je klíčové v úspěšném řešení BI, proto je kvalita extrakce velmi důležitá pro kvalitu výsledných dat.

4.2 Transformace (Transform)

Po extrakci následuje transformace dat do požadované podoby, neboť takto získaná data zpravidla obsahují hned několik problémů, které je nutné vyřešit, jinak

získaná data nebudou mít plnou vypovídající hodnotu. Podle [5] a [9] to jsou:

- Vybírat pouze sloupce mající data (nebo nevybírat ty, které mají prázdnou nebo null hodnotu)
- Přeložit data do společného formátu (pokud například obsahují 1 pro muže a 2 pro ženu a datový sklad používá pro určení pohlaví hodnoty M a Z, pak je přeložit jako M popř. Z apod.) - očištění dat
- Sjednotit různá označení pro stejnou věc (1, Muž, pan uložit jako M)
- Odstranit duplicity dat.
- Rozlišit stejné označení pro různé věci (IS – injektážní souprava, instalatérská sada)
- Dodat spočítané údaje
(např. vynos = počet * (prodejní_cena – náklad))
- Sčítání víceřádkových dat (např. celková prodejnost pro každou pobočku a každý region)
- Generovat náhradních identifikátory položek
- Transponovat nebo otáčet (např. změnit sloupce na řádky nebo obráceně)
- Rozdělit sloupec do více sloupců (např. vložit řetězec do jednoho sloupce a hodnoty do dalších)

4.3 Import (Load)

Tato část nahrává extrahovaná případně i transformovaná data do datového skladu. Může probíhat v týdeních cyklech ale také každou hodinu. Data v datovém skladu mohou být jednak nahrazena nebo mohou být přidávána k již existujícím.

Nastavení importu je pak závislé od potřeb analytických nástrojů.

4.4 Historie ETL/ELT

Krátký přehled historie ETL nástrojů podle [1]:

První generace (generátory kódu) ETL byla založena na vytváření kódu, nejčastěji v jazyce COBOL, přímo pro operační systém popř. platformu, na které procesy data integration běžely. Takové nástroje data integration byly na vývoj jednodušší, a využívaly centralizovaného nástroje pro generování procesů data integration. Výkon těchto nástrojů byl velmi dobrý, ale byla zapotřebí rozsáhlá znalost platformy, na které tyto procesy běží. V té době byla tato architektura výkonná, protože data byla uložena ve flat-file souborech nebo hiearchistických databázích. Dobře tak procesy běžely na mainframech, méně úspěšné byly pro práci s relačními databázemi s velkým obsahem dat.

Druhou generací (vlastní ETL nástroje) byly nástroje, které byly umístěny mezi datovými zdoji a datovými sklady. Výhoda spočívala v tom, že bylo zapotřebí pouze vlastní jazyk ETL nástroje a nebylo tak nutné užívat více programovacích jazyků pro různé platformy. Tyto nástroje měly však nevýhodu ve formě snížení výkonu při zpracovávání všech transformací a stávaly se tak uzkým hrdlem zpracování dat. Všechna data, která přicházela do ETL nástrojů z různých zdrojů byla zpracovávána po řádcích (row-by-row), což je velmi pomalé při velkém objemu dat.

Třetí generace ETL (E-LT architektura) Vychází z výhod a řeší nevýhody předchozích dvou generací. Od druhé generace výrobci zlepšují vlastnosti vlastních SQL jazyků, což mělo za následek generování a vykonávání vysoce

optimalizovaných procesu data integrity v nativnim SQL nebo v dalších jazycích těchto procesech. Třetí generace tak nabízí grafické prostředí s možností generovat nativní SQL pro spouštění transformací na serveru datového skladu. Není tak nutné umisťovat ETL server mezi datové zdroje a datový sklad, použitím RDBMS pro spouštění transformace umožňuje hromadné zpracování dat, které je 1000 krát rychlejší než při zpracování row-by-row.

4.5 Rozdíly mezi ETL a ELT

Při použití postupu ETL se během transformace se podle [2] v současnosti data převádí na transformační server, kde probíhá ukládání dat do dočasných struktur pro konečné přepracování (někdy se označuje jako „staging“, příslušné adresáře popř. databáze jako „Stage“).Tento transformační server obsahuje i SW prostředky pro transformaci a není součástí datových repositorů.

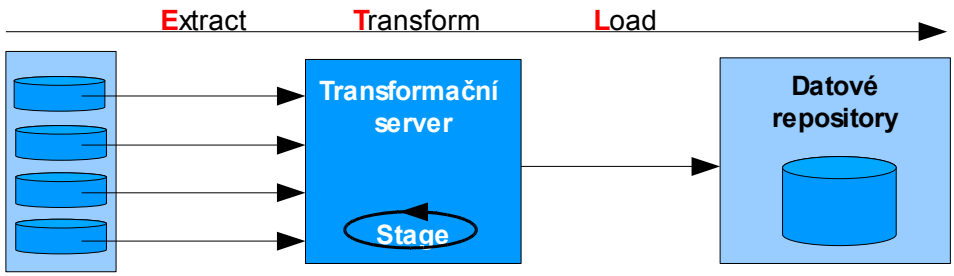


Schéma 6 – ETL architektura, převzato [2]

Mezi výhody tohoto řešení pak patří uložení dat mimo datové repozitory a tím zpřístupnění i dalším uživatelům. Další výhodou je lepší implementace některých složitějších algoritmů. Nevýhodou však je náročnost datové transformace a tedy nutnost přiřadit transformačnímu serveru dostatečný výkon, který se už nikde jinde

nevyužije. V případě, že extrahovaná data v data repository budeme dále transformovat, stane se tak celý proces implementačně složitější.

Naproti tomu ELT využívá výkon cílového data repository, protože extrahovaná data jsou uložena přímo v data repository a „Stage“ se stává jedou z databází data repository, čímž snižuje nároky na transformační server. Nevýhodou tohoto řešení je nepřístupnost dat během transformace, protože vyžaduje veškerý výkon data repozitory a nutnost integrace s databázovým systémem data repozitory.

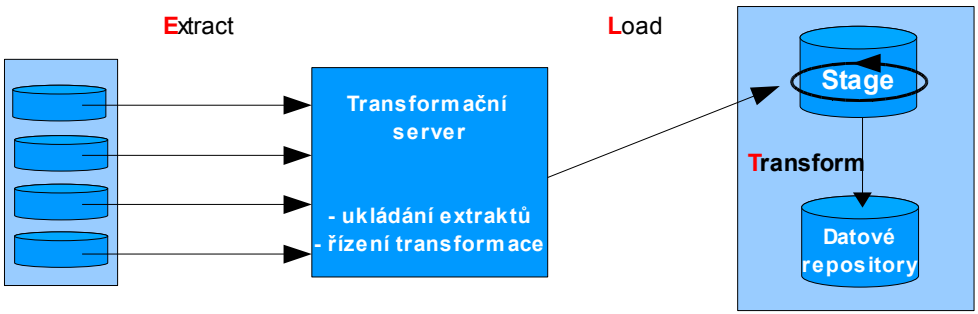


Schéma 7 – ELT architektura, převzato [2]

Volba architektury datové transformace tak závisí na potřebách jednotlivých společností. Každý z těchto postupů má svá pro a proti. Pro výběr je nutné zvážit kopatibilita ETL/ELT nástroje s cílovou databází, možnost monitoringu, správy provozu transformačních procesů a správu metadat.

4.6 Paralelní zpracování (Parallel Processing)

V současnosti některé ETL/ELT nástroje používají paralelní zpracování, což dovoluje zvýšení výkonu celého transformačního procesu. Existují tři typy paralelismu [5] :

- Data – rozdělení sekvenčního souboru do více datových souborů pro paralelní přístup.
- Pipeline – dovolují současně spouštět několik komponent na stejném datovém toku (např. sledujeme hodnoty v prvním záznamu a ve stejný čas přidáváme dohromady dvě pole do záznamu 2).
- Komponet – současný běh více procesů na rozdílných datových proudech. Můžeme tak například třídit jeden vstupní soubor, zatím co odstraňujeme duplicitní hodnoty.

4.7 Čistění dat (Data Cleansing)

Při čištění dat dochází k nalezí a opravě nebo odstranění chybných nebo neúplných dat. Po očištění jsou data konzistentní s ostatními záznamy. Tento proces garantuje, že data jsou jednoznačná, správná a úplná. K čištění dat se používají opravy překlepů nebo opravy podle seznamu známých chyb. Nekonzistence může být způsobena uživatelovým chybným zadáním, ale také přesunem nebo uložením dat.

4.8 Open source ETL/ELT nástroje

V této části jsou popsány některé z open source ETL/ELT nástrojů. Kvalita, dokumentace a podpora těchto nástrojů je podle toho velmi rozličná. Některé nabízejí velmi kvalitní dokumentaci ať už formou názorných videosekvencí

zobrazujících například nastavení, kompletních administrátorských, vývojářských nebo uživatelských příruček, podpory formou fóra a FAQ (zpravidla u open source projektů vyvýjených společnostmi). Jiné pak nenabízí ani popis základních funkcí, dostupný je pouze zdrojový kód.

První ETL/ELT nástroje byly pouze generátory skriptu a v některých tato architektura přetrvává dodnes. Jejich výhodou je jednodušší a rychlejší implementace, na druhou stranu jsou však méně vhodné pro rozsáhlejší nebo velké systémy z důvodu složité administrace. V těchto případech se více uplatňují frameworkové nebo serverové řešení, která mají však nevýhodu ve složitější implementaci do BI systému jiných společností.

Hlavními programovacími jazyky open source ETL/ELT nástrojů jsou Java a PERL, v případě generátorů se lze setkat s oběma z těchto jazyků. Java je více používaná i přes to, že programovací jazyk PERL disponuje větším množstvím propojení s databázemi.

Většina níže popsaných ETL/ELT nástrojů používá grafické prostředí pro modelování úloh, pro navržení transformací (Transformation Designer, Transformation Mapper) a je většinou řešeno buď jako plug-in pro Eclipse nebo vlastním grafickým prostředím.

4.8.1 Apatar

Verze:	**1.0**
Typ:	**Server**
Programovací jazyk:	**Java**
Operační systém:	**UNIX/Linux, Windows**
Připojení:	**JDBC, ODBC, XML**
Licence:	**GPL**
Výrobce:	**Apatar, Inc.**
Odkazy:	**http://www.apatar.com**

Společnost Apatar nabízí ETL servery ve dvou edicích: zdarma dostupná Apatar Community Edition a zpoplatněná Apatar Enterprise Edition. Obě nabízí grafický Job Designer, Mapping and Transformation Designer, Business Modeler, Metadata Configuration Wizards, Job Scheduling, datové a aplikační propojení, podporu Flat File, reportování transformací, logování chyb, instalační a administrátorskou dokumentaci a komunitní podporu. Placená verze pak navíc obsahuje obsáhlejší podporu, školení a podporu API rozhraní dalších výrobců. Dokumentace je kvalitní, nechybí zde ani video ukázky instalace nebo nastavení.

4.8.2 CloverETL

Verze:	**2.3.0**
Typ:	**Framework**
Programovací jazyk:	**Java**
Operační systém:	**UNIX/Linux, Windows**
Připojení:	**XML, JDBC**
Licence:	**LGPL**
Výrobce:	**OpenSys, a.s.**
Odkazy:	**www.cloveretl.org** **www.opensys.eu**

CloverETL pražské společnosti OpenSys a.s. je na operačním systému nezávislý framework a díky programovacímu jazyku Java a je dle výrobce schopen běhu na následujících operačních systémech: Linux, AIX, Solaris, HP-UX a Windows. Pro svá data nabízí podporu Unicode, což umožňuje použít jakoukoli znakovou sadu a dále konvertovat například do ASCII, UTF-8, ISO-8859-1 nebo ISO-8859-2. Podporuje jakékoliv databáze s JDBC, dále pak také dBase, FoxPro nebo FlashFilter.

Transformace dat je vykonávána pomocí nezávislých součástí, každá využívá vlastní vlákno a lze vyžít více procesorů. Framework implementuje

pipeline-parallelism – zpracovaná položka je ihned poslaná následující komponentě k dalšímu zpracování. Nástavbou je placené pro komerční užití Clover.GUI, které je řešeno jako plugin do Eclipse, kde lze nastavit transformační schéma v grafickém prostředí.

Dokumentace a podpora je kvalitní, nabízí jednak instalační, uživatelské a příručky, FAQ a také online fórum.

4.8.3 Enhydra Octopus

Verze:	**3.6-3**
Typ:	**Script generator**
Programovací jazyk:	**Java**
Operační systém:	**UNIX/Linux, Windows**
Připojení:	**JDBC, XML**
Licence:	**GPL, LGPL**
Výrobce:	**Enhydra.org**
Odkazy:	**http://forge.objectweb.org/projects/octopus** **http://www.enhydra.org**

Enhydra Octopus je vytvořen v jazyce Java, k datovým souborům se připojuje pomocí JDBC. Nastavení transformace je řešeno pomocí konfiguračního souboru XML, který se vytváří v grafickém nástroji Octopus Generator. Octopus Loader pak provádí transformaci mezi různými typy databází (MS SQL, MySQL, Oracle, DB2, PostgreSQL, Sybase, Paradox a JDBC-ODBC s Excelem a Accessem). Dále jsou podporovány XML (XML-JDBC) a CSV (CSV-JDBC) soubory. Chybí zde propojení s flat-file soubory a dalšímy databázemi bez podpory JDBC. Součástí je velmi kvalitní a rozsáhlá dokumentace, která je však dostupná až po stažení celého balíku.

4.8.4 KETL

Verze:	**2.1.12**
Typ:	**Server**
Programovací jazyk:	**Java**
Operační systém:	**UNIX/Linux, Windows (testování)**
Připojení:	**XML, JDBC, SOAP**
Licence:	**GPL, LGPL**
Výrobce:	**Kinetic Networks, Inc**
Odkazy:	**www.ketl.org** **www.kineticnetworks.com**

KETL (Kinetic Network Extract Transform and Load) je ETL server vyvýjený v jazyce Java (zapotřebí Java Virtual Machine verze 1.5 nebo vyšší), což umožňuje jeho nasazení na libovolné platformě. Podporována je prozatím instalace pouze pro platformu UNIX/Linux, instalace poslední verze tohoto ETL/ELT nástroje pro Windows se v současnosti testuje. Běh serveru je možný také v systémech s více procesory a na 64-bit serverech. Rozšířená správa běhu úloh pak nabízí plánování na základě událostí nebo pomocí času. Úlohy jsou přístupné přes KETL API a dělí se do tří hlavních částí: SQL s předdefinovanými SQL příkazy přes JDBC, XML s definicemi úloh (v XML) a Operačního systému s příkazy OS.

KETL úlohy jsou definován pomocí XML, což umožňuje snadné vytváření nových úloh, modifikaci a správu verzí. Další výhodou je také sledování výkonu, které sbírá údaje o dřívějších a současných úlohách pro celkovou analýzu těch problemových. K uložení metadat lze využít jakoukoli relační databázi s funkcí row-level locking, testovány jsou PostgreSQL (8.1 a vyšší), Oracle (9.1 a vyšší) a MySQL (5.1 a vyšší).

Součástí je také kvalitní dokumentace a to: Instalační průvodce, Průvodce administrátora – vše v angličtině. V dokumentaci jsou odkazy také na Příručku programátora, která však spolu s funkčním FAQ v době vzniku této práce není k .

4.8.5 Pentaho Data Integration (KETTLE)

Verze:	**2.5.1**
Typ:	**Framework**
Programovací jazyk:	**JAVA**
Operační systém:	**UNIX/Linux, Windows, MacOS**
Připojení:	**XML, JDBC-ODBC**
Licence:	**LGPL**
Výrobce:	**Pentaho Corp.**
Odkazy:	**http://www.pentaho.org**

Pehtaho Data Integration (dříve KETTLE: Kettle Extraction, Transport, Transformation and Loading Environment) je založen na platformě Java (JRE 1.4 nebo vyšší). Tento framework umožňuje spojení s relačními databázemi pomocí JDBC, podporovány jsou také databáze formou XML, XLS souborů a flat-file soubory. Architektura tohoto ETL nástroje se skládá z pěti komponent:

- Spoon: grafický nástroj pro modelování datového toku ze zdrojové databáze, transformace dat a výstupu této transformace do databáze.
- Pan: nástroj pro provádění transformace modelované ve Spoon a to pomocí příkazové řádky a nabízí možnost spouštění i v časových intervalech.
- Chef: grafický nástroj pro modelování úloh, jako například transformace nebo FTP download.
- Kitchen: nástroj pro spouštění modelovaných úloh v nástroji Chef pomocí příkazové řádky.
- Carte: jednoduchý web server, který umožňuje vzdálené spouštění transformací. Je založenen na přijímání XML (pomocí servletu), který obsahuje transformace ke spuštění a nastavení. Dále také nabízí vzdálenou správu, spouštění a ukončení transformací běžících na Carte serveru.

Tento nástroj je součástí Pentaho Open BI Suite. Díky své uživatelské přívětivosti konfiguračních nástrojů Spoon a Chef lze tento produkt bez dalších úprav použít přímo koncovými uživateli. Dokumentace je velmi obsáhlá a doplněna podporou formou FAQ a fóra. Za poplatek lze získat i profesionální podporu. Tento ETL nástroj se tak řadí mezi nejlepší v tomto srovnání.

4.8.6 Talend a JasperETL

Verze:	**2.0**
Typ:	**Script generator**
Programovací jazyk:	**JAVA, PERL**
Operační systém:	**UNIX/Linux, Windows**
Připojení:	**XML, ODBC**
Licence:	**GPL**
Výrobce:	**JasperSoft, Talend**
Odkazy:	**http://www.jasperforge.org** **http://www.jaspersoft.com** **http://www.talend.com**

JasperETL je vyvíjen na technologii jiného open source ETL nástroje Talend (nabízí také ELT), který byl vůbec prvním dostupným open source ETL/ELT produktem. JasperETL je součástí JasperSoft Open Source Business Intelligence Suite, Talend je pak užit například v SpagoBI. Nejedná se o klasický framework (jako např. KETTLE), ale o generátor Java/PERL skriptů. Je založen na Eclipse RCP a skládá se z následujících komponent:

- Job Designer: nabízí grafický editor a funkční zobrazení ETL procesů.
- Transformation Mapper: poskytuje náhledy a možnost editace celkového schématu transformace.
- Business Modeler: umožňuje grafické náhledy na toky obchodních informací.

Dále také nabízí Real-time ladění, který umožňuje sledování statistik

a krokování během celého transformačního procesu v reálném čase. Tento ETL nástroj umožňuje připojení k mnoha datovým zdrojům jednak pomocí ODBC, ale také přístup k XML souborům a dalším. Konfigrační průvodci metadat (Wizards) napomáhájí konfiguraci různorodých datových zdrojů a komplexních datových souborů včetně CSV a XML.

Celý ETL nástroj je doplněn o kvalitní dokumentaci (rozsáhlá Uživatelské příručka), FAQ a fórum. Existuje také placená verze Profesional, která obsahuje rozšířenou podporu operačních systémů, web/aplikačních serverů, databází, profesionální vývojářskou a uživatelskou podporu.

4.9 Hodnocení

Kritérií pro hodnocení reportovacích nástrojů jsou:

- Dokumentace a podpora (Dokumentace), především pak její kvalita o obsáhlost, 30%.
- Propojení s databázemi a dalšími datovými soubory (Propojení), kde je sledována možnost vedle standartního JDBC případně ODBC je sledována možnost použití i jiných datových zdrojů (XML, XLS), 30%.
- Nastavení konfiguračního souboru (Nastavení) pro transformaci, kde je sledováno užití grafického rozhraní, či pouhá editace konfiguračního souboru na textové úrovni, 10%.
- Plánování spouštění transformace (Plánování), kde kde je kritériem ovladatelnost v grafickém rozhraní či pouhá tvorba skriptů, 10%.
- Operační systém (Operační systém), kde je sledován počet podporovaných operačních systémů, 20%.

Podle těchto kritérií se tak nejlepším posuzovaným ETL nástrojem stal KETTLE, který všechna kritéria splnil. Talend pak ztrácí pouze v podporovaných operačních systémech, kde je pouze absence podpory operčního systému MacOS. Třetí místo zastává KETL, který je podporován pouze pro Linux a má horší grafické rozhraní oproti KETTLE a Talend.

5 Datové repository

Datové repository slouží k uložení dat po ETL procesu a jsou zde uložena k dalšímu zpracování pomocí OLAP nástrojů nebo k přímému čtení uživately. Součástí datových repositor mohou být dočasná úložiště dat (DSA) a operativní úložište (ODS). Datové sklady mohou být řešeny několika způsoby v závislosti na potřebách a struktuře společnosti, ale také na investicích do vývoje datového skladu:

- Formou samostatných datových tržišť (R. Kimball)
- Konsolidovaným datovým skladem (W.H. Inmon)

Setkáváme se tu tedy se dvěma základními pohledy na datové sklady. První, formulovaný R. Kimballem, který představuje datový sklad jako sjednocení několika dílčích skladů tzv. datových tržišť (Data Mart), jejiž tématicky orientovaná databáze slouží potřebám pouze jedné skupiny uživatelů. Jedná se tak o rychlé řešení konkrétních požadavků, avšak za cenu redundance dat a rozdílných výsledků analýzy dat jednotlivými skupinami. Zobrazeno na schématu 8.

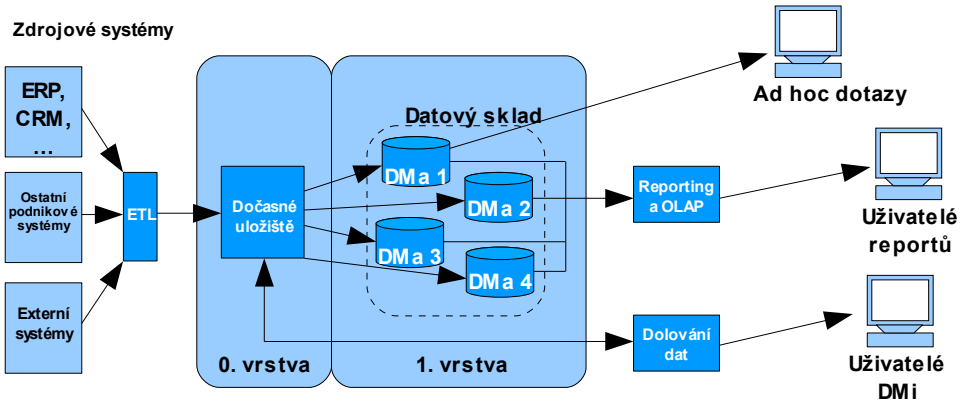


Schéma 8 – Postupné budování datových tržišť (Kimball), převzato z [4]

Druhý pohled W. H. (Billa) Inmona na tuto problematiku vede k jednotnému datovému skladu se závislými datamarty, ke kterým pak přistupují všechny skupiny. Jedná se o třívrstvou architekturu, neboť vedle datového skladu a datových tržišť se zde může objevit i dočasné datové uložiště. Takový datový sklad je subjektově orientovaný, integrovaný, stálý a časově rozlišený. Toto řešení však vyžaduje v začátku důkladnou analýzu potřeb jednotlivých oddělení. Zobrazeno na schématu 9.

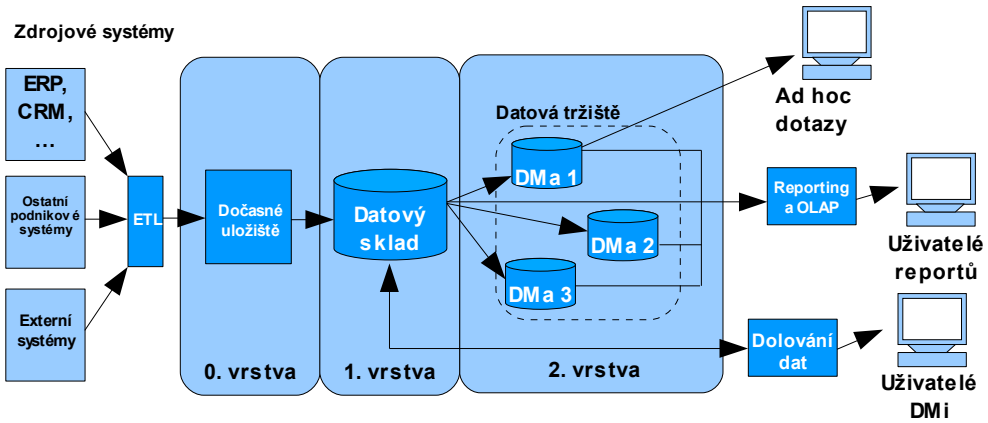


Schéma 9 – konsolidovaný datový sklad (Inmon), převzato z [4]

5.1 Datové tržiště (DMA, Data Mart)

Datová tržiště jsou de facto jednotlivé datové sklady, které jsou určené jednotlivým skupinám uživatelů (popř. oddělením , úsekům) ve společnosti. Pro každou skupinu tak existuje jedno datové tržiště. Každé takové tržiště obsahuje data potřebná pro jednotlivé skupiny uživatelů v agregované formě.

V případě budování datového skladu formou samostatných datových tržišť (dvouvrstvá architektura) mohou jednotlivá datová tržiště mít vlastní ETL, OLAP, Data Mining a další nástroje. Nevýhodou této architektury pak je redundance dat všech tržišť, ale hlavně potenciální nejednotnost údajů poskytovaných jednotlivými datovými tržišti, čímž dochází k rozdílným výsledkům stejné analýzy dat prováděné různými skupinami. Na druhou stranu se datová tržiště lépe implementují, v případě potřeby lze jednoduše přidat další tržiště bez složitého zásahu do celého systému. Lze také lépe sledovat návratnost investic do budování DMA.

Budování samostatných datových tržišť bylo navrženo v osmdesátých letech minulého století R. Kimballem. Tento koncept byl pak na základě potřeb práce s velkými objemy dat přepracován v devadesátých letech a vzniká tak sběrnicová architektura (Bus Architekture), která se snaží o budování jednotlivých datových tržišť integrovaně pomocí sdílených dimenzí (dimenzionální tabulky), které jsou užity v dalších (později budovaných) datových tržištích.

5.2 Konsolidovaný datový sklad

Konsolidovaný (EDW, Enterprise Data Warehouse) datový sklad je tak oproti postupu s použitím pouze DMA navrhován jako jednotná struktura pro všechny skupiny. Návrh potřeb všech oddělení se tak řeší hned na začátku, což má zvýšené

nároky na vývoj tohoto řešení. Výhodou však je, že údaje v EDW jsou stejné pro všechna oddělení celého podniku. Získaná data jsou tak jednotná a nedochází k redundanci údajů. Data v tomto skladu jsou detailní, tj. na úrovni jednotlivých transakcí. Nad tímto skladem jsou zřízeny jednotlivá datová tržiště, která však operují nad společnými daty konsolidovaného datového skladu a výstupy jsou tak stejné pro různé skupiny.

Data jsou v konsolidovaném datovém skladu uložena v relační databázi. Vzhledem ke složitosti čtení ERD se objevují metody pro jeho zjednodušení, přizpůsobení koncovými uživateli. Užívá se k tomu Dimenzionální model, kde jsou fakta a dimenze. Fakta jsou údaje, které jsou měřitelné a měnící se v čase. Dimenze jsou pak konstanty. Další výhodou dimensionálního modelování je dobrá srozumitelnost koncovým uživatelům nebo optimalizace pro složité analýzy a vyhledávání dat. Používá se zde denormalizace a redundance za použití schématu hvězdy (Star Scheme). Schéma sněhové vločky (Snowflake scheme) je oproti tomu v 3. normální formě. Srovnání je zobrazeno na schématu 10. Rozdíl mezi datovými modely OLTP a datovými sklady je vyobrazen na schématu 11. Zvětšený model včetně vzorových dat je v příloze 2.1 – OLTP.

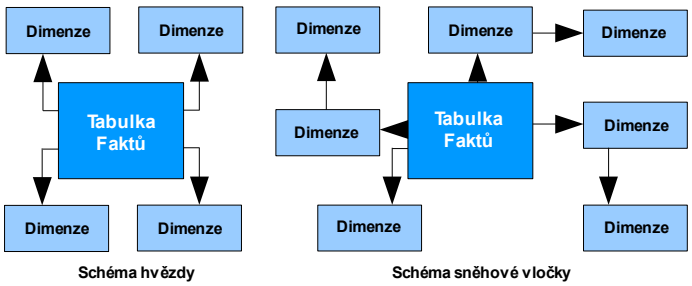
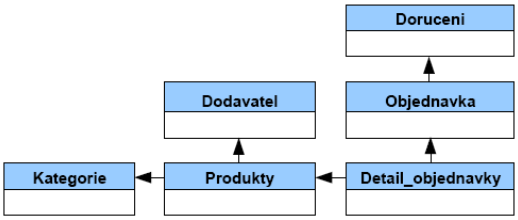
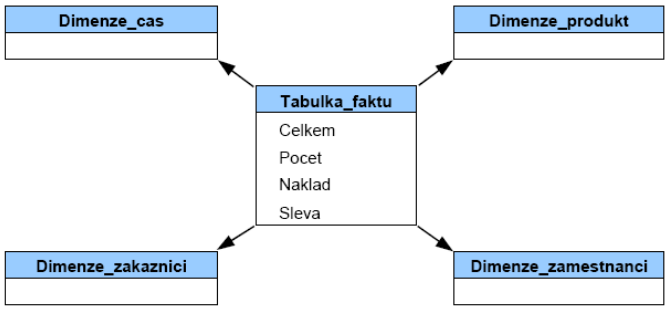


Schéma 10 –Srovnání schématu hvězdy a sněhové vločky



Datový model OLTP systémů



Datový model OLAP systémů

Schéma 11 – Srovnání datových modelů OLTP a OLAP

5.3 Přírůstkový konsolidovaný sklad

Přírůstkový konsolidovaný sklad vychází z konsolidovaného skladu s tím rozdílem, že jednotlivé části jsou budovány postupně, ale je však nutný detailní návrh už na začátku řešení. Tento přístup je nejmladší a stále více oblíbený. Zobrazen na schématu 12.

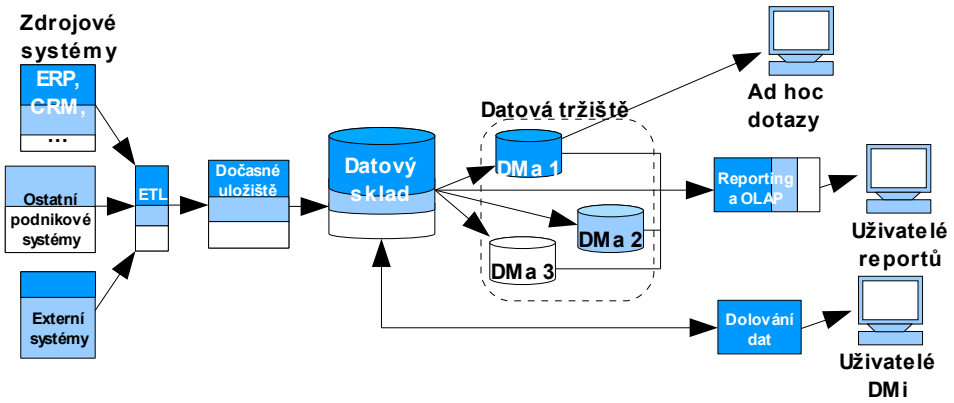


schéma 12 – Přírůstkový konsolidovaný sklad, převzato z [4]

5.4 Dočasné uložiště dat

Hlavním úkolem dočasného uložiště dat (DSA, Data Staging Areas) je urychlit proces vlastní extrakce dat. Je tak přímo spojen s ETL nástroji, slouží k prvnímu uložení dat po extrakci. Pro celkové řešení BI není DSA povinná, používá se pokud je potřeba provádět transformace dat bez dopadu na výkonnost produkčních systémů. Uplatnění DSA je také v případech, kdy je potřeba před zpracováním konvertovat data do jiného formátu (např. z textových souborů do relačních databází). Data v DSA jsou podle [3] detailní (nejsou agregovaná), nekozistentní (kontrolovaná až v datovém skladu), neobsahují historii (pouze aktuální data), měnící se (jsou nahrazována po jejich zpracování) a jsou ve stejné struktuře jako ve zrojových systémech.

5.5 Operativní uložiště dat

Stejně jako DSA je i operativní uložiště dat (ODS, Operational Data Store) nepovinou součástí BI. Slouží k uchovávání aktuálních dat, která jsou přístupná

s velmi nízkou dobou odezvy a jsou například určena pro analýzu malého objemu dat nebo pro aktuální získávání informací o profilu zákazníka v support centrech. ODS však ale někdy bývá pouze řešením špatně navrženého modelu datového skladu či problém nekompatibility aplikací s datovým skladem anebo nevhodné platformy.

5.6 Open source a freeware databázové systémy

Open source databázové systémy jsou zpravidla vyvýjeny spolu s komerčními produkty, které jsou doplněny o rozšířenou podporu, školení apod., nebo jsou vyvýjeny za přispění dalších společností. Počet open source databází je samozřejmě větší, ale především pro chybějící či špatnou dokumentaci v prostředí téměř nepoužitelné.

Další skupinou (Freewarové databázové systémy) jsou systémy, které jsou freewarovou verzí komerčních produktů. Jsou zpravidla omezeny na velikost podporované databáze, počet procesorů, velikost podporované RAM a další. Jejich podpora je obdobná jako u komerčních produktů – velmi kvalitní dokumentace, FAQ, placená podpora a školení je ve většině případů zpoplatněné.

Všechny databázové systémy mají přístup pomocí ODBC případně JDBC, pro zprávu lze použít i grafické rozhraní, které bývá buď součástí nebo je řešeno formou produktů třetích stran.

Limity databází jsou zpravidla ovlivněny operačním systémem, všechny níže zmiňované open source databázové systémy mohou obsahovat tabulky až v řádech terabajtů. U Freewarových produktů se vyskytují limity databází, hlavně pak omezení velikosti databáze, které jsou zavedeny hlavně z marketingových důvodů, což je znevýhodňuje pro použití v datových skladech a na větším objemu dat

obecně. Vyjímkou je zde IBM DB/2 Express-C 9, který má limity nastaveny tak, že ho lze použít i pro větší rozsáhlejší databáze.

5.6.1 MySQL

Verze:	**5.0**
Standardy:	**ANSI-SQL92/SQL99, ACID**
Operační systém:	**UNIX/Linux, Windows**
Maximální velikost dat:	**pouze podle OS**
Max. počet procesorů:	**pouze podle OS**
Max. velikost RAM:	**pouze podle OS**
Licence:	**GPL**
Výrobce:	**MySQL AB.**
Odkazy:	**http://www.mysql.com**

MySQL Community server 5.0 je jedním z nejčastěji používaným open source databázovým serverem. Pro grafickou zprávu zde není implementován žádný grafický klient, užívají se klienti buď přímo od MySQL (MySQL Administrator 1.2, MySQL Query Browser 1.2 a MySQL Migration Toolkit 1.1 vše pod GPL licencí) nebo produkty třetích stran (např. pro komerční užití zpoplatněný Navicat www.navicat.com nebo phpMyAdmin www.phpmyadmin.net pod licencí GPL). Používá několik modulů pro zpracování dat (storage engine), jsou to např. InnoDB nebo MyISAM. Mezi jeho výhody patří podpora mnoha operačních systémů. Další nepostradatelnou výhodou je kvalitní dokumentace, která je dostupná na stránkách výrobce, FAQ, internetová diskuzní fóra a v neposlední řadě také publikace o MySQL, které jsou dostupné i v češtině.

5.6.2 PostgreSQL

Verze:	**8.2**
Standardy:	**ANSI-SQL92/SQL99, ACID**
Operační systém:	**UNIX/Linux, Windows**
Maximální velikost dat:	**pouze podle OS**
Max. počet procesorů:	**pouze podle OS**
Max. velikost RAM:	**pouze podle OS**
Licence:	**BSD**
Výrobce:	**PostgreSQL Global Development Group**
Odkazy:	**http://www.postgresql.org** **http://www.pgadmin.org**

ORDBMS databázový server PostgreSQL je dalším oblíbeným databázovým systémem, který je vyvýjen již patnáct let a byl postaven na systému Postgress. V současnosti podporuje také velkou část standardů ASCI-SQL2003. Pro grafickou správu lze použít mnoho nástrojů jako např. PgAdminIII, PhpPgAdmin nebo komerční Navicat. Vyskytuje se zde opět kvalitní dokumentace, FAQ, internetová fóra a publikace.

5.6.3 Firebird

Verze:	**2.0.3**
Standardy:	**ANSI-SQL92/SQL99, ACID**
Operační systém:	**UNIX/Linux, Windows**
Maximální velikost dat:	**pouze podle OS**
Max. počet procesorů:	**pouze podle OS**
Max. velikost RAM:	**pouze podle OS**
Licence:	**IPL, IDPL**
Výrobce:	**Firebird Foundation Inc.**
Odkazy:	**http://www.firebirdsql.org**

Firebird je dalším open source RDBMS systémem, který vychází z InterBase 6.0. Firebird existuje ve dvou variantách Classic a Super server. Classic nabízí více aplikacím současný přímý přístup do databáze a také vzdálený přístup k databázi. Super server poskytuje pouze serverové procesy, klient tak nemůže přistupovat k databázi a všechny SQL požadavky jsou provedeny přes server s použitím

socketu. Pro grafickou správu lze použít DbVisualizer verze 1.1 pod LGPL licencí nebo Flame Robin 0.8.0 pod IDPL licencí. Dokumentace je u tohoto produktu dobrá, využít lze také diskuzních fór a FAQ.

5.6.4 Apache Derby

Verze:	**10.3**
Standardy:	**ANSI-SQL92/SQL99/SQL2003, ACID**
Operační systém:	**UNIX/Linux, Windows**
Maximální velikost dat:	**pouze podle OS**
Max. počet procesorů:	**2 (r.2004)**
Max. velikost RAM:	**pouze podle OS**
Licence:	**Apache License v. 2**
Výrobce:	**Apache Software Foudation**
Odkazy:	**http://www.apache.org** **http://db.apache.org/derby**

Předchůdce Apache Derby databázový systém společnosti Cloudscape byl odkoupen společností IBM a v roce 2004 byl uvolněn zdrojový kód a věnován Apache Software Foudation. Tento produkt je založen na jazyce Java. Vedle nízkých nároků na vlastní běh a instalaci je výhodou také podpora XML souborů a rozsáhlá dokumentace. Jako další zdroje dat mohou posloužit i FAQ a diskuzní fóra.

5.6.5 HSQLDB

Verze:	**1.8.0**
Standardy:	**ANSI-SQL92/SQL99/SQL2003**
Operační systém:	**UNIX/Linux, Windows**
Maximální velikost dat:	**8 GB**
Max. počet procesorů:	**neuvedeno**
Max. velikost RAM:	**neuvedeno**
Licence:	**vlastní**
Výrobce:	**The HSQL Development Group**
Odkazy:	**http://hsqldb.org**

HSQLDB je RDBMS databázový stroj naprogramovaný v jazyce Java. Pro

připojení používá JDBC, který však není plně implementován. Další nevýhodou je maximální limit objemu dat 8 GB v poslední verzi. Na druhou stranu nabízí velmi vysoký výkon, který dosahuje díky držení dat v operační paměti a relativně malou velikost instalace. Umožňuje ale také pracovat s daty uloženými na disku. Dokumentace je relativně rozsáhlá, dostupná je i podpora formou FAQ nebo fóra.

5.6.6 Freewarové databázové systémy

Název:	IBM DB/2 Express-C 9	MS SQL Server 2005 Express	Oracle XE	Sybase ASE Express Edition 15
Verze:	9.1	2005-sp1	10.2	15.0
OS:	UNIX/Linux, Windows	Windows	UNIX/Linux, Windows	Linux
Max. dat:	pouze podle OS	4 GB	4 GB	5 GB
Max. proces.:	2	1	1	1
Max. RAM:	4 GB	1 GB	1 GB	2 GB
Licence:	Freeware	Freeware	Freeware	Freeware
Výrobce:	IBM Corp.	Microsoft Corp.	Oracle Corp.	Sybase Inc.
Odkazy:	http://www.ibm.com	www.microsoft.com	www.oracle.com	www.sybase.com

Freewarové systémy jsou nabízeny jako alternativa komerčních databázových systémů. Nejdále zde zachází IBM, kdy nabízí neomezenou velikost databáze, limity jsou pouze na počet procesorů a RAM, ale i zde se jedná o velmi vysoký výkon. Tyto produkty jsou odvozeny od vyšších verzí, nabízí tak kvalitní technologický základ.

5.7 Hodnocení

Kritérii pro hodnocení databázových systémů jsou:

- Dokumentace a podpora (Dokumentace), především pak její kvalita o obsáhlost a také dostupnost české literatury, 20%.
- Podporované standardy (Standardy), kde je posuzováno plnění standardů SQL92/99/2003 a případně dalších (ACID) 20%.

- Omezení velikosti databáze, počtu procesorů a velikosti operační paměti RAM (Omezení), kde za každé významné omezení je odečten jeden až dva body, 30%.
- Rozhraní (Rozhraní), kde kritériem je především intuitivní a funkční grafické ovládání pro správu databázového systému, 10%.
- Operační systém (Operační systém), kde je sledován počet podporovaných operačních systémů, 20%.

Na základě těchto kritérií se nejlepe umístil MySQL, který pouze nepodporuje více operačních systémů. Na dalších místech se umístily databázové systémy PostgreSQL, Firebird a IBM DB/2, který díky omezení pouze počtu procesorů (2 CPU) je tak nejlepším freewarovým produktem v této kategorii.

6 Analýza OLAP

OLAP (OnLine Analytical Processing) je proces analýzy podnikových dat v reálném čase. K tomuto účelu se používá multidimenzionální kostky, ve které každá dimenze představuje hodnoty jako jsou například čas, region či pobočka nebo produkt – schéma 13. Spojením těchto parametrů tak například dostáváme prodej jednotlivých produktů v regionech za určité období. Sledované číselné hodnoty se označují míry (Measure), mohou jimi být například cena, náklady, počet kusů apod. Obdobně tak lze přistupovat i k dalším ukazatelům s využitím dalších dimenzí (věk a pohlaví zákazníků nebo sledování dle dodavatele – prodej dané elektroniky od daného dodavatele v daném regionu, kterou kupují muži ve věku od 25 do 35 let).

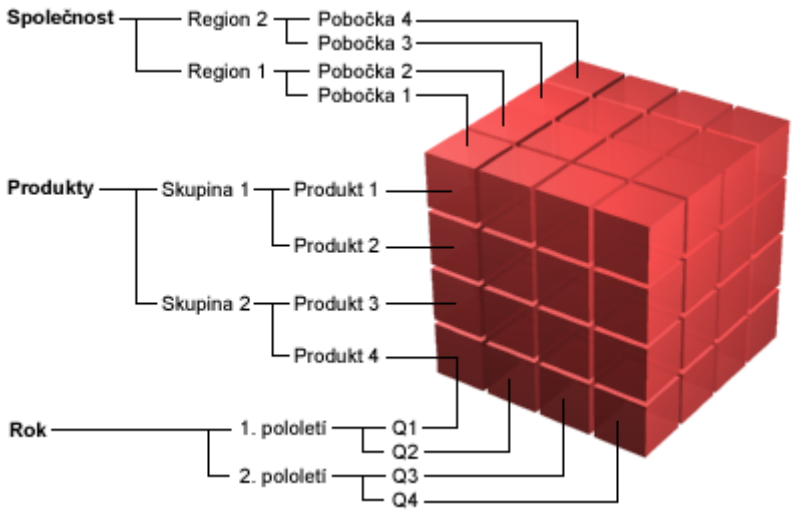


schéma 13 – OLAP kostka

Data pro OLAP analýzu se získávají z datových skladů resp. datových tržišť, kde jsou uložena v dimenzionálním datovém modelu – viz kapitola 5. Vlastní analýzu by bylo možné také provádět nad transakčními systémy a dalšími zdroji dat, výsledek by však nad větším objemem dat nebyl dosažitelný v přijatelném čase. Data jsou tak již v datových tržištích agregovaná, pro rapidní zkrácení doby výpočtu dané analytické úlohy – dle [5-3] se jedná 0,1% doby běhu nad transakčními systémy.

6.1 Základní operace

Mezi základní operace nad OLAP kostkou se řadí Nesting, Drill-Down, Drill-Up (Roll-Up), Dicing, Slicing a Pivot. Tyto operace umožňují získávat mnohem detailnější nebo naopak mnohem obecnější údaje o společnosti, tak i umožňují různé úhly pohledu na data v OLAP kostce.

6.1.1 Nesting

Nesting zobrazovací technika, která umožňuje zobrazit tři a více dimenzí do jedné tabulky. Je to řešeno přidáním dalších řádků popř. sloupců s novou dimenzí k jednotlivým již zobrazeným. Lze tak zvýšit výpovědní hodnotu tabulky. V tabulce 1 je zobrazen prodej za první pololetí, v tabulce 2 za druhé pololetí a v tabulce 3 je zobrazen prodej za celý rok podle pololetí.

Prodej produktů v ks	Produkt 1	Produkt 2	Produkt 3	Produkt 4
Region 1	365	423	95	12
Region 2	215	301	55	10

Tabulka 1 - Prodej za první pololetí

Prodej produktů v ks	Produkt 1	Produkt 2	Produkt 3	Produkt 4
Region 1	388	498	108	52
Region 2	220	300	64	34

Tabulka 2 - Prodej za druhé pololetí

Prodej produktů v ks		Produkt 1	Produkt 2	Produkt 3	Produkt 4
Region 1	1. pololetí	365	423	95	12
	2. pololetí	388	498	108	52
Region 2	1. pololetí	215	301	55	10
	2. pololetí	220	300	64	34

Tabulka 3 – Prodej za celý rok

6.1.2 Drill-Down

Drill-Down umožňuje detailnější náhled snížením agragační úrovně. Lze tak získat například při procházení údajů o prodejnosti v regionech přehled o prodejnosti jednotlivých poboček a tím zpřesnit celou analýzu – viz schema 14 a schema 15.

schema 14 – Drill-Down, Drill-Up

6.1.3 Drill Up (Roll Up)

Drill-Up (označován také jako Roll-Up) je obrácenou operací k Drill-Down. Zde se užívá vyšší agregace dat, získáváme tak souhrnější pohledy.

Prodej produktů v ks		Produkt 1	Produkt 2	Produkt 3	Produkt 4
Region 1	1. pololetí	365	423	95	12
	2. pololetí	388	498	108	52
Region 2	1. pololetí	215	301	55	10
	2. pololetí	220	300	64	34

Drill-Up

Drill-Down

Prodej produktů v ks			Produkt 1	Produkt 2	Produkt 3	Produkt 4
Region 1	Pobočka 1	1. pololetí	130	200	62	3
		2. pololetí	150	230	58	21
	Pobočka 2	1. pololetí	235	223	33	9
		2. pololetí	238	268	50	31
Region 2	Pobočka 3	1. pololetí	100	149	31	2
		2. pololetí	104	165	33	15
	Pobočka 4	1. pololetí	115	152	24	8
		2. pololetí	116	135	31	19

Schéma 15 – Drill Down, Drill Up

6.1.4 Slicing a Dicing

Tyto operace provádí omezení na počtu dimenzí. Lze tak například uzamknout výběr pouze pro daný region, produkt či časové rozmezí. Pokud uzamykáme pouze jednu dimenzi, jedná se o slicing – tabulka 4 , pokud takto uzamykáme více dimenzí, pak se jedná o dicing – tabulka 5.

Prodej produktů v ks		Produkt 1
Region 1	1. pololetí	365
	2. pololetí	388
Region 2	1. pololetí	215
	2. pololetí	220

Tabulka 4 - Silicing

Prodej produktů v ks		Produkt 1
Region 1	1. pololetí	365
	2. pololetí	388

Tabulka 5 - Dicing

6.1.5 Pivot

Operace Pivot provede přetočení dimenzí, čímž poskytuje další pohled na sledovaná data.

6.2 Architektury OLAP

V současnosti jsou nejčastěji používáné architektury OLAP jsou odvozené od způsobu uložení dat. Existují tak architektury Multidimenzionální OLAP (MOLAP), Relační OLAP (ROLAP), Hybridní OLAP (HOLAP) a architektura Desktop OLAP (DOLAP).

6.2.1 MOLAP

Architektura MOLAP je charakteristická speciálním ukládáním dat v multidimenzionálních-binárních OLAP kostkách. Výhodou této architektury je vysoký dotazovací výkon díky optimalizovanému uložení, multidimenzionálnímu indexování a cachování, menší velikost dat oproti uložení v relační databázi, automatizované dopočítávání vyšších agregací a učinná extrakce dat dosažená předuspořádáním agregovaných dat. Mezi nevýhody této architektury se řadí pomalý výkon procesu s velkým objemem dat a problematické přistupování některými MOLAP nástroji k většímu počtu dimenzí (10 a více).

6.2.2 ROLAP

ROLAP je architektura, založená na řešení uložení multidimenzionálních dat v relační databázi. Výhodou tohoto řešení je dobrý výkon při práci nad mnoha dimenzemi, nevýhodou je však oproti MOLAP nízký výkon při dotazování a vyšší objem celkových dat. Srovnání složitosti, výkonu a velikosti dat je na schématu 16.

6.2.3 HOLAP

Architektura HOLAP spojuje výhody obou předchozích, kdy detailní data jsou uložena v relačních databázích, agregovaná a často používaná data jsou uložena v multidimenzionálních-binárních OLAP kostkách. Tato architektura je v současnosti nejpoužívanější.

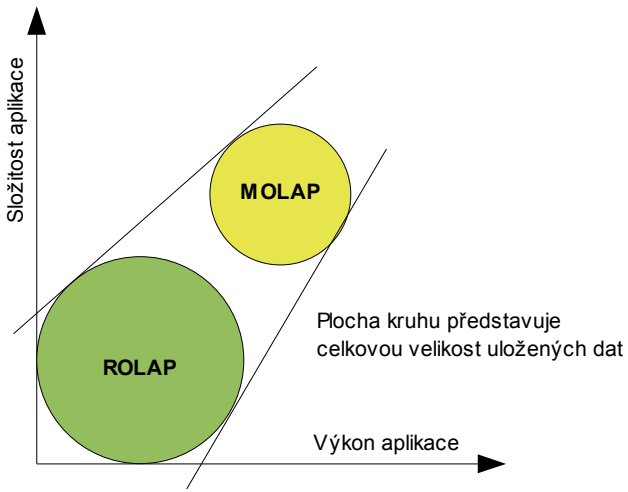


Schéma 16 – srovnání ROLAP a MOLAP, převzato [7]

6.2.4 DOLAP

Tato architektura se objevuje od konce devadesatých let. Umožňuje připojení k centrálnímu úložišti OLAP dat a stažení potřebné podmnožiny kostky na lokální počítač, kde se s daty nadále pracuje nad lokální kostkou. Tento způsob je výhodný pro obzvláště pro mobilní aplikace a uživatele, neboť není zapotřebí stálého připojení k serveru.

6.2.5 Další Architektury

Dále existují další architektury OLAP, které jsou již odvozeny od výše uvedených. Mezi takové patří Web-based OLAP (WOLAP), Real-Time OLAP (RTOLAP) a Spatial OLAP (SOLAP).

6.3 MDX a mdXML

MDX (Multidimensional Expressions) je dotazovacím jazykem pro OLAP databáze, který je velmi podobný jazyku SQL. Původně byl vyvinut ve společnosti Microsoft v roce 1997 spolu se specifikací OLE DB for OLAP (ODBO) a stal se součástí Microsoft OLAP Services 7.0 – nejednalo se však o otevřený formát. Později se stal standardem a byl nabídnut i dalším výrobcům analytických nástrojů. Na MDX je založen mdXML. Zatímco MDX je jazyk specifikovaný pro ODBC, mdXML je jazykově nezávislý a je více využívá výhod XMLA.

6.4 XMLA

XML for Analysis (označované taky jako XML/A) je rozhraní pro přístup k datům OLAP analýzy a data miningu a je založena na standardech XML, SOAP a

HTTP. Jedná se o rozhraní umožňující klientským aplikacím komunikaci s multidimenzionálními nebo OLAP datovými zdroji umístěmými v internetu. Toto rozhraní bylo představeno společností Microsoft v devadesátých letech minulého století.

6.5 Open Source OLAP

V prostředí open source se vyskytuje spousta OLAP serverů a klientů, ale jen málo je jich kvalitních. Lze se tak sice setkat se spoustou zajímavých projektů, ale absence stabilní verze, nekvalitní (či dokonce žádná) dokumentace, popřípadě absence anglické verze nasazení v podnikové sféře znemožňuje. Do představení OLAP nástrojů se tak nedostal zajímavý projekt OpenOLAP určený pro databázi PostgreSQL nebo ve verzi pro MySQL, který užívá jak MOLAP tak i ROLAP, z důvodu absence jakékoliv anglické dokumentace (pouze v japonštině) tak neumožňuje jeho nasazení. Spousta projektů se také vyskytuje ve verzích Alpha nebo Beta, nebo nejsou již aktivně dál vyvíjeny. Na druhé straně tak stojí server Mondrian, který s kvalitní dokumentací, doprovodnými grafickými nástroji a podporovanými standarty je nejrozšířenějším open source OLAP serverem. Do srovnání se tak dostávají pouze dva OLAP servery a to Mondrian a Palo. Spolu s nimi jsou představeny i klienti pro interpretaci dat, kteří umožňují základní operace jako Drill-down, Drill-up a další.

6.5.1 Mondrian a jPivot

Verze:	**2.4.2**
Typ:	**Server, klient (jPivot)**
Architektura:	**ROLAP**
Programovací jazyk:	**Java**
Operační systém:	**UNIX/Linux, Windows**
Licence:	**GPL**
Výrobce:	**Pentaho corp.**
Odkazy:	**http://www.pentaho.org**

OLAP server Mondrian napsaný v jazyce Java je spolu s klientem jPilot pod názvem Pentaho Analyst součástí Pentaho BI Suite, SpagoBI nebo také JasperSoft BI Suite. Jedná se o ROLAP server, který je dále rozložen do čtyř vrstev: Prezentační vrstva je určená k interakci s uživatelem, kde slouží nejen pro zobrazování výsledků reportů, ale také k vytváření nových dotazů, změně dimenzí nebo pro export získaných dat do některého z dalších formátů jako např. JPG, GIF nebo XML. Dimenzionální úroveň je určená pro práci s jazykem MDX. Provádí tak parsování, validování a vlastní vykonání jazyka MDX. Třetí úrovní je vrstva hvězdy, která je odpovědná za údržbu agregovaných dat v paměti, která pak zasílá do dimenzionální (druhé) vrstvy. Čtvrtou vrstvou je pak RDBMS, kde jsou uložena jednotlivá data připravena pro agregaci.

Server Mondrian je v řešení Pentaho dále doplněn sadou nástrojů (Schema Workbench nebo Cube Designer), které umožňují grafickou správu OLAP kostky. Mondrian je nejčastěji používaný OLAP server a to také díky velmi kvalitní a detailní dokumetaci a rozsáhlé podpoře formou fóra a FAQ.

jPilot je klient pro práci se serverem OLAP, se kterým komunikuje pomocí XMLA. Nejčastěji se jPilot aplikuje spolu se serverem Mondrian, ale díky rozhraní XMLA je možné nasazení pro jiné servery. Nabízí základní fukce jako Drill-down, Drill-up, Slice, Dice a Nesting. Dále nabízí práci s mapami a tvorbu MDX.

Je využíván v Pentaho BI Suite nebo OpenI. Dokumentace je základní, existuje však několik diskuzních fór. Z projektu jPivot vychází i další OLAP klient JRubik, který využívá knihovny jPivot.

6.5.2 Palo

Verze:	**2.0**
Typ:	**Server, Klient (Excel add-in)**
Architektura:	**MOLAP**
Programovací jazyk:	**C++**
Operační systém:	**UNIX/Linux, Windows**
Licence:	**GPL, Freeware**
Výrobce:	**Jedox GmbH**
Odkazy:	**http://www.jedox.com** **http://www.palo.net**

Tento nástroj slouží pro analýzu dat v Microsoft Excel. Skládá se ze dvou částí: Server, který je ditribuován pod GPL licencí a přístupný ve verzi pro UNIX/Linux nebo Windows. Další částí je klient, který je řešen formou add-in pro Microsoft Excel, který je však již distrbuován jako freeware. Modelování OLAP kostky a správa tohoto MOLAP serveru probíhá pomocí přehledného grafického rozhraní pro webový prohlížeč. Přístup k serveru probíhá pomocí XMLA, což umožňje jeho implementaci i do dalších řešení mimo MS Excel. V současnosti je připravován i add-in pro OpenOffice Calc. Součástí je také kvalitní dokumentace a podpora formou fóra.

6.6 Hodnocení

Kritéria pro hodnocení reportovacích nástrojů jsou:

- Dokumentace a podpora (Dokumentace), především pak její kvalita a obsáhlost, 30%.

- Propojení s dalšími aplikacemi (Propojení), kde je sledováno propojení pomocí standardů XMLA a jazyka MDX, 20%.
- Hodnocení preferovaného klienta (Klient), kde je slodován jeho typ (tenký klient – web, desktopový produkt), jeho dokumentace a rozhraní 30%.
- Operační systém (Operační systém), kde je sledován počet podporovaných operačních systémů, 20%.

Jelikož byly v tomto srovnání posuzovány pouze dva OLAP produkty s rozličnou architekturou (MOLAP a ROLAP), volba vhodného produktu se tak omezí pouze na potřebu dané architektury. Lepším na základě výše zmiňovaných kritérií je Mondrian, avšak pouze díky preferovanému klientovi.

7 Dolování dat

Dolování dat (Data Mining) je další součást Business Intelligence. Objevuje se dnes jako součást všech významných komerčních databázových produktů, což odráží mimo jiné důležitost pro správu a řízení podniku. V Business Intelligence se začalo objevovat v devadesátých letech minulého století, tedy až v době, kdy výpočetní výkon a disková kapacita už umožňovala zpracovat dostatečné množství podnikových dat pro dolování. Data mining tak využívá velký objem dat přímo v datovém skladu a zpravidla nepřistupuje k agregovaným datům v datových tržištích.

Oproti OLAP analýze, kdy se pro analytické úlohy používá deduktivní způsob práce (předem vytvořené množiny hypotéz, které se pomocí OLAP analýzy potvrzují nebo vyvrací), analýza založená na dolování dat se opírá o induktivní způsob práce, tedy pomáhá tyto hypotézy vytvářet.

Pro zpřesnění pojmu dolování dat použijeme definici [4]: *Dolování dat lze charakterizovat jako proces extrakce relevatních, předem neznámých nebo nedefinovaných informací z velmi rozsáhlých databází. Důležitou vlastností dolování dat je, že se jedná o analýzy odvozované z obsahu dat, nikoli předem specifikované uživatelem nebo implementátorem, a jedná se především o odvozování predikativních informací, nikoli pouze deskriptivních. Dolování dat slouží manažerům k objevování nových skutečností, čímž pomáhají zaměřit jejich pozornost na podstatné faktory podnikání, umožňují testovat hypotézy, odhalují*

ve stále se zrychlujícím a složitějším obchodním prostředí skryté korelace mezi ekonomickými proměnnými apod.

Donedávna bylo nasazení Data mining pouze otázkou pro velké společnosti, které byly schopny plně využít tohoto nástroje. Malé a střední společnosti však postupem času byly schopné také nasbírat dostatečné množství dat potřebných pro tuto analýzu. Jako zdroj zde mohou sloužit nákupní košíky zákazníků v elektronických obchodech, monitorování chování při návštěvě webových stránek nebo nespokojenost zákazníků projevovaná zvýšenou návštěvností technické podpory. Tyto společnosti tak většinou pořizují pouze komerční moduly pro jednotlivé úlohy (převážně statistické), neboť nákup obsáhlejšího nástroje je neefektivní.

7.1 Typy úloh dolování dat

Mezi úlohami dolování dat se můžeme setkat s Exploračními analýzami dat, které prozkoumávájí data bez předchozích znalostí vymezení problému. Oproti tomu Deskriptivní úlohy pracují s předem vymezenými kritérii, kde jako příklad lze uvést dedukci shluků. Predikativní úlohy předpovídají na základě znalosti ostatních hodnot veličin dosud neznámou hodnotu veličiny. Hledání vzorů a pravidel je hledání vstahů mezi jednotlivými veličinami popř. chováním. Lze tak například provádět analýzu nákupního košíku či detekce praní špinavých peněz v bankovnictví. Posledním typem úloh je hledání podle vzorů, kdy se na vstup použije vzor, ke kterému se hledají v datech stejné vzory.

7.1.1 Úlohy v podnikatelském prostředí

Mezi nejčastěji používané úlohy v podnikatelském prostředí se podle [13] řadí:

- Analýza úvěrového rizika – na základě doposud nashromážděných dat chování současných klientů lze určit míra rizika nesplácení úvěru.
- Vyhodnocení marketingových kampaní – na základě vzorku zákazníků umožňuje vybrat skupinu zákazníků, kde se marketingová kampaň setká s nejlepší odezvou, čímž se sníží vynaložené náklady na kampaň.
- Analýza odchodu zákazníků – nabízí možnost nalézt zákazníky s největším sklonem ke změně dodavatele.
- Segmentace zákazníků – rozdělení zákazníků do skupin pro marketingové účely nebo pro účely stanovení limitů pro úvěr.
- Analýza chování zákazníků – určení na základě historických dat vedených o chování zákazníků lze zlepšit nabídku služeb.
- Analýza produktů – sledování produktů u zákazníka a nalezení dalších, pro tohoto zákazníka vhodných, produktů, čímž lze cíleně směřovat marketingovou kampaň.

Obdobně lze přistupovat například k analýzám dodavatelů, kde je možné sledovat kvalitu dodaných produktů (kazovost) k nákupní ceně, zpoždění dodávek, či na základě reklamace zákazníků.

7.2 Techniky dolování dat

Pro dolování dat se používá množství technik, obsahujících statistické a matematické funkce. Pojmenování a případně rozdělení těchto technik se liší.

Jsou to například:

- Rozhodovací stromy a indukce (Decision Trees and Rule Induction) – data jsou zobrazeny v podobě stromu, každý uzel pak definuje podmínky (kritéria) pro nasledné rozdělení. Data jsou rozdělena do segmentů, kde každý list obsahuje segment dat, který byl definován v předchozích uzlech. Tato data se vyznačují shodnými popř. obdobnými charakteristikami. Díky snadné interpretaci jsou rozhodovací stromy často používanou technikou.
- Neuronové sítě (Neural Networks) – nejčastěji se používají pro tvorbu predikativních modelů. Je zde uplatněna zjednodušená anologie s neurony v lidském mozku, tedy navzájem propojené velké množství elementů. Každý takový elemet na základě přijatých tréninkových dat má vlastní nastavení parametrů, aby výsledná konfigurace pak nejlépe vyhovovala následné konfiguraci a predikci.
- Genetické algoritmy (Genetic Algorithm) – využívá mechanismy genetiky a přirozeného výběru pro nalezení optimální množiny parametrů. Tato technika neslouží přímo k predikci určitých hodnot, ale k vývoji popř. parametrizaci dalších modelů.
- Nejbližší soused a dedukce shluků (Nearest Neighbor and Clusteing) – technika, při níž se vyhledávají nejbližší prvky k prvku, který byl již predikován, nebo se vyskytuje v historii a to podle stanovené klasifikace.
- Analýza nákupního košíku (Market Basket Analysis) – tato analýza, založená na clusteringu, se opírá o produkty, požadované společně v jedné transakci. Typickým příkladem může být sekce „s tímto produktem zákazníci nejčastěji kupovali“ v elektronických obchodech.
- Analýza závislostí (Link Analysis) – sleduje prvky na základě vstahů

mezi prvky, nikoli na základě jejich vlastností.

- Dedukce (Memory-Based Reasoning) – sledováním známých skutečností slouží tato technika k predikci, kdy sleduje nejbližší okolí známých instancí a kombinuje tyto hodnoty k odhadu predikovaných hodnot.

7.3 Metodologie CRIPS-DM

Data miningová metodologie CRIPS-DM (CRoss Industry Standard Process for Data Mining) definuje šest základních etap, které se dále dělí do jednotlivých kroků. Vznik této metodologie byl inicializován Evropskou komisí pro standartizování postupů data minigových úloh. Později přešla pod CRIPS-DM konzorcium společností NCR Systems Engineering Copenhagen, DaimlerChrysler AG, SPSS Inc. a OHRA Verzekeringen en Bank Groep B.V. Jednotlivými etapami tedy podle [12] jsou :

- Porozumění problematice – tato počáteční etapa se zaměří na porozumění projektových cílů a požadavků z obchodní perspektivy, převádění této znalosti do popisu problému těžby dat a navržení úvodního projektu.
- Porozumění datům – etapa porozumění datům začíná počátečním sběrem dat, postupuje s aktivitami za účelem obeznámení se s daty, identifikuje problémy kvality dat, objevuje první náhledy na tato data, nebo zjišťuje zajímavé podmnožiny tvořící hypotézy pro skryté informace.
- Příprava dat – do této etapy se zahrnují všechny činnosti vedoucí ke konstruování cílového datasetu (tj. data, která se budou dodávat do modelovacích nástrojů) ze zdrojových dat. Úlohy spojené s přípravou

dat budou vykonávány zpravidla vícekrát, neexistuje žádný předem určený postup. Úkoly zahrnují tabulky, záznamy a volba atributů, ale také transformaci a čištění dat pro modelovací nástroje.

- Modelování – v této etapě jsou vybírány a aplikovány různé modelovací techniky a jejich parametry jsou nastaveny optimálními hodnotami. Většinou existuje několik data miningových technik pro stejný typ úlohy dolování dat. Nějaké techniky však mají specifické požadavky na formu dat, proto je zde možnost vrácení se do etapy přípravy dat často potřebná.
- Hodnocení výsledků – v této etapě projektu je hotový model (nebo modely), které se jeví, že mají vysokou kvalitu z perspektivy analýzy dat. Před konečným nasazením modelu je však důležité důkladněji ohodnotit model a zhodnotit všechny kroky vykonané k tomu, aby konstruovaly model řádně dosahující podnikových cílů. Podstatné je určit zda neexistuje nejaký důležitý obchodní problém, který nebyl doposud dostatečně zvážen. Na konci této etapy je rozhodnutí o použitelnosti výsledků tohoto data miningu, kterých mělo být dosaženo.
- Impementace modelu – vytvoření modelu obecně nepatří na konec projektu. I když účelem modelu je zvýšit znalost dat, tato získaná znalost musí být organizovaná a představená způsobem, kterému zákazník porozumí a může ho použít. V závislosti na požadavcích, etapa implementace modelu může být stejně tak velmi snadná jako komplikovaná. V mnoha případech to bude zákazník, nikoliv analytik, kdo provede implementační kroky. Důležité je, aby zákazník porozuměl především činnostem potřebých pro používání takového modelu.

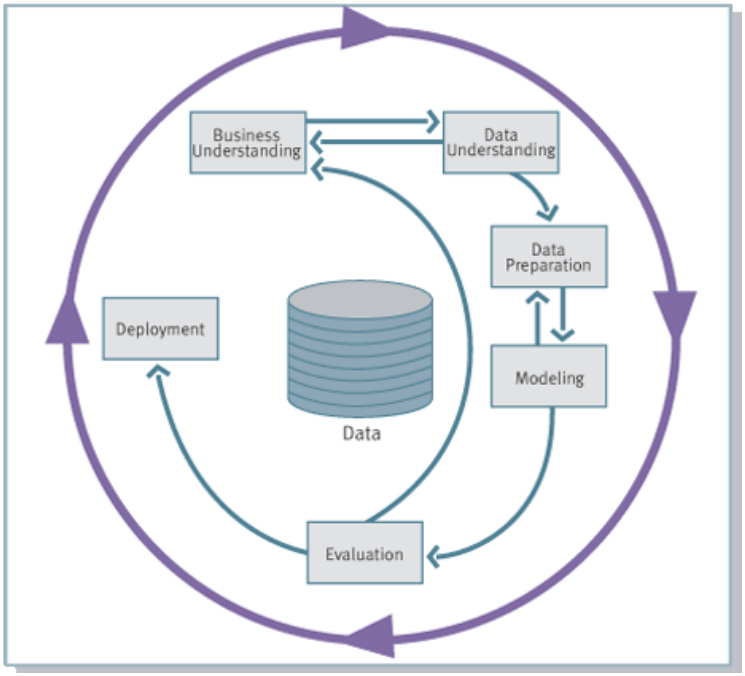


Schéma 17 – CRIPS-DM

7.4 Open Source Data mining nástroje

Open source Data mining nástroje jsou zpravidla vyvýjeny akademickou sférou (WEKA, Orange) nebo z těchto produktů odvozeny (RapidMiner). Všechny dále popisované nástroje užívají pro ovládání přehledné grafické rozhraní (obrázek %%7), ale kvalita provedení tu však kolísá. Všechny nástroje nabízí základní statistické funkce, které jsou důležité pro případné nasazení v malých a středních společnostech. Nevýhodou těchto nástroju je v porovnání s ostatními komponentami Business Intelligence mnohdy slabší dokumentace.

7.4.1 WEKA

Verze:	**3.5.6.**
Propojení:	**JDBC, flat-file**
Programovací jazyk:	**Java**
Operační systém:	**UNIX/Linux, Windows, MacOS**
Licence:	**GPL**
Výrobce:	**University of Waikato**
Odkazy:	**http://www.cs.waikato.ac.nz** **http://www.cs.waikato.ac.nz/ml/weka**

WEKA (Waikato Environment for Knowledge Analysis) je data miningový nástroj napsaný v jazyce Java. Nabízí možnost užití několika technik dolování dat jako jsou dedukce shluků, klasifikace, rozhodovací stromy nebo neuronové sítě. Pracovat s tímto nástrojem lze pomocí čtyř nástrojů s rozličným grafickým rozhraním: Explorer, Experimenter, KnowledgeFlow a příkazová řádka (ozn. Simple CLI). Jako vstup lze použít databáze připojené pomocí JDBC nebo flat-file soubory. Grafické prostředí je pěkně zpracované, dokumentace je obsáhlá, existuje i podpora formou internetových diskuzních fór a FAQ.

7.4.2 R-Project

Verze:	**2.6.0**
Propojení:	**JDBC (RJDBC), ODBC (RODBC)**
Programovací jazyk:	**C, C++, Fortran**
Operační systém:	**UNIX/Linux, Windows, MacOS**
Licence:	**GPL**
Výrobce:	**The R Foundation for Statistical Computing**
Odkazy:	**http://www.r-project.org/**

R je jazyk a zároveň prostředí pro statistické výpočty a grafy. Vznikl na projektu S, vyvýjeného od druhé poloviny sedmdesátých letech minulého století v Bell Labs. Nabízí statistické funkce, shlukování, analýzu časových řad, klasifikaci a další. Rozšíření funkcí lze pomocí add-ons, nebo pomocí programů napsaných v C, C++ nebo Fortran. Jako výstup mohou být použity dokumenty

PDF, HTML, LaTeX, textové soubory a další. Grafické prostředí je základní. Dokumentace je obsáhlá, rozsáhlá je i podpora formou FAQ a diskuzních fór.

7.4.3 Orange

Verze:	**0.9**
Propojení:	**XML, flat-file, CSV**
Programovací jazyk:	**C++, Python**
Operační systém:	**UNIX/Linux, Windows, MacOS**
Licence:	**GPL**
Výrobce:	**Faculty of Computer and Information Science, University of Ljubljana**
Odkazy:	**http://magix.fri.uni-lj.si/orange/**

Orange je vyvýjen na Faculty of Computer and Information Science, University of Ljubljana v jazyce C++ (core) a Pyton (moduly). Nabízí techniky klasifikace, shlukovaní, rozhodovací stromy a další. Rozšíření je možné pomocí dalších přídavných modulů, kdy tak lze získat například podporu neuronových sítí. Podporuje import/export do nejrůznějších formátů jako XML, textové soubory nebo CSV. Výhodou je kvalitně provedené předzpracování dat. Dokumentace je základní, doplněná internetovým diskuzním fórem.

7.4.4 RapidMiner

Verze:	**4.0**
Propojení:	**JDBC**
Programovací jazyk:	**Java**
Operační systém:	**UNIX/Linux, Windows**
Licence:	**GPL**
Výrobce:	**Rapid-I : Mierswa & Klinkenberg GbR**
Odkazy:	**http://www.rapidminer.com/**

RapidMiner (dříve YALE – Yet another Learning Environment) vychází z data miningového nástroje WEKA, odkud využívá knihovny a je nabízen v GPL nebo v komerční licenci. Stejně jako WEKA nabízí dedukci shluků, rozhodovací stromy,

klasifikaci nebo neuronové sítě. Součástí je také přehledný grafický průvodce pro nastavení úloh. Dokumentace k tomuto nástroji je kvalitní.

7.5 Hodnocení

Kritéria pro hodnocení reportovacích nástrojů jsou:

- Dokumentace a podpora (Dokumentace), především pak její kvalita a obsáhlost, 30%.
- Propojení s datovými zdroji (Propojení), kde je sledováno propojení k datovým zdrojům, především pomocí JDBC a dalších, 10%.
- Hodnocení rozhraní (Rozhraní), kde je posuzována kvalita a intuitivnost grafického rozhraní, 30%.
- Nabízené funkce (Funkce), pro provádění analýz, především statistických, 20%
- Operační systém (Operační systém), kde je sledován počet podporovaných operačních systémů, 10%.

V tomto hodnocení nejlépe dopadl nástroj Weka, následovaný nástrojem RapidMiner. Oba tak nabízejí statistické funkce a další jako například neuronové sítě. Dokumentace je u všech produktů obsáhlá, pouze Orange má dokumentaci základní.

8 Reportovací nástroje

Reportovací nástroje slouží k vizualizaci dat z datového skladu nebo z jiných systémů společnosti. Zobrazují data nejčastěji formou tabulek nebo grafů, lze se však setkat i s jinými formami, jako například zobrazení prodejnosti v regionech pomocí map. Realizují se nejčastěji pomocí SQL dotazů nad databázemi. Reporty se rozdělují do dvou základních kategorií:

- Standardní reporting – v předdefinovaných časových intervalech se spouští předpřipravené dotazy
- Ad hoc reporting – dotazy jsou spouštěny a definovány přímo uživatelem dle jeho potřeby

8.1 Standardní a Ad hoc reporting

Standartní reporting slouží například k měsíčním, kvartálním a ročním souhrnům prodeje, definování je provedeno již během implementace nebo na počátku provozu systému. Lze postupně doplňovat a není tolik závislé na znalostech uživatele. Možnost ad hoc reportingu je velmi závislá na kvalitě uživatelského prostředí, které musí umožňovat intuitivní ovládání reportovacího nástroje, čímž se snižuje nutnost znalosti jazyka SQL a databází. Výstupem těchto reportů mohou být HTML stránky, textové soubory (PDF, DOC) nebo tabulkové soubory (XLS). Použité formáty jsou velmi důležité pro dodatečné úpravy připravených sestav.

8.2 Open Source reportovací nástroje

Open Source reportovací nástroje nabízí všechny podstatné funkce, někdy jsou doplněny i o nadstandardní, jako například tvorba čárkových kódů. Všechny nástroje nabízí zobrazení dat přímo na obrazovce, možnost přípravy tiskových sestav, většinou ve formátech US Letter nebo A4 a export do dalších formátů jako například PDF, XLS, XML a HTML. Formát DOC a následná editace v textovém editoru není podporována, což je asi největší nedostatek těchto nástrojů. Grafické rozhraní je zpracováno kvalitně a nabízí intuitivní ovádání. Dokumentace je zpravidla kvalitní, ale do výběru se například nedostal produkt Agata Report 7.5, neboť k němu v současnosti není dostupná dokumetace v angličtině, pouze v portugalštině.

8.2.1 JFreeReport

Verze:	**0.9.1**
Zdroje:	**JDBC, HTML**
Výstupy:	**HTML, PDF, XLS, XML, text**
Jazyk:	**Java**
Operační systém:	**Linux/Unix, Windows**
Licence:	**LGPL**
Výrobce:	**Enterprise House**
Odkazy:	**http://www.object-refinery.com** **http://www.jfree.org**

JFreeReport je použit vrámci Pentaho Business Intelligence Suite pod názvem Classic-Reporting Engine. Konfigurace reportování probíhá pomocí XML souboru, který lze editovat například Pentaho Report Designer nebo Pentaho Report Design Wizard. Nabízí export do formátů HTML, PDF, XLS, XML a textových souborů. Dokumentace pro reportovací nástroj i pro konfigurátory je kvalitní, podpora je doplněna FAQ a internetovými diskuzními skupinami.

8.2.2 DataVision

Verze:	**1.1**
Zdroje:	**JDBC, Text**
Výstupy:	**DocBook, HTML, LaTeX, PDF, XLS, XML**
Jazyk:	**Java**
Operační systém:	**Linux/Unix, Windows**
Licence:	**Apache**
Výrobce:	**Jim Menard**
Odkazy:	**http://datavision.sourceforge.net**

Reportovací nástroj DataVision je programován v jazyce Java a pro konfiguraci výstupu používá XML. Tvorba konfiguračního XML souboru probíhá ve vlastním grafickém rozhraní, je však samozřejmě možné konfigurovat i pomocí jiných editorů. Lze tak nastavit vzhled výsledné stránky včetě písma, loga, podkladu, záhlaví a zápatí. Provádí jednoduché příkazy (SELECT) jazyka SQL s možností doplnění dalších podmínek (WHERE) nebo agregace. Dokumentace je základní, doplněna o FAQ, což je vzhledem ke skutečnosti, že je tento produkt vyvíjen pouze jednou osobou, vcelku pochopitelné.

8.2.3 iReport

Verze:	**2.0.2**
Zdroje:	**JDBC**
Výstupy:	**PDF, HTML, XLS, CSV, RTF**
Jazyk:	**Java**
Operační systém:	**Linux/Unix, Windows**
Licence:	**GPL**
Výrobce:	**JasperSoft**
Odkazy:	**http://jasperforge.org/sf/projects/ireport**

iReport je součástí Jasper BI suite. Nabízí export do PDF, HTML, XLS a dalších. Dále nabízí možnost generování čárkových kódů. Tvorba reportů probíhá v kvalitním grafickém rozhraní. K databázi přistupuje pomocí JDBC, pro konfiguraci reportů využívá XML nebo speciální Jasper soubor. Dokumentace

je kvalitní, doplněná o názorné animace. Podpora je doplněna FAQ a diskuzním fórem. Je distribuován pod LGP licencí, za poplatek pak lze iReport získat i pod LGPL licencí.

8.2.4 Eclipse BIRT

Verze:	**2.1.3**
Zdroje:	**JDBC, XML, Flat-File, webové služby**
Výstupy:	**PDF, HTML**
Jazyk:	**Java**
Operační systém:	**Linux/Unix, Windows**
Licence:	**EPL**
Výrobce:	**Eclipse Foundation**
Odkazy:	**http://www.eclipse.org/birt/phoenix** **http://www.eclipse.org**

BIRT (Business Intelligence Reporting Tool) je nástavba open source vývojářského nástroje Eclipse, která je určená k tvorbě reportů. Nabízí tak kvalitní a intuitivní uživatelské prostředí s možností tvorby reportů (v Eclipse) pomocí průvodce a běhovou komponentu pro přidání do vlastního aplikačního serveru. Export je možný pouze do PDF a HTML. Dokumentace je velmi kvalitní, doplněná nazornými animacemi, podporou formou FAQ a diskuzního fóra.

8.3 Hodnocení

Kritéria pro hodnocení reportovacích nástrojů jsou:

- Dokumentace a podpora (Dokumentace), především pak její kvalita a obsáhlost, 20%.
- Zdroje dat pro tvorbu reportů (Zdroje), kde je posuzován počet vstupních zdrojů, kde vedle základního JDBC připojení je sledována možnost i dalších zdrojů. 20%.

- Formáty výstupu reporů (Výstupy), kde je vedle obrazových výstupů (Nejčastěji v HTML) sledovaná možnost exportu do dalších formátů, obzvláště pak PDF, XLS, XML a DOC, 20%.
- Rozhraní (Rozhraní), kde kritériem je především intuitivní a funkční ovládání, 20%.
- Operační systém (Operační systém), kde je sledován počet podporovaných operačních systémů, 20%.

Všechny hodnocené produkty mají svá slabá místa na různých pozicích. DataVision tak ztrácí hlavně kvůli dokumentaci, iReport pak z důvodu jediného zdroje dat (JDBC), Eclipse BIRT pak zase z důvodu exportu pouze do PDF a HTML. Nejvyváženější je pak jFreeReport, který s předešlými dvěma sdílí první místo.

9 Komplexní BI řešení

Další možností přístupu k řešení je použití komlexních balíků nástrojů BI. Výhodou tohoto řešení je již provedená integrace jednotlivých komponent do celku, čímž se snižuje potřebný čas pro doladění celého systému. Další výhodou je existence vlastního frameworku, který tak není nutné vyvíjet. Vše je postaveno na jednotném programovacím jazyce Java. Stačí zde pouze přiřadit datové zdroje, nakonfigurovat ETL/ELT, OLAP, Data mining, připravit reporty a upravit dle potřeby dashboard. Všechny balíky obsahují také demonstrační data, pomocí kterých se lze lépe seznámit s celým projektem. Dokumentace je rozsáhlá a kvalitní, obsahuje i mnoho demonstrativních animací. Lze se také setkat se specializovanými nástroji (např. Pentaho Design Studio) pro vývoj nebo modifikaci těchto balíků.

9.1 SpagoBI

Verze:	**1.9**
Programovací jazyk:	**Java, JSP (J2EE)**
Operační systém:	**Linux, Windows**
Datové zdroje:	**JDBC**
ETL/ELT:	**Talend OpenStudio**
Data mining:	**WEKA**
OLAP:	**Mondrian/jPivot**
Reportovací nástroje:	**JasperReport**
Licence:	**LGPL**
Výrobce:	**Engineering Ingegneria Informatica S.p.A.**
Odkazy:	**http://spagobi.eng.it**

SpagoBI je jedním z produktů vyvíjených italskou společností Engineering Ingegneria Informatica S.p.A., která se zabývá produkty pro řízení podniků.Celý

systém je procesně orientovaný, což umožňuje lepší řízení systému pomocí integrovaného workflow. Základem je aplikační Java framework Spago, který je doplněn o komponenty (FOSS) dalších výrobců. Komplexním řešením BI je pak SpagoBI, což je balík vybraných komponent a je rozdělen do tří vrstev: vrstva dat a metadat, analytická vrstva a prezenční vrstva (schéma 18). Dokumentace SpagoBI je rozsáhlá a kvalitní, ale obsahuje však některé nesrovnalosti, způsobené různými verzemi produktu, pro který jsou psané, neboť poslední verze nemá plnou dokumentaci. Dokumentace je také doplněná o názorná videa a v internetu dostupnou a funkční demoverzi produktu. Podpora je řešena FAQ a diskuzními fóry. Instalace je relativně snadná a dobře zdokumentovaná.

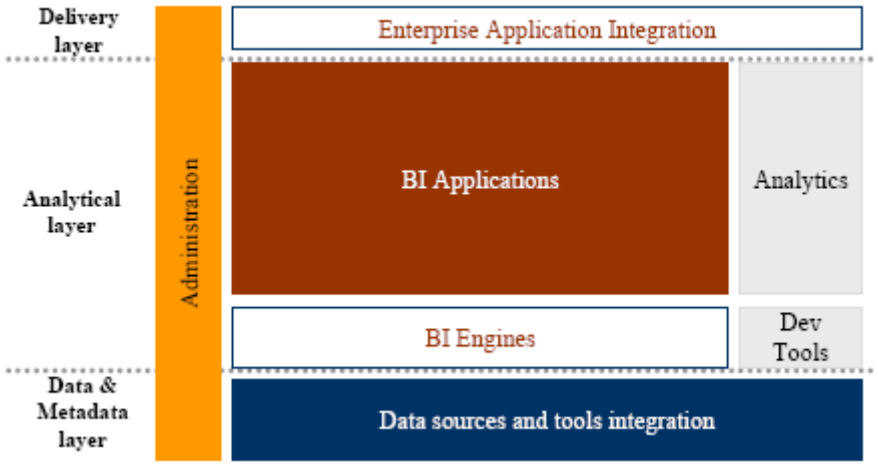


Schéma 18 – Základní architektura SpagoBI, převzato od výrobce

Datová vrstva

Datová vrstva obsahuje ETL nástroje (Talend OpenStudio[5]), datový sklad,

5 Kapitola 4 ETL nástroje

metadata a service repository (MySQL, Oracle, MS SQL server a PostgreSQL[6]).

Analytická vrstva

V analytické vrstvě je použit soubor komponent BI engine – Reportovací nástroje (JasperReport, BIRT[7]), OLAP (Jpivot, Mondrian[8]), dolování dat (Weka[9]), Dashboard (SWF, OpenLaszlo), QbE (Hibernate) a Geo (CartoWeb, MapServer, SVG). Další částí jsou modelování běhu a správy analytických nástrojů (plánování spouštění systému - Quartz[7], role v systému apod.).

Prezenční vrstva

Všechny důležité BI služby jsou přístupné pomocí portletů ve standardu JSR-168 (BIPortlet), což umožňuje jednoduchou integraci i do dalších portálů či systémů v rámci podniku. Distribuční vrstva dále obsahuje zpřístupnění pomocí webových služeb (BIService), zasílání výsledků a textů ve formátu XML (BIXCube) a systém zasílání zpráv pomocí JMS (BIMessage).

Většina komponent v datové a analytické vrstvě byla představena v předchozích kapitolách. Podrobnější schéma architektury je na schématu 19). .

Grafické rozhraní je přehledné a nabízí relativně intuitivní ovládání, které je i dobře představeno v názorných videosekvencích. Díky aplikaci Dossier, která nabízí možnost sdílení dokumentů vytvořených ať už ve SpagoBI či v jiných (i desktopových) aplikacích. Je zde také propracovaná podpora správy uživatelů

6 Kapitola 5 Datové sklady
7 Kapitola 8 Reportovací nástroje
8 Kapitola 6 OLAP nástroje
9 Kapitola 7 Dolování dat

a přístupových práv. Nevýhodou je absence jakékoli integrované nápovědy při používání tohoto produktu.

Pro běh internetového portálu je potřebný Exo nebo IBM Websphere Portal, aplikačními servery mohou být JOnAS, JBoss, Tomcat nebo IBM Websphere.

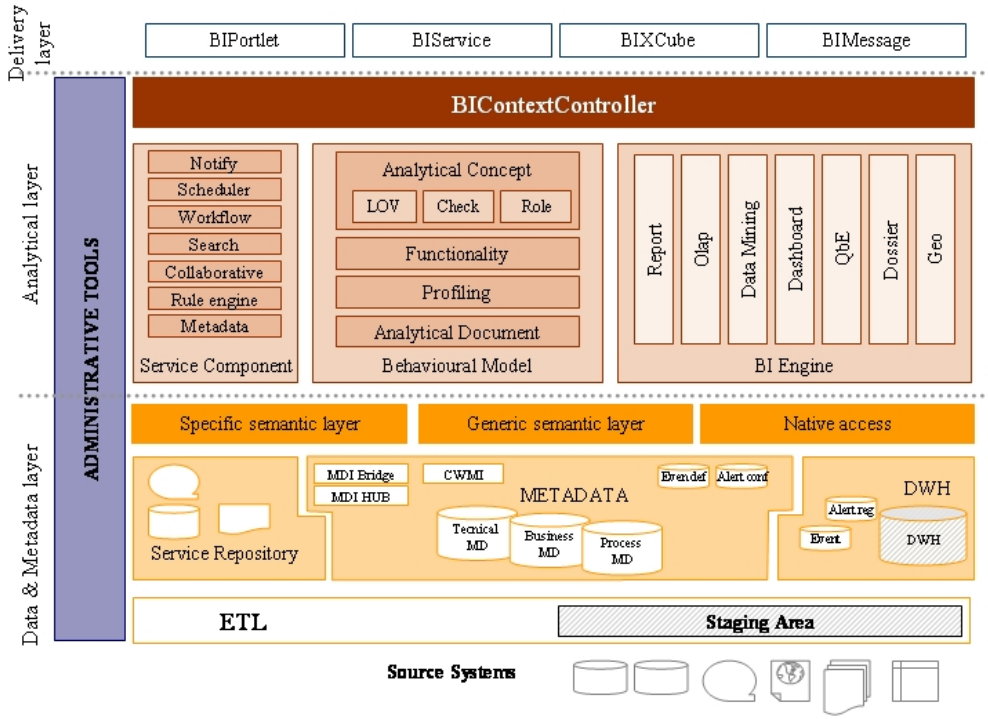


Schéma 19 – Detail architektury SpagoBI verze 1.9, převzato od výrobce

9.2 OpenI

Verze:	**1.3**
Programovací jazyk:	**Java, JSP (J2EE)**
Operační systém:	**Linux/Unix, Windows**
Datové zdroje:	**JDBC**
ETL/ELT:	**-**
Data mining	**R project**
OLAP:	**Mondrian/jPivot**
Reportovací nástroje:	**JasperReports**
Licence:	**MPL**
Výrobce:	**Loyalty Matrix, Inc.**
Odkazy:	**http://openi.sourceforge.net** **http://www.loyaltymatrix.com**

OpenI (vyslovované „Open eye“) je analytický balík nabízející OLAP analýzu (Mondrian/jPivot[10]), tvorbu reportů (JasperReports[11]) a dolování dat (R project[12]). Je naprogramován v jazyce Java, pro běh používá jakýkoliv aplikační framework postavený na J2EE (například Tomcat). Umožňuje základní nastavení tří uživatelských rolí: Aplikační administrátor, Projektový administrátor a Projektový uživatel. Pro OLAP analýzu používá XMLA a lze tak použít i Microsoft Analysis Services. ETL nástroj není součástí tohoto balíku. Dokumentace je základní, lze však vycházet i z dokumentace přiložených komponent. Výhodou je užití portletů (JSR-168), což umožňuje začlenění i do stávajících systémů. Jedná se o řešení určené spíše malým podnikům, kde nebudou kladeny větší nároky na tento produkt. Architektura OpenI je na schématu 20.

10 Kapitola 6 OLAP nástroje
11 Kapitola 8 Reportovací nástroje
12 Kapitola 7 Dolování dat

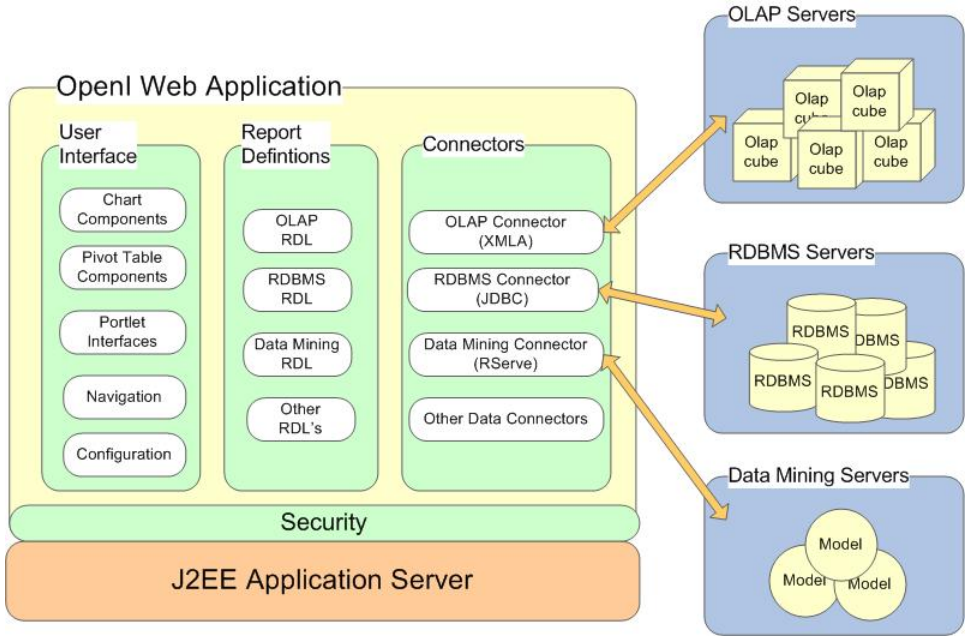


schéma 20 – Architektura OpenI 1.3, převzato od výrobce

9.3 JasperSoft BI Suite

Verze:	**2.0**
Programovací jazyk:	**Java, JSP (J2EE)**
Operační systém:	**Linux, Windows**
Datové zdroje:	**JDBC**
ETL/ELT:	**JasperETL (Talend OpenStudio)**
Data mining	**-**
OLAP:	**JasperAnalyst**
Reportovací nástroje:	**JasperReports, JasperStudio (iReport)**
Licence:	**GPL**
Výrobce:	**JasperSoft Corporation**
Odkazy:	**http://jasperforge.org** **http://www.jaspersoft.com**

JasperSoft BI Suite nabízí komplexní řešení BI, který společnost JasperSoft nabízí v open source a komerční verzi. Komerční verze oproti open source verzi nabízí rozšířenou podporu a možnost použít i jiné licence. Dokumentace tohoto

produktu je obsáhlá a kvalitní, obsahuje také dokumentaci integrovaných produktů a je doplněna o animace. Podpora je tvořena FAQ a internetovými diskuzními skupinami.

Architektura

Architektura JasperSoft BI Suite je rozčleněna do tří částí: datová, analytická a reportovací. Přehled základní architektury JasperSoft BI Suite je na schéma 20.

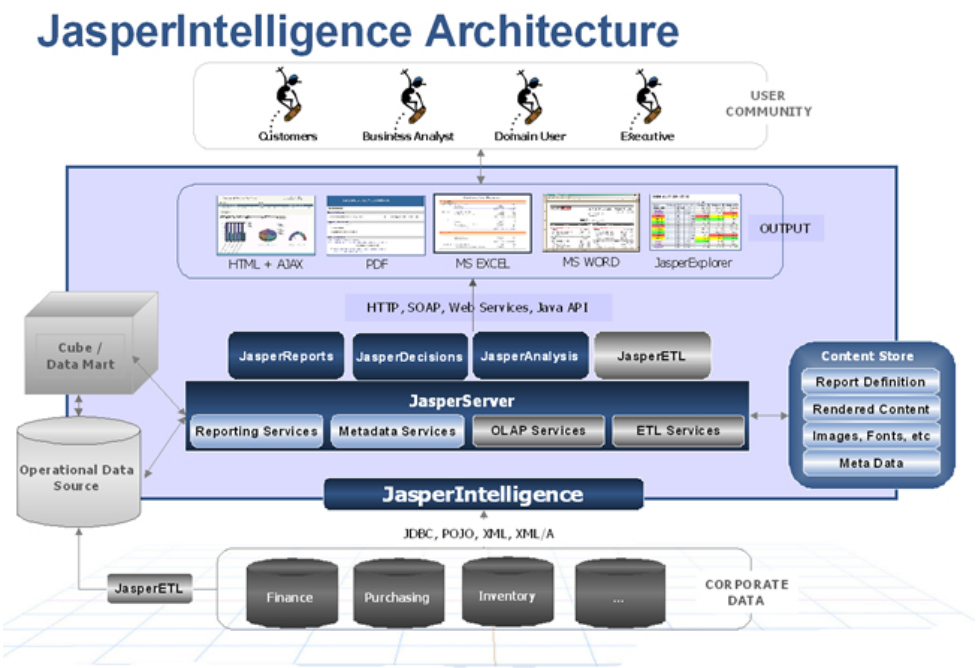


Schéma 20 – Architektura JasperSoft BI Suite, převzato od výrobce

Datová část

V datové části je použit ETL nástroj JasperETL[13], pro uložení dat je použit MySQL[14]. V komerční verzi lze použít také databázové stroje Oracle nebo Microsoft SQL Server.

Analytická část

Analytická část obsahuje JasperServer - interaktivní reportovací server, který nabízí správu uživatelů, zabezpečení, plánování a ukládání reportů. JasperAnalyst , provádějící analýzu dat/OLAP, založená na serveru Mondrian[15]. Od verze 2.1 jsou JasperServer a JasperAnalyst sloučeny do jedné komponenty.

Reportovací část

Prezenční část obsahuje JasperReport, nabízející tvorbu plánovaných, interaktivních a ad-hoc reportů. Výstupem tak mohou být soubory ve formátu CSV, DOC, RTF, TXT, XML, XLS a HTML včetně podpory technologie AJAX. JasperSoft BI Suite také umožňuje uložení vytvořených reportů pro další zpracování, tvorba Reportů je prováděna v JasperStudio, který je založen na iReport.

Data miningový nástroj není součástí tohoto balíku, lze však integrovat například produkt Weka.

Jedná se o J2EE aplikaci, pro běh JasperSoft BI Suite tak lze použít aplikační server Apache Tomcat nebo JBoss, v komerční verzi také IBM WebSphere. Přístup uživatelů je řešen pomocí tenkého klienta internetového prohlížeče. Podpora

13 Kapitola 4 ETL/ELT nástroje (Talend Open Studio)
14 Kapitola 5 Datové sklady
15 Kapitola 6 OLAP nástroje

operačních systémů v open source verzi je omezena na Linux a Windows, v komerčí verzi pak doplněna o MacOS, Sun Solaris, HP-UX a další. Od verze 2.1 je podporován standard JSR-168, což umožňuje práci s portlety a tak i integraci do dalších systémů. Nasazení tohoto produktu je vhodné do malých a středních společností.

9.4 Pentaho Open BI Suite

Verze:	**1.6**
Programovací jazyk:	**Java**
Operační systém:	**Linux, Windows, MacOS**
Datové zdroje:	**JDBC, XML**
ETL/ELT:	**Pentaho Data Integration (KETTLE)**
Data mining	**Pentaho Data Mining (Weka)**
OLAP:	**Pentaho Analyst (Mondrian/jPivot)**
Reportovací nástroje:	**Pentaho Reporting (jFreeReport)**
Licence:	**MPL**
Výrobce:	**Pentaho Corp.**
Odkazy:	**http://www.pentaho.com**

Pentaho Open BI Suite je nejkomplexnějším balíkem BI nástrojů v tomto srovnání. Vedle ETL, OLAP analýzy, reportů a data miningu nabízí například přímé propojení s mapovými podklady (Google Maps), výstrahy zasílané přes RSS, správu analytických procesů nebo podporu workflow analytických procesů. Dokumentace je velmi precizně zpracovaná, obsahuje vedle uživatelských, programátorských a instalačních příruček také návody krok za krokem pro specifické problémy a to i pro jednotlivé komponenty tohoto balíku. Součástí je také podpora formou internetových diskuzních skupin nebo FAQ. Další výhodou je také možnost instalace plně funkční zkušební verze včetně všech potřebných komponent jako je přednastavený aplikační server JBoss a v neposlední řadě i kvalitního návodu instalace této verze.

Architektura

Pentaho Open BI Suite je založen na architektuře využívající aplikačního serveru a tenkého klienta (internetový prohlížeč). Architektura je rozdělena do tří základních vrstev: vrstva datové a aplikační integrace, obsahující nástroje pro správu metadat, ETL (Pentaho Data Integration - KETTLE[16]) a integraci podnikových informací. Druhou vrstvou je platforma Business Intelligence, obsahující správu zabezpečení, administraci systému, podnikovou logiku a správu úložišť. Mezivrstva mezi vrstvou platformy a prezenční vrstvou obsahuje reportovací nástroje (Pentaho Reporting - jFreeReport[17]), nástroje analýzy (Pentaho Analyst – Mondrian/jPivot[18]) a data miningu (Pentaho Data Mining - Weka[19]), dashboardy a správu procesů. Prezenční vrstva podporuje komunikaci s uživatelem pomocí internetového prohlížeče, elektronickou poštou, portálem, exportem do kancelářských aplikací nebo komunakci pomocí webových služeb. Architektura Pentaho Open BI Suite je na schématu 21.

Pro běh Pentaho Open BI Suite je zapotřebí libovolný J2EE aplikační server, doporučovan je Jboss, ale lze použít i Apache, WebSphere, WebLogic a Oracle AS. Pro integraci do dalších systémů lze využít portletů (JSR-168). Pentaho Open BI Suite lze spustit na operačních systémech Linux, Windows nebo MacOS. Balík open source nástrojů od Pentaho je nejkomplexnějším balíkem v tomto srovnání a jeho nasazení je vhodné do malých a středních společností.

16 Kapitola 4 ETL/ELT nástroje
17 Kapitola 8 Reportovací nástroje
18 Kapitola 6 OLAP nástroje
19 Kapitola 7 Dolování dat

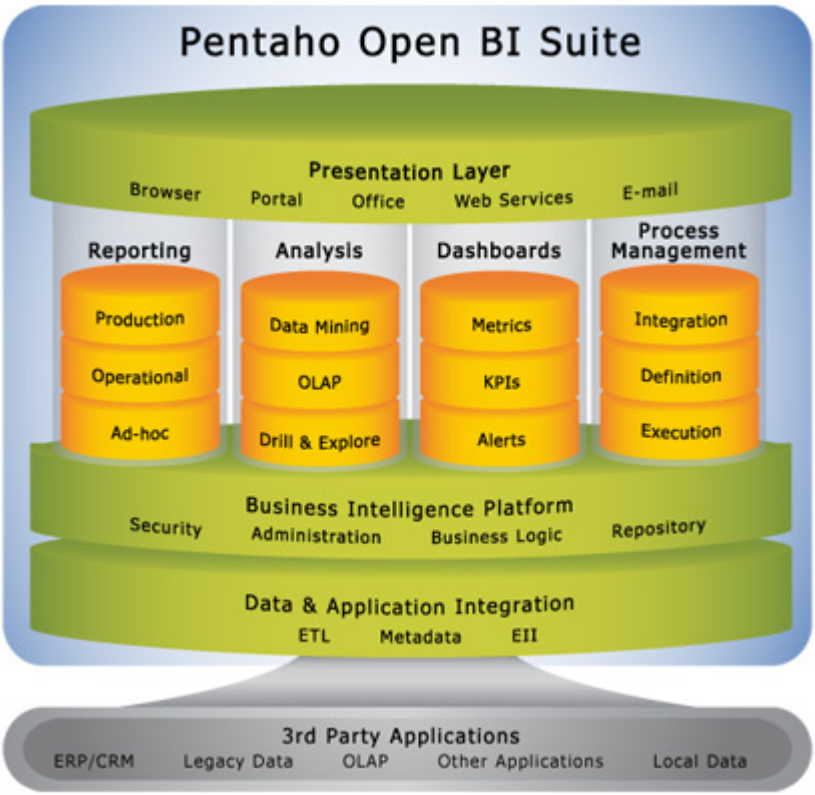


Schéma 21 – Architektura Pentaho Open BI Suite, , převzato od výrobce

9.5 Hodnocení

Kritéria pro hodnocení komplexních balíků nástrojů jsou:

- Dokumentace a podpora (Dokumentace), především pak její kvalita a obsáhlost, 30%.
- Datové zdroje (Zdroje), kde je sledováno propojení pomocí JDBC a případně dalších, 10%.
- Uživatelské rozhraní (Rozhraní), kde je posuzována kvalita

uživatelského rozhraní, 30%

- Obsažené komponenty (Komponenty), kde je hodnoceno obsažení základních a případně dalších komponent balíků (propojení s mapami, administrace apod.), 20%.
- Operační systém (Operační systém), kde je sledován počet podporovaných operačních systémů, 10%.

Ve srovnání se tak na pomyslné první místo dostává Pentaho BI Suite, který vyniká oproti ostatním ve všech sledovaných kritériích. Je následován JasperSoft BI Suite, který má nedostatek především v absenci nástroje pro dolování dat, což však lze řešit integrací dataminingových nástrojů Weka nebo RapidMiner. Oba tyto nástroje mohou konkurovat i komerčně nabízeným produktům. Další dva produkty (OpenI a SpagoBI) tak zaostávají v horším uživatelském prostředí a OpenI také v horší dokumentaci.

10 Navrhovaný systém

Pro představení možností open source manažerských informačních systémů byl vybrán MIS pro menší stavební firmu. V této kapitole bude nejprve představen segment trhu, ve kterém se společnost nachází, představení společnosti, požadavky na MIS a výběr vhodného open source základu pro MIS. Navržený mamažerský informační systém bude určen pro stavební výrobu ve společnosti.

10.1 Trh ve stavebnictví

Stavební výroba v České republice zažívá v současnosti velký růst, který byl ovlivněn hned několika faktory. Prvním bylo snížení úrokové míry a tím i snížení cen půjček a hypoték poskytovaných občanům. Dalším významným faktorem, který v posledních letech zvyšoval poptávku po výstavbě rodinných domů a bytů bylo riziko zvýšení DPH pro tento segment trhu ze snížené sazby 5% na základní sazbu 19% k 1. 1. 2008. Díky prodloužení vyjímky Evropské komise se tak až do roku 2010 může uplatňovat snížená sazba, která se od příštího roku zvyšuje na 9%. Posledním významným faktorem růstu byla také poslední zima na přelomu let 2006 a 2007. Počasí je ostatně faktorem, který ovlivňuje většinu stavebních prací v průběhu celého roku.

Produktem většiny menších i větších stavebních firem je služba, neboť je po stránce řešen smlouvou o dílo. Dalším specifikem je skutečnost, že každá provedená stavební činnost je svým způsobem unikátní. Vždy závisí na použitých technologiích, prostředí, ročním období nebo dadatečných změnách

inicializovaných zákazníkem nebo jinými okolnosti. Tato skutečnost se projevuje i při realizaci katalogových domů nebo bytových jednotek. Z tohoto důvodu je přímo srovnávat provádění konkrétních stavebních činností mezi sebou.

10.2 Specifikace společnosti

Společnost se vyvinula postupně z malého rodinného podniku s několika zaměstnanci v dobře fungující a rostoucí stavební firmu. Vlastníci společnosti si tak uvědomují potřebu efektivnějšího řízení společnosti a to i s využitím informačních technologií. Ve společnosti prozatím neběží podnikové informační systémy, jejíž dat lze pro MIS využít, ale s jejich nástupem se v nejbližší době počítá. Navržený MIS tak nabízí možnosti manažerského informačního systému a je připravený při spuštění zdrojových systémů a nastavení takto získaných datových zdrojů pro nasazení. Jako zdroj dat bude sloužit databáze postavená na současné evidenci.

Společnost v současnosti zaměstnává okolo dvaceti zaměstnanců, další lidské zdroje jsou řešeny smluvně formou subdodávek. Společnost sice vlastní skladovací prostory, jsou však primárně využívány jako prostory pro skladování pracovních strojů, skladovaní stavebního materiálu je zde pouze nárazové.

Předmětem podnikání vybrané společnosti je:

- provádění staveb, jejich změn a odstraňování
- zprostředkovatelská činnost v oblasti obchodu, služeb a výroby
- koupě zboží za účelem jeho dalšího prodeje a prodej

Ve společnosti jsou vlastníky a zaměstnanci vykonávány role, popsané v následující tabulce 6.

Role	**Popis**
Manažer	Vlastník společnosti, řízení společnosti, monitoring
Rozpočtář	Příprava rozpočtu, oceňování jednotlivých etap výstavby
Přípravář	Plánování etap stavby (přidělování lidských zdrojů, materiálů, strojů a subdodávek)
Stavbyvedoucí	řízení staveb, kontakt s dodavateli a pracovníky na místě stavby
Stavební dozor	dozorování staveb, potvrzování jednotlivých etap a bodů, porovnávání rozestavěnosti jednotlivých staveb s harmonogrmy.
Účetní	vedení účetnictví společnosti
Dělník	další členění dle profesí: bagrista, řidič, elektrikář, přidavač, obkladač, tesař, zedník a další

Tabulka 6 – Role ve společnosti

10.2.1 Hierarchie

Hierarchie těchto rolí včetně přístupů do podnikových a manažerských informačních systémů je znázorněna na schématu 22. Manažer zde má přístup jak k MIS, tak i k podnikovým systémům. Je to důsledek požadavků na monitoring a na možnost zasahovat do řízení podniku. Do podnikových systémů mají dále přístup zaměstnanci v rolích Stavbyvedoucí, Rozpočtář, Přípravář, Stavební dozor a Účetní. Zaměstnanci v roli Dělník nemají přístup do žádného z podnikových systémů, neboť se zde nedá předpokládat přínos jak pro společnost tak i pro tyto zaměstnance samotné.

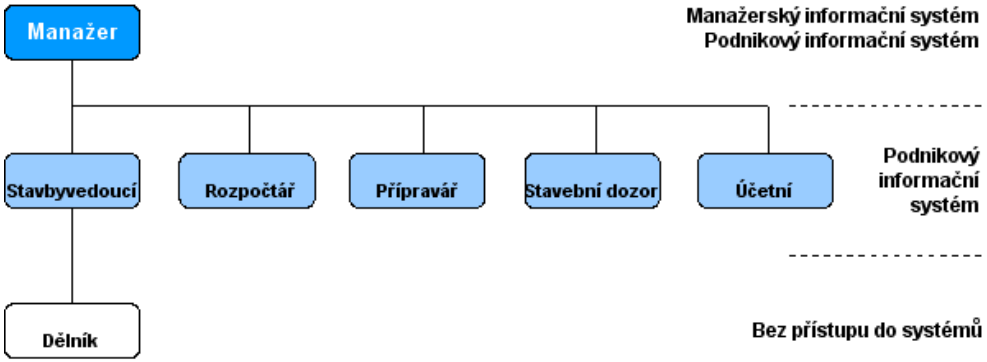


Schéma 22 – Hierarchie ve společnosti

10.2.2 Evidence dat

Ve společnosti jsou evidovány následující údaje, které mohou sloužit jako zdroje dat pro manažerský informační systém. Seznam evidovaných údajů včetně krátkého popisu a využití pro navrhovaný MIS je v tabulce 7.

Evidence	Popis	Využito
Zakázky	Evidence zakázek včetně harmonogramů etap, přidělení strojů, materialu, lidských zdrojů, subdodávek, prostavěnosti, zisku/ztráty	ANO
Účetnictví	Evidence účetnictví společnosti	NE
Zaměstnanci	Evidence zaměstnanců	ANO
Stroje	Evidence strojů včetně automobilů	ANO
Materiál	Evidence objednaného a dodaného materialu	ANO
Dodavatelé	Evidence dodavatelů materiálu, lidských zdrojů, strojů, subdodávek	ANO
Zákazníci	Evidence zákazníků společnosti	ANO
Reklamace	Evidence reklamací	ANO
Revize	Evidence revizí a kontrol strojů a automobilů	ANO
Nabídky	Evidence nabídek na zakázku	NE

Tabulka 7 – Evidované údaje

Zakázky

Evidence zakázek obsahuje údaje o zakázce, jednotlivých etapách zakázky, potřebném materiálu, strojích a zaměstnanců pro tyto etapy. Dále je zde evidován stav zakázky a etapy, dodavatelé a odběratelé. Z těchto zdrojů budeme získávat informace pro MIS.

Účetnictví

Účetní evidence společnosti obsahuje účetnictví, daňovou evidenci, objednávky, fakturace, evidenci majetku a skladů. Tato evidence nebude prozatím implementována do MIS z důvodu její současné nevyužitelnosti.

Zaměstnanci

Evidence zaměstnanců a externích spolupracovníků včetně jejich profesních znalostí a přiřazení k jednotlivým etapám realizace zakázek. Tato evidence může sloužit k zobrazení aktuálního stavu vytíženosti zaměstnanců.

Stroje a Revize

Evidence strojů a automobilů včetně jejich plánovaných revizí. Data zde uložená mohou sloužit jako podklad pro varování před možným překročením lhůty pro revizi případně Státní technickou kontrolu.

Dodavatelé a Zákazníci

Evidence dodavatelů včetně oboru podnikání a hodnocení jeho kvalit. Pro evidenci iniciálů se využívá společné databáze subjektů, kde jsou vedle dodavatelských iniciálů vedeny i iniciály odběratelů – zákazníků.

Reklamace

Evidence reklamací zakázek může sloužit jako zdroj dat pro hodnocení kvality provedených prací a pro zjištění aktuálního stavu reklamace.

10.3 Požadavky na MIS

V následující tabulce 8 jsou vyjádřeny prvotní požadavky na manažerský informační systém. Jsou doplněny prioritou jednotlivých požadavků v rozmení 0 (žádná) až 3 (nejvyšší). Všechny tyto požadavky byly formulovány s manažerem společnosti.

Požadavek	Popis	Význam	Priorita
1	Přehled probíhajících zakázek a etap	3	3
2	Přehled volných pracovních sil včetně profesí.	3	2
3	Přehled stavu materiálu	3	2
4	Analýza zakázek a etap	2	3
5	Přehled revizí strojů	2	2
6	Přehled průběhu reklamací	2	2
7	Analýza vývoje cen materiálu	2	1

Tabulka 8 – Požadavky na systém

- Požadavek 1 – Přehled probíhajících zakázek a etap slouží k zjištění aktuálního stavu, případného zpoždění, chybějícího materiálu nebo profesí a zjištění jejich stavu (přijatá, řešená, pozastavená, dokončená, zrušená).
- Požadavek 2 – Přehled volných pracovníků včetně profesí pro jejich případné přidělení na etapy, kde hrozí zpoždění**.**
- Požadavek 3 – Přehled dodaného materiálu (Dodáno vše, část, nic) pro zjištění stavu materiálové připravenosti k realizaci etapy.

- Požadavek 4 –Analýza zakázek a etap formou pro zjištění nákladů, zisku a fakturované ceny pomocí OLAP kostky.
- Požadavek 5 – Přehled blížících se revizí strojů a automobilů.
- Požadavek 6 – Přehled průběhu reklamací dle jejich stavu (přijatá, řešená, pozastavená, dokončená, zrušená).
- Požadavek 7 – Analýza vývoje cen materiálu dle dodavatelů.

10.4 Model Datového skladu

Datový sklad bude realizován ve schématu hvězdy, kde tabulka faktů (Etapa_Fakt) obsahuje náklady(naklady) a fakturovanou cenu (cena). Dimenzemi v tomto datovém skladu jsou:

- Region (Město, Okres, Kraj, Stát) realizace,
- Region (Město, Okres, Kraj, Stát) zákazníka,
- Zákazník (Název),
- Čas (Měsíc, Kvartál, Rok) zahájení,
- Čas (Měsíc, Kvartál, Rok) dokončení,
- Stav (Započato, Probíhá, Přerušeno, Dokončeno, Zrušeno) etapy,
- Zakázka (Etapa, Zakázka)
- Typ Etapy (Terénní úpravy, Betonáž, Zednické práce, Krovy, Instalatérství, Okna a Dveře, Podlahy, Obkady, Úpravy, Bourání, Úklid)

Pro dimenzionální analýzu použijeme MDS (Multidimensional Domain Structure) od Erika Thomsena, která je označována jako jeden ze základních přístupů k modelování dimenzí. Modely dimenzí jsou na schématu 23.

Schéma 23 – Dimenze

Ačkoli již okresy jako územněprávní neexistují, území jimi vyznačené představuje významné členění obzvláště pro organizace, působící v rámci kraje či na menším území, proto byly zařazeny do dimenzí regionu.

Tabulka faktů obsahuje náklady na etapu (naklady) a účtovanou cenu (cena). Na základě těchto ukazatelů lze také určit zisk, případně jeho procentuální hodnotu.

10.5 Architektura

Architektura navrhovaného systému je uvedena na schématu 24. MIS bude získávat data jak z vytvořeného datového skladu, tak i přímo z produkčních systémů. Přístup do produkčních systémů je zde umožněn díky malému zatížení, neboť zde probíhá malý počet transakcí – řádově sto denně. Přesun dat do datového

skladu by tak znamenalo zbytečnou redundanci. V případě výrazného zvýšení zátěže však musí dojít k přesouvání analyzovaných dat z produkčních databází do datového skladu. Uživatel manažerského informačního systému zde bude komunikovat pouze s MIS.

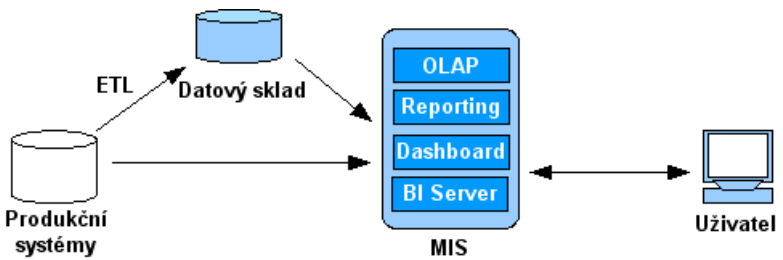


Schéma 24 – Architektura navrhovaného MIS

10.5.1 Databáze

Jako databázový stroj pro produkční systémy a datový sklad je použit MySQL 5.0 Community Edition. V přiložené zkušební verzi obsahuje mj. databáze JSZDROJ (simulující produkční systémy společnosti) a JSMIS (datový sklad). MySQL 5.0 obsahuje několik typů databázových strojů. Pro analytické účely je vhodný MyISAM, který je použit i pro JSMIS, pro transakční systémy pak InnoDB, který je použit pro JSZDROJ. Pro přístup do databází je použito JDBC.

Produkční systémy

Tato databáze představuje zdoj dat, který je dále pomocí ETL nástrojů transformován do datového skladu, či přímo přístupný z manažerského informačního systému. Modifikace těchto dat není pro MIS umožněna. Jako

vzorová databáze pro navrhovaný produkt zde byla použita databáze JSZDROJ. Model této databáze je v příloze 2.1.

Datový sklad

V datovém skladu jsou data uložena pomocí schématu hvězdy, tato data slouží pro OLAP analýzu, mamažerský pult a tvorbu některých reportů. Tuto databázi MIS pouze čte, o naplnění daty se stará ETL nástroj. Vzorovou databází je JSMIS. Zjednodušený model této databáze je na schématu 25.

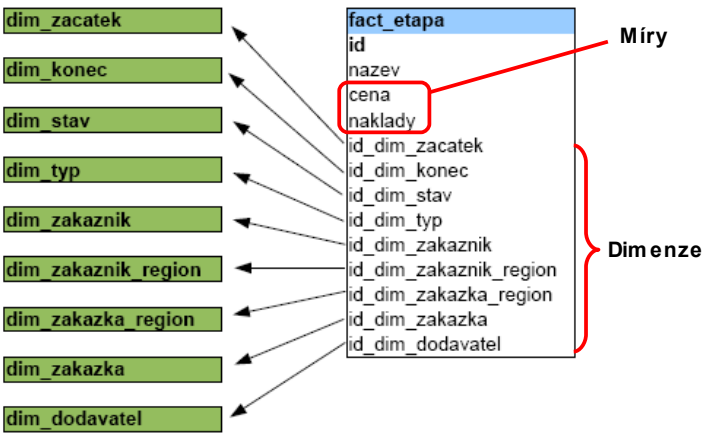


Schéma 25 – Zjednodušený model JSMIS

Pro tabulku faktů zde byla použita tabulka fact_etapa, která vychází z tabulky etapa v databázi JSZDROJ.

Datový sklad je řešen podle R. Kimballa, tedy metodou postupného budování datových tržišť, kde se z důvodu malého zatížení produkčních systémů nevyužívá dočasného úložiště, které je nahrazeno přímým přístupem do databází produkčních systémů.

10.5.2 ETL

Pro datovou transformaci mezi produkčnímy systémy a datovým skladem je použito Pentaho Data Integration (KETTLE) verze 3.0.1. Oproti verzi 2.5 se liší především v grafickém rozhraní, které je zde přehlednější. Komponenta pro datovou transformaci (Pan) je spouštěna dávkovým souborem, jako parametr je zde soubor ve formátu XML, obsahující instrukce pro tuto transformaci. Tento soubor tak bude spouštěn v denních intervalech ve vhodně nastavených hodinách (nejlépe nočních). Schéma datové transformace je zobrazeno v příloze 2.2, konfigurační soubor je přiložen na CD.

10.5.3 MIS

Jako výchozí část pro tvorbu manažerského informačního systému byl vybrán Pentaho BI Suite 1.6, obsahující především Pentaho BI server, pro webový server je použito JBoss 2.6.1. Tento balík má velkou výhodu ve snadné instalaci a následné modifikaci. Instalace Pentaho BI Suite 1.6, MySQL5.0 a běhového prostředí je uvedena v instalační příručce na přiloženém CD.

Vybraným operačním systémem pro MIS je Microsoft Windows XP Profesional SP2, neboť se jedná o operační systém používaný na všech ostatních pracovních stanicích ve společnosti. Tímto krokem se tak nenaruší homogenita prostředí a tím nezvýší nároky na správce podnikové sítě, což může do budoucna zvýšit náklady na vlastnění produktu (TCO). Tento operační systém umožňuje současný přístup až deseti uživatelům (pro více uživatelů je nutný již serverový produkt, např. Microsoft Windows Server), což je i do budoucna plně dostačující.

10.5.4 OLAP

Pro OLAP server Mondrian 2.4 je v tomto MIS nastavena OLAP kostka JSCube, její konfigurační soubor je přiložen na CD. Pro vytváření nových OLAP kostek je určen Pentaho Cube Designer, který poměrně snadným způsobem pomocí průvodce umožňuje vytváření nových OLAP kostek. Definice kostky je uložena v souboru ve formátu XML.

10.5.5 Reporting

Pro reportovací služby je zde využito Pentaho Classic Reporting, pro modifikaci a vytváření nových reportovacích souborů uložených ve formátu XML, lze použít Pentaho Report Designer 1.6, který umožňuje práci metodou WYSIWYG. Reporty lze také vytvářet pomocí průvodce.

10.5.6 Dashboard

Manažerský pult je zde vytvořen pomocí knihoven JFree, které jsou součástí instalace a jsou využívány například i k vytváření grafů v reportech. Jsou zde shromážděny funkce monitoringu probíhajících staveb, volných kapacit či blížící se revize.

10.5.7 Uživatel

Uživatelem systému tak bude pouze osoba v roli Manažer, tedy vlastník společnosti. Pro přístup k manažerskému informačnímu systému bude používat tenkého klienta ve formě internetového prohlížeče, čímž je umožněno flexibilnější užívání MIS, které není vázané na pracovní stanici. K MIS přistupuje pomocí

protololu HTML.

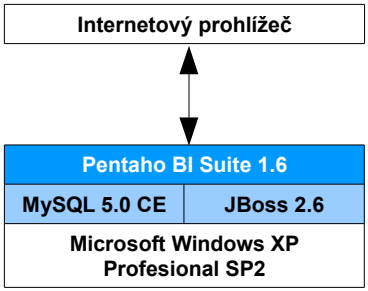


Schéma 26 – Architektura aplikací

10.6 Realizace

Pro realizaci navrhovaného MIS byl zvolen model inkrementálního životního cyklu, který je vzhledem k charakteru realizace nejvýhodnější. Důvodem je skutečnost, že uživatel doposud žádný MIS ve společnosti doposud nepoužíval a není tak na začátku schopen plně definovat své požadavky. Může tak také v co nejkratší době začít používat vlastní MIS. Dalšími požadavky lze později manažerský informační systém rozšiřovat o nově potřebné komponenty, případně ho zpřístupnit i dalším uživatelům – zaměstnancům. MIS přiložený na CD je tak funkční verzí, která je vytvořena na základě prvotních požadavků, provozovanou na vzorové databázi.

Přístup k realizaci spočíval v nastavení potřebných parametrů pro prováděné analýzy, reporty, manažerský pult a lokalizace často používaných komponent. Nejedná se tak o vývoj či modifikaci binárního kódu produktu, ale o jeho nastavení a přizpůsobení požadavkům. Pentaho BI Suite nabízí toto nastavení pomocí jednoduchých konfiguračních souborů nebo souborů ve formátu XML. Pro tvorbu těchto souborů je použito nástavby pro Eclipse – Pentaho Design Studio, kde pomocí předem připravených formulářů probíhá nastavení. Pro vytvoření MDX

souboru OLAP kostky je vhodné použít Pentaho Cube Designer, který však není v konečné verzi, což se projevuje na jeho stabilitě a částečně omezené funkčnosti. Všechny tyto konfigurační nástroje jsou spouštěné přímo na počítači, na kterém MIS běží.

10.7 Navrhovaný systém

Na přiloženém CD je navržený manažerský informační systém včetně potřebných komponent pro jeho spuštění. Součástí je také krátká instalační a uživatelská příručka, obsahující postup instalace, spuštění, přehled navrženého MIS, jeho funkcí a odkazy na plnou dokumentaci výrobce.

Přidání nových součástí systému je možné pomocí konfiguračních nástrojů jako je Cube Designer, Report Designer nebo Design Studio. Přidání nové části (např. OLAP kostka nebo report) nebo její modifikace spočívá ve vytvoření popř. úpravě .xaction souboru ve formátu XML s konfigurací, případně dalšího souboru ve formátu XML (např. specifikace reporu). Systém pro organizaci složek přejímá adresářovou strukturu, lze tak snadno přidat novou složku vytvořením adresáře a přidáním konfiguračního souboru index.xml. Všechny konfigurační nástroje jsou ke stažení včetně dokumentace ze stránek výrobce.

10.8 Licence

Pentaho BI Suite 1.6 a modifikace jsou distribuovány pod MPL 1.1 nebo vyšší licencí, JBoss 2.6.1 je distribuován pod LGPL licencí, MySQL 5.0 CE pod GPL licencí, Java Runtime Enviroment 1.5 pod freeware licencí. Text licence je obsažen na přiloženém CD.

11 Závěr

Realizace manažerského informačního systému pomocí open source je alternativou pro komerčně distribuované produkty. Ačkoli náklady na pořízení systému jsou minimální, konfigurace a následná správa takového systému se může být ve výsledku dražší než komerční verze. Toto riziko je si tak nutné uvědomit při výběru komponent nebo celého systému. Na druhou stranu nabízí bezpečné řešení, plně konfigurovatelné pro potřeby společnosti, což komerční verze mnohdy nemohou nabídnout.

Pentaho BI Suite je rozsáhlým produktem, nabízející oproti svým konkurentům velmi kvalitní dokumentaci ke každé komponentě, což značně snižuje čas pro implementaci. Jednotlivé komponenty lze s dodatečným úsilím nahradit jinými, které více vyhovují. Nevýhodou je však fakt, že velká část komponent se musí konfigurovat samostatně, neboť Pentaho BI Suite je spojením několika základních komponent přes BI server. Nastavení korektního zobrazení české znakové sady se tak musí provádět hned na několika místech. Ačkoli BI server zobrazuje stránky v UTF-8, některé komponenty (např. Report Designer) však pracují s ISO-8859-1, které neobsahuje plně českou znakovou sadu. Obdobný problém je i s nastavením spojení s databází, který se musí konfigurovat jak pro JBoss tak i pro vlastní Pentaho BI server. Jsou to drobné problémy, které však stojí polovinu času vývoje.

Během realizace MIS nedošlo k využití všech částí Pentaho BI Suite, neboť se jedná o velmi rozsáhlý systém. Byla zde provedena konfigurace potřebných částí, které jsou potřebné pro splnění požadavků. Pentaho BI Suite je však dobrým základem, neboť lze do budoucna očekávat rozšíření požadavků na manažerský

informační systém (např. upravit chování podle uživatelských rolí), což toto řešení plně umožňuje. Přes drobné nedostatky je Pentaho BI Suite nejobsáhlejším open source produktem na trhu a spolu s kvalitní dokumentací a komponentami je nejvhodnějším řešením pro segment malých a středních společností.

12 Seznam použité literatury

[1] Montecheuil, Dupupet: Third Generation ETL: Delivering Best Performace, 2006, Sunopsis (White Paper)

[2] Seige, Viktor a kol.: Příručka manažera IV. - Business Intelligence, 2007, Tate International s.r.o

[3] Voříšek, Jiří: Strategické řízení informačního systému a systémová integrace, 2003, Management Press, ISBN 80-85943-40-9

[4] Novotný, Pour, Slanský: Business Intelligence – Jak využít bohatství ve vašich datech, 2005, Grada Publishing, ISBN 80-247-1094-3

[5] Wikipedia.org, citováno 1.1.2008, www.wikipedia.org

1. Extract, Transform, Load: http://en.wikipedia.org/extract_transform_load
2. Online analytical processing: http://en.wikipedia.org/wiki/Online_analytical_processing
3. Data warehouse: http://en.wikipedia.org/wiki/Data_warehouse
4. Third normal form: http://en.wikipedia.org/wiki/third_normal_form
5. MDX: http://en.wikipedia.org/wiki/Multidimensional_Expressions
6. XMLA: http://en.wikipedia.org/wiki/XMLA
7. WEKA: http://en.wikipedia.org/wiki/Weka_(machine_learning)
8. Competitive Intelligence: http://en.wikipedia.org/wiki/Competitive_Intelligence
9. Business Intelligence: http://en.wikipedia.org/wiki/Business_intelligence

[6] DWReview.com http://www.dwreview.com

1. Introduction to OLAP: http://www.dwreview.com/OLAP/Introduction_OLAP.html

[7] Berson, Smith: Data Warehousing, Data Mining & OLAP, 1997, McGraw-Hill Inc., ISBN 0-07-006272-2

[8] Pospíšil, Nemrava: Dolování dat a jeho aplikace, 2006, http://axpsu.fpf.slu.cz/~sos10um/trendy/DM.pdf

[9] Oral, Tomáš: Manažerské informační systémy a jejich úloha v řízení podniku, 2006, FI MU, diplomová práce

[10] XMLA: http://www.xmla.org

[11] Lacko, Ľuboslav : Business Intelligence v SQL Serveri 2005, 2006, Microsoft .r.o.

[12] CRISP-DM: http://www.crisp-dm.org

[13] Zpravodajský portál časopis IT systems, Systém On Line: http://www.systemonline.cz

[14] Witten, Frank: Data Mining: Practical machine learning tools and techniques, 2nd Edition, 2005, Morgan Kaufmann, San Francisco

[15] R Development Core Team: R: A Language and Environment for Statistical Computing, 2007, R Foundation for Statistical Computing, ISBN 3-900051-07-0

[16] Demsar, Zupan, Leban: Orange: From Experimental Machine Learning to Interactive Data Mining, 2004 White Paper, Faculty of Computer and Information Science, University of Ljubljana

[17] D. J. Power: A Brief History of Decision Support Systems, 2007, DSSResources.COM, http://dssresources.com/history/dsshistory.html

[18] Česká společnost pro systémovou integraci: http://www.cssi.cz

[19] Král, Jaroslav: Informační systémy: specifikace, realizace, provoz, 1998, Science, ISBN 80-86083004

[20] Štědroň, Bohumír: Manažerské řízení a informační technologie, 2007, Grada Publishing, ISBN 978-248-2052-4

[21] Kroenke, David: Management Information Systems, 1992, McGraw-Hill,

ISBN-0-07-035787-0

[22] Berka, Petr: Dobývání znalostí z databází, 2003, Academica, ISBN-80-200-1062-9

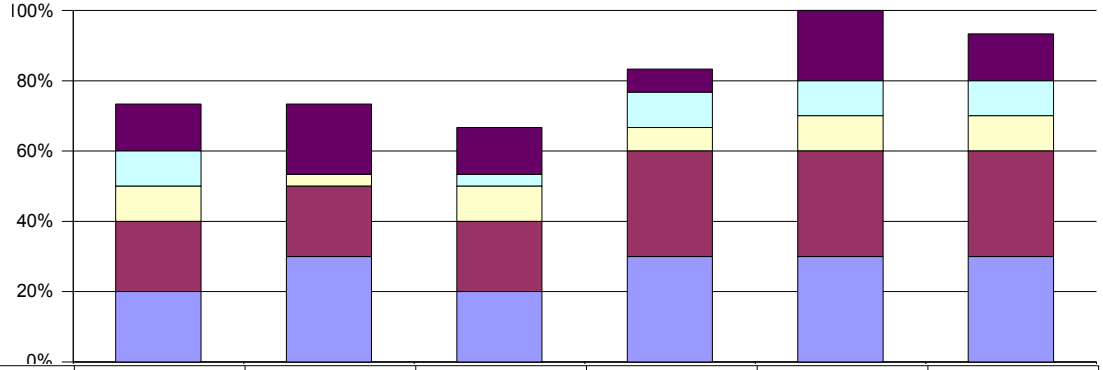
[23] Rud, Olivia Parr: Data Mining, 2001, Computer Press, ISBN-80-7226-577-6

[24] Arlow, Neustadt: UML2 a unifikovaný proces vývoje aplikací, 2007, Computer Press, ISBN-978-80-251-1503-9

[25] Portál Společnosti pro výzkum a podporu Open source: http://www.oss.cz

[26] Šilberský J.: Manažerské informační systémy, 2001, FI MU, diplomová práce.

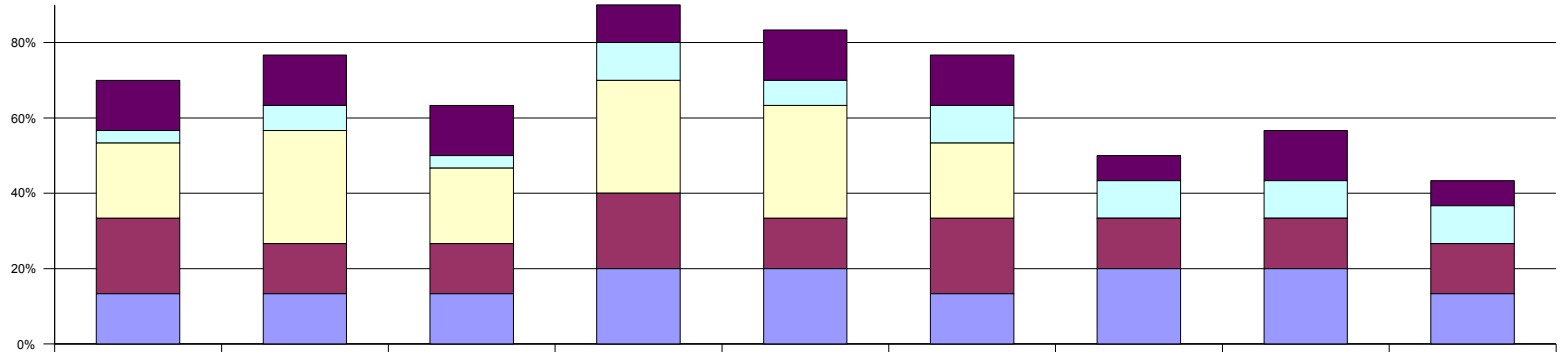
Příloha 1.1 – Srovnání ETL/ELT



Srovnání		Apatar	CloverETL	Octopus	KETL	KETTLE	Talend
Dokumentace	30%	2	3	2	3	3	3
Propojení	30%	2	2	2	3	3	3
Nastavení	10%	3	1	3	2	3	3
Plánování	10%	3	0	1	3	3	3
Operační systém	20%	2	3	2	1	3	2
Celkem		**73%**	**73%**	**67%**	**83%**	**100%**	**93%**

Hodnocení		3 body	2 body	1 bod	0 bodů	Barva
Dokumentace	30%	úplná, kvalitní	rozsáhlá	základní	žádná	
Propojení	30%	více	dvě	jedno	-	
Nastavení	10%	grafické, kvalitní	grafické	script/XML	žádné	
Plánování	10%	grafické, kvalitní	grafické	script	žádné	
Operační systém	20%	více systémů	dva systémy	jeden systém	-	

Příloha 1.2 – Srovnání Databázových systémů

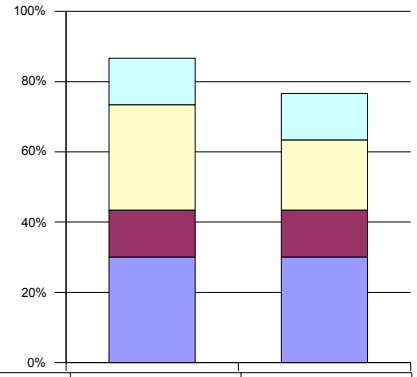


Srovnání		Derby	Firebird	HSQLDB	MySQL	PostgreSQL	IBM	MSSQL	Oracle	Sybase
Dokumentace	20%	2	2	2	3	3	2	3	3	2
Standarty	20%	3	2	2	3	2	3	2	2	2
Omezení	30%	2	3	2	3	3	2	0	0	0
Rozhraní	10%	1	2	1	3	2	3	3	3	3
Operační systém	20%	2	2	2	2	2	2	1	2	1
Celkem		**70%**	**77%**	**63%**	**93%**	**83%**	**77%**	**50%**	**57%**	**43%**

Hodnocení		3 body	2 body	1 bod	0 bodů	Barva
Dokumentace	20%	úplná,kvalitní,CZ	rozsáhlá	základní	žádná	
Stadardy	20%	více	SQL99/2003	SQL92	-	
Omezení	30%	-	-	-	-	
Rozhraní	10%	grafické, kvalitní	grafické	příkazový řádek	-	
Operační systém	20%	více systémů	dva systémy	jeden systém	-	

Omezení		- 0 bodů	- 1 bod	- 2 body
Velikost DB		bez omezení	≤ 10 GB	≤ 5 GB
Počet CPU		bez omezení	≤ 2 CPU	1 CPU
Velikost RAM		bez omezení	≤ 2 GB	≤ 1 GB

Příloha 1.3 – Srovnání OLAP

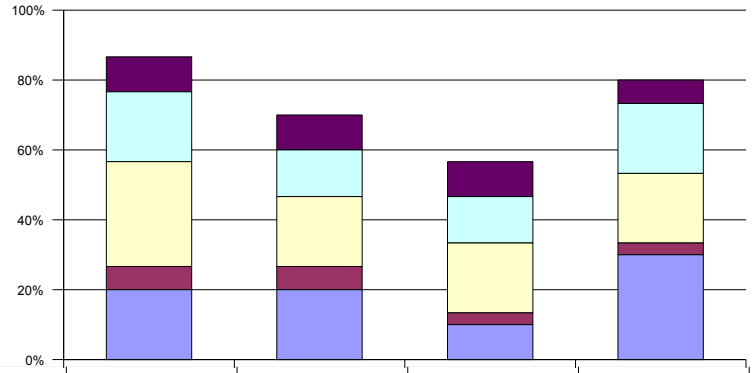


Srovnání		Mondrian	Palo
Dokumentace	30%	3	3
Propojení	20%	2	2
Klient	30%	3	2
Operační systém	20%	2	2

Celkem		**87%**	**77%**

Hodnocení		3 body	2 body	1 bod	0 bodů	Barva
Dokumentace	30%	úplná, kvalitní	rozsáhlá	základní	žádná	
Propojení	20%	další	XMLA	vlastní	-	
Klient	30%	grafický, web	grafický	-	-	
Operační systém	20%	více systémů	dva systémy	jeden systém	-	

Příloha 1.4 – Srovnání nástrojů pro dolování dat



Srovnání		WEKA	R-Project	Orange	RapidMiner
Dokumentace	30%	2	2	1	3
Propojení	10%	2	2	1	1
Rozhraní	30%	3	2	2	2
Funkce	20%	3	2	2	3
Operační systém	10%	3	3	3	2
Celkem		**87%**	**70%**	**57%**	**80%**

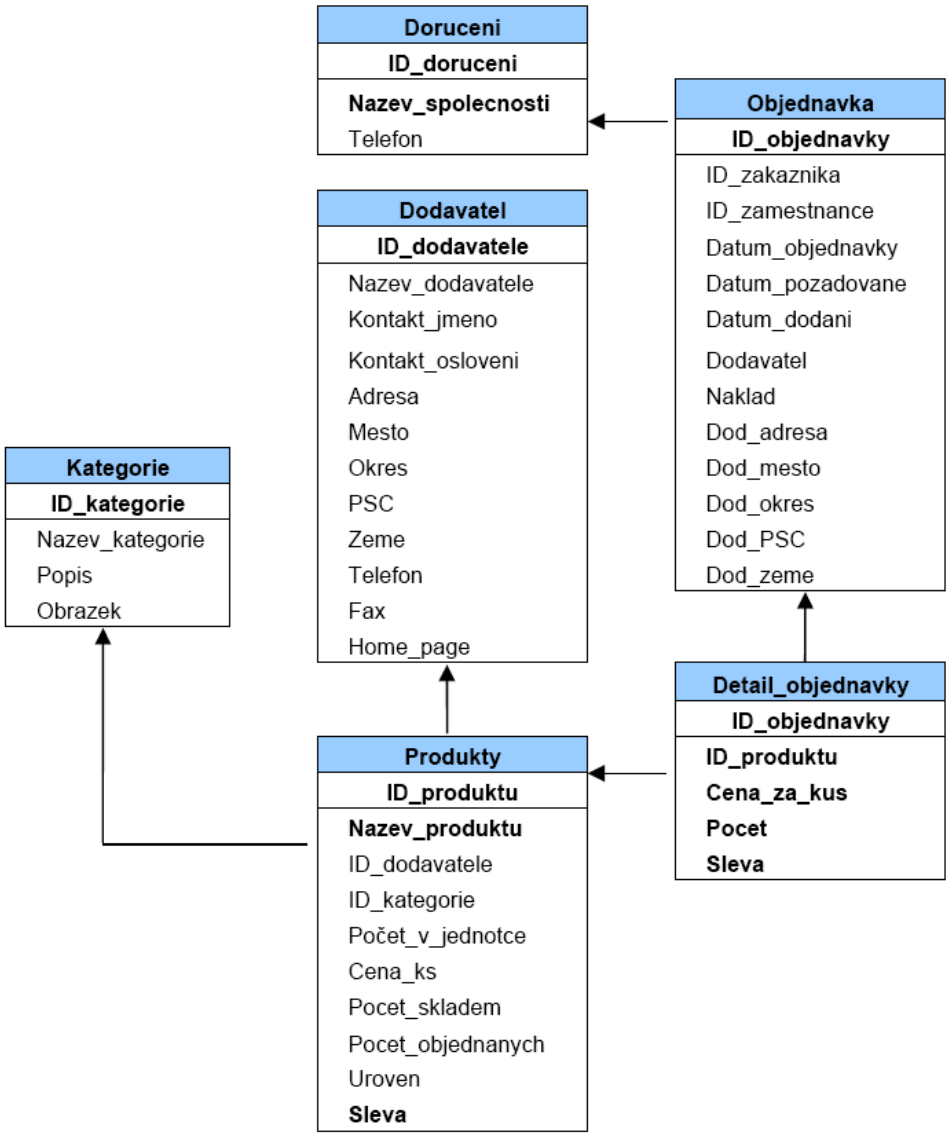
Hodnocení		3 body	2 body	1 bod	0 bodů	Barva
Dokumentace	30%	úplná, kvalitní	rozsáhlá	základní	žádná	
Propojení	10%	více	dvě	jedno / bez JDBC	-	
Rozhraní	30%	grafické, kvalitní	grafické	-	-	
Funkce	20%	více	statistické	-	-	
Operační systém	10%	více systémů	dva systémy	jeden systém	-	

Příloha 1.5 – Srovnání reportovacích nástrojů

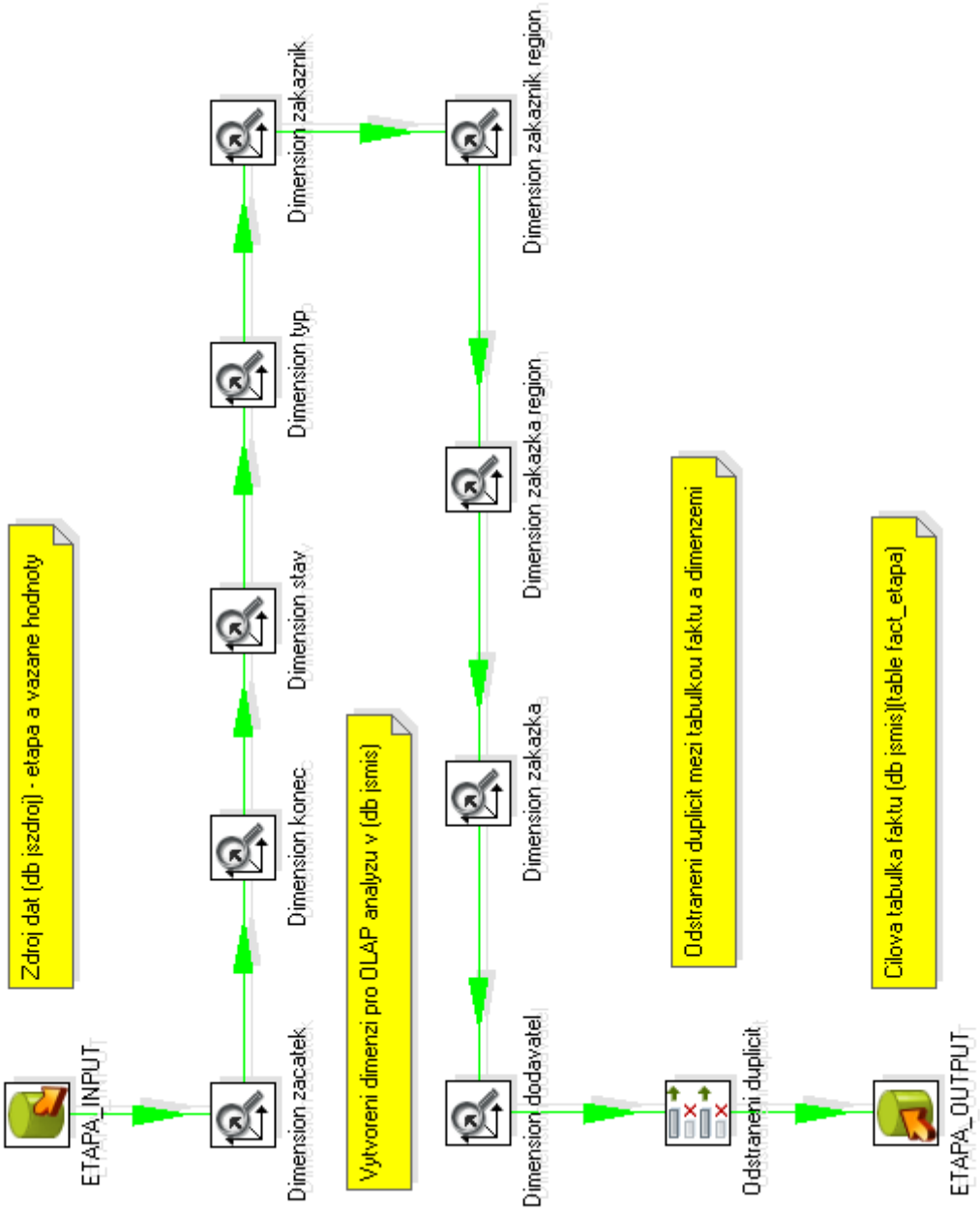
Srovnání		JfreeReport	DataVision	iReport	Eclipse BIRT
Dokumentace	20%	3	1	3	3
Zdroje	20%	2	2	1	3
Výstupy	20%	2	3	3	1
Rozhraní	20%	3	2	3	3
Operační systém	20%	2	2	2	2
Celkem		**80%**	**67%**	**80%**	**80%**

Hodnocení		3 body	2 body	1 bod	0 bodů	Barva
Dokumentace	20%	úplná,kvalitní	rozsáhlá	základní	žádná	
Zdroje	20%	více	JDBC, XML/HTML	JDBC	-	
Výstupy	20%	více	navíc XLS/DOC	PDF/XML	obrazovka/HTML	
Rozhraní	20%	grafické, kvalitní	grafické	-	-	
Operační systém	20%	více systémů	dva systémy	jeden systém	-	

Příloha 2.1 – Příklad OLTP databáze



Příloha 2.3 Zobrazení transformačního souboru



Příloha 2.2 Fyzický model JSZDROJ