

**Univerzita Pardubice**

**Fakulta ekonomicko-správní**

**Využití datových skladů v malém a středním podniku**

**Štěpán Jahoda**

**Bakalářská práce  
2016**

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Štěpán Jahoda**  
Osobní číslo: **E13173**  
Studijní program: **B6208 Ekonomika a management**  
Studijní obor: **Management podniku: Management malých a středních podniků**  
Název tématu: **Využití datových skladů v malém a středním podniku**  
Zadávací katedra: **Ústav matematiky a kvantitativních metod**

### Z á s a d y p r o v y p r a c o v á n í :

Cílem práce je popsat možnosti datových skladů (velikost, rozsah, frekvence ukládání) a jejich využití k analýze dat v malém a středním podniku. V teoretické části budou popsány základní principy a pravidla a SW platformy umožňující s daty pracovat. V praktické části bude zpracováno několik analýz na základě skutečných dat ve vybraném podniku, včetně praktických zkušeností týkajících se získání dat, ošetření chyb atd. Vybraná data budou zpracována také s užitím metod analýzy časových řad (trendy, klouzavé průměry atd.)

Osnova:

- Vymezení pojmu datový sklad.
- Implementace datových skladů.
- Využití datových skladů.
- Příklad datových skladů v malém a středním podniku.

Rozsah grafických prací: –  
Rozsah pracovní zprávy: cca 35 stran  
Forma zpracování bakalářské práce: tištěná/elektronická  
Seznam odborné literatury:

CONOLLY, T., BEGG, C., HOLOWCZAK, R. Databáze: profesionální průvodce tvorbou efektivních databází. Brno: Computer Press, 2009. ISBN 978-80-251-2328-7.

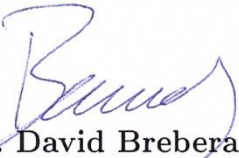
DOGDE, M. Microsoft Excel 2010: Kompletní průvodce do posledního detailu. Brno: Computer Press, 2010. ISBN 978-80-251-3354-5.

HUMPHRIES, M. Data warehousing. Praha: Computer Press, 2001. ISBN 80-7226-560-1.

LACKO, L. Databáze: datové sklady. OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle. Brno: Computer Press, 2003. ISBN 80-7226-969-0.

SODOMKA, P., KLČOVÁ, H. Informační systémy v podnikové praxi. 2. aktualizované a rozšířené vydání. Brno: Computer Press, 2010. ISBN 978-80-251-2878-7.

Vedoucí bakalářské práce:



Mgr. David Brebera

Ústav matematiky a kvantitativních metod

Datum zadání bakalářské práce: 29. září 2015

Termín odevzdání bakalářské práce: 29. dubna 2016



doc. Ing. Renáta Myšková, Ph.D.  
děkanka

L.S.



doc. Ing. Marcela Kožená, Ph.D.  
vedoucí ústavu

V Pardubicích dne 29. září 2015

## **Prohlášení**

Prohlašuji, že tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou, nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Nesouhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 28. 4. 2016

Štěpán Jahoda

## **Poděkování**

Na tomto místě bych rád poděkoval vedoucímu bakalářské práce panu Mgr. Davidu Breberovi. Můj velký dík také patří Ing. Monice Papouškové za cenné rady a připomínky.

Dále děkuji panu Ing. Zdeňku Kuntovi za umožnění přístupu k firemním datům, za poskytnutí konzultací, cenných rad a věnovaný čas.

Závěrem děkuji celé své rodině za velkou podporu při mém studiu.

## **Anotace**

Předmětem této bakalářské práce je popsání datových skladů a business intelligence. Mezi hlavní body práce patří především zpracování dat ze systému v podniku a jeho následná interpretace. Teoretická část práce pojednává o vývoji business intelligence a jeho využití. Na to navazují datové sklady. Praktická část je věnována popsání podniku, popsání problému, který podnikový systém má a následně jeho řešení.

## **Klíčová slova**

Datový sklad, business intelligence, analýza dat, reporting

## **Title**

Utilization of data warehouses in small and medium-sized enterprise

## **Annotation**

The subject of this bachelor thesis is to describe the data warehouse and business intelligence. The main points of my thesis are data analysis from a company and their interpretation. The theoretical part deals with the development business intelligence and its utilization. Next part is data warehouse and how its link with business intelligence. The practical part is devoted to describing the company and their problem with enterprise system and its solution.

## **Keywords**

Data warehouse, business intelligence, data analysis, reporting

# Obsah

<b>ÚVOD .....</b>	<b>10</b>
<b>I. TEORETICKÁ ČÁST .....</b>	<b>11</b>
<b>1 ZÁKLADNÍ POJMY .....</b>	<b>11</b>
1.1 DATA .....	11
1.2 INFORMACE .....	11
<b>2 BUSINESS INTELLIGENCE .....</b>	<b>11</b>
2.1 DEFINICE BI .....	12
2.2 HISTORIE BI .....	12
2.3 POSTAVENÍ BI V IS/ICT .....	13
2.4 NÁSTROJE A APLIKACE BI .....	14
<b>3 DATOVÉ SKLADY (DWH – DATA WAREHOUSE).....</b>	<b>16</b>
3.1 DEFINICE .....	16
3.2 DŮLEŽITÉ POJMY SPOJENÉ S DATOVÝMI SKLADY.....	18
3.3 NÁVRH DATOVÉHO SKLADU, HARDWARE A SOFTWARE.....	19
3.3.1 <i>Metody budování datového skladu</i> .....	20
3.4 ARCHITEKTURA.....	21
3.4.1 <i>Dvouvrstvá architektura</i> .....	21
3.4.2 <i>Třívrstvá architektura</i> .....	21
<b>4 OLAP .....</b>	<b>21</b>
4.1 FAKTA A DIMENZE.....	23
4.2 SCHÉMATA DATABÁZÍ .....	24
4.3 MOLAP, ROLAP, HOLAP .....	26
4.3.1 <i>MOLAP (Multidimensional Online Analytical Processing)</i> .....	26
4.3.2 <i>ROLAP (Relational Online Analytical Processing)</i> .....	26
4.3.3 <i>HOLAP (Hybrid Online Analytical Processing)</i> .....	26
4.4 REPORTING.....	27
<b>II. PRAKTICKÁ ČÁST .....</b>	<b>28</b>
<b>5 SEZNÁMENÍ S PODNIKEM .....</b>	<b>28</b>
<b>6 ŘEŠENÍ PROBLÉMU .....</b>	<b>28</b>

6.1 ZÍSKÁNÍ DAT.....	29
6.2 ZPRACOVÁNÍ DAT.....	33
<b>ZÁVĚR.....</b>	<b>42</b>
<b>SEZNAM POUŽITÝCH ZDROJŮ .....</b>	<b>44</b>



## Seznam ilustrací

<b>Obrázek 1:</b> Postavení BI v aplikační architektuře v IS/ICT .....	13
<b>Obrázek 2:</b> Komponenty Business Intelligence .....	15
<b>Obrázek 3:</b> Komponenty BI a jejich vazby .....	16
<b>Obrázek 4:</b> Datový sklad.....	18
<b>Obrázek 5:</b> Vztah transakčních a analytických databází v PIS.....	18
<b>Obrázek 6:</b> OLAP kostka, příklad dimenzí.....	24
<b>Obrázek 7:</b> Příklad hvězdicového schéma .....	25
<b>Obrázek 8:</b> Příklad schéma sněhové vločky .....	25
<b>Obrázek 9:</b> Připojení ke zdroji v MS Excel .....	30
<b>Obrázek 10:</b> Přihlášení k serveru .....	30
<b>Obrázek 11:</b> Výběr databáze .....	31
<b>Obrázek 12:</b> Nastavení SQL příkazu .....	31
<b>Obrázek 13:</b> Možnost importování dat .....	32
<b>Obrázek 14:</b> Importovaná databáze ze serveru .....	33
<b>Obrázek 15:</b> Rozdělené datum .....	34
<b>Obrázek 16:</b> Rozdělené číslo skladové položky .....	34
<b>Obrázek 17:</b> Výpočty prodejů a výpočet slev .....	35
<b>Obrázek 18:</b> Nastavené bonusy pro jednotlivé odběratele.....	36
<b>Obrázek 19:</b> Výpočet bonusu a konečná prodejní cena .....	37
<b>Obrázek 20:</b> Celkový přehled .....	37
<b>Obrázek 21:</b> Vytvoření kontingenční tabulky.....	38
<b>Obrázek 22:</b> Výběr dat pro kontingenční tabulku.....	38
<b>Obrázek 23:</b> Výsledná kontingenční tabulka .....	39
<b>Obrázek 24:</b> Porovnání jednotlivých let u celkového prodeje, prodeje katalogových výrobků a atypů a prodej katalogových výrobků .....	40
<b>Obrázek 25:</b> Porovnání TOP 10 firem podle obrátu mezi roky 2013 a 2014 .....	41

## Úvod

Pro zpracování mé bakalářské práce (dále jen práce) jsem si zvolil téma Využití datových skladů v malém a středním podniku. Datové sklady jsou důležitou součástí celé společnosti, neboť díky správné údržbě datového skladu, může firma jednoduše zpracovávat data, která právě potřebuje, dále s nimi pracovat a na základě výstupů rozhodovat.

V dnešní době mají společnosti spoustu cenných dat, která mohou pomoci k dalšímu rozvoji firmy či k rozhodování. Díky datovým skladům jsou data přehlednější, dostupnější a mají větší potenciál uspokojit uživatele, který bude hledat data ke své potřebě. Již zmíněné datové sklady jsou hlavními prvky tzv. Business Intelligence (BI).

Hlavním cílem této bakalářské práce je popis datových skladů, ke kterým se váže problematika Business Intelligence a v návaznosti na to provedení výzkumu použití v malém a středním podniku. Ke splnění tohoto cíle, je potřeba splnit dílčí cíle, které vycházejí z hlavního cíle. Těmito podcíli jsou: základní pojmy, definice BI, historie a současnost BI, nástroje a aplikace BI, definice datových skladů, pojmy spojené s datovými sklady, architektura datových skladů, definice OLAP databáze, struktura a reporting.

První kapitola bude úvod do celé problematiky a budou zde vysvětleny základní pojmy. Druhá kapitola se zabývá pojmem BI. V této kapitole bude definován tento pojem, jeho vývoj v historii a jeho vnímání v současnosti a nakonec jeho nástroje. Třetí kapitola je zaměřena na datový sklad. Budou zde vysvětleny nejdůležitější pojmy spojené s datovými sklady a k závěru této kapitoly zde bude uvedena architektura datových skladů. Čtvrtá kapitola se zaměřuje na problematiku spojenou s OLAP databázemi. Bude zde uvedena definice, jejich struktura a reporting.

Pátá kapitola bude zaměřena na praktickou část. Zde bude uveden popis podniku, jehož data budou využita v praktické části. Stručně bude popsán jejich informační systém, který používají. V šesté kapitole bude ukázáno, jak se data získají a jak se upraví pro potřeby daného podniku. Pro lepší orientaci, bude tato část doplněna několika obrázky.

# I. Teoretická část

Majetkem podniku jsou mimo jiné informace, kterými disponuje. A tyto informace bývají někdy nejcennějším majetkem, který podnik vlastní. Tyto informace jsou buď uloženy v primárním systému, nebo v datových skladech, ale v podobě dat, která se nejdříve musí zpracovat a správně interpretovat a teprve se z nich stanou informace.

## 1 Základní pojmy

### 1.1 Data

Pojem data má mnoho definic a vysvětlení. Data jsou údaje a fakta, která mají různé formy, např. numerické, textové nebo obrázkové. Odrážejí stav reality v určitém okamžiku, a proto je nelze měnit, ale lze pouze získat nová data o realitě v jiném časovém okamžiku. V databázi výrobního podniku může být název výrobku, nebo počet kusů a jeho cena. Smyslem zpracování dat je vytvoření **informace**.

Existují dvě skupiny lidí. První skupina jsou uživatelé dat, kteří využívají informace v nich obsažené, nejčastěji se jedná o manažerskou funkci. Druhá skupina je ta, která zodpovídá za údržbu a správu těchto dat.

### 1.2 Informace

Informace má různá pojetí (jako „každodenní“ laické, matematické, filozofické, komunikační, kybernetické, ...) a definice. Jedna z definic od Kuhlena zní: „*Informace je podmnožina poznatků, která je někým použita v konkrétní situaci pro řešení problémů*“ [10] Takže informace má jednoduchou funkci. Ulehčit dané rozhodnutí, aby rozhodnutí bylo kvalitní a správné. Z předchozí kapitoly víme, že před informacemi jsou data. Když se k datům přidá význam, vznikne informace.

## 2 Business Intelligence

Podnik v dnešní době využívá mnoho systémů, které pomáhají podniku v jeho aktivitách a zájmech. Jako příklad zmíněných systémů může být CRM (Customer Relationship Management), ERP (Enterprise Resource Planning), nebo SCM (Supply Chain Management). Mezi tyto systémy se řadí i Business Intelligence.

Tato kapitola se zabývá Business Intelligence.

## 2.1 Definice BI

Při hledání na internetu anebo v knihách, lze nalézt mnoho definic týkajících se pojmu Business Intelligence (dále jen BI). Tyto interpretace se různě liší a proto vhodné pro příklad uvést několik z nich.

*„BI představuje specifický typ úloh informatiky, které téměř výlučně podporují analytické, plánovací a rozhodovací činnosti podniků a organizací a jsou postaveny na principech, které těmto činnostem nejvíce odpovídají.“ [10]*

*„Sada procesů, aplikací a technologií, jejichž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Podporují analytické a plánovací činnosti podniku a organizací a jsou postaveny na principech multidimenzionálních pohledů na podniková data.“ [8]*

Business Intelligence System v IBM Journal of Research and Development tento pojem definoval jako: *„schopnost pochopit vzájemné vztahy prezentovaných faktů takový způsobem, který umožní provést akci k dosažení požadovaného cíle.“ [2]*

*„Business intelligence můžeme chápat jako ucelený a efektivní přístup k práci s firemními daty, který má vliv na správnost strategických rozhodnutí, a tím i na obchodní úspěch společnosti. V současném vysoce konkurenčním prostředí představuje informovanost jednu z hlavních konkurenčních výhod. Tato výhoda spočívá ve schopnosti efektivně využít data nashromážděná ve firmách k tvorbě informací a znalostí, na základě kterých můžeme reagovat na rychle se měnící požadavky trhu a našich zákazníků.“ [9]*

## 2.2 Historie BI

Hans Peter Luhn, počítačový vědec z IBM, byl první, kdo poprvé zmínil termín Business Intelligence, a to v roce 1958. Definoval BI jako *„schopnost pochopit vzájemné vztahy z prezentovaných faktů takovým způsobem, který umožní dovést akci k požadovanému cíli.“* Samozřejmě v té době bylo BI spíše jen teoreticky vymezeno a patřilo více do kategorie systémů pro podporu rozhodování.

Koncem 70. let se začíná objevovat řešení manažerských a analytických úloh v podniku v souvislosti s online zpracováváním daty. Firma Lockheed se snažila o aplikace z této oblasti. Brzy se ukázalo, že tyto řešení jsou velmi efektivní pro podnik, a proto se na trhu začaly objevovat první produkty založené na multidimenzionálním uložení a zpracování dat neboli EIS (Executive Information System).

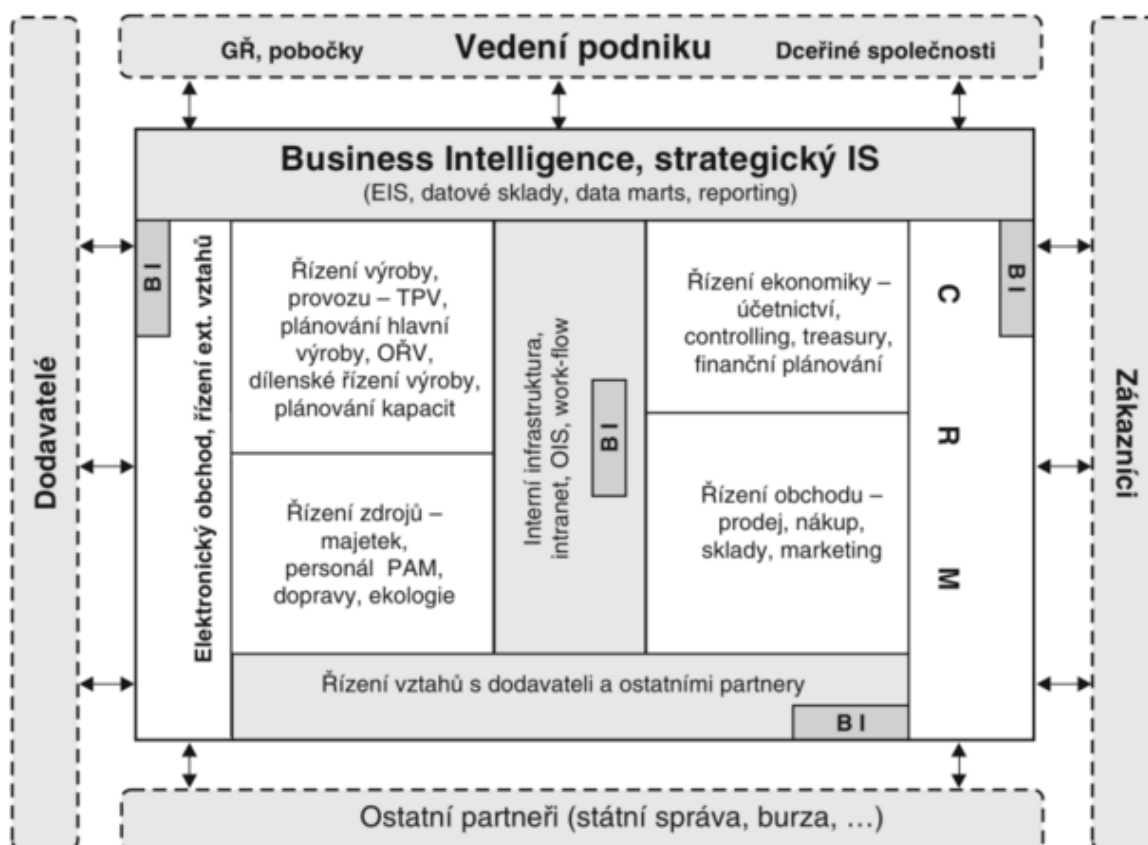
Na přelomu osmdesátých a devadesátých let se začal s nárůstem objemu dat rozmáhat trend zavádění datových skladů. Na základě toho, Howard J. Dresner, zaměstnanec konzultační společnosti Gartner Group, vymezil pojem Business Intelligence a to jako „sadu konceptů a metod pro zlepšení rozhodování firem“. [8]

Do dnešní doby se tento termín rozšířil po celém světě a jeho technologie jsou využívány ve většině úspěšných společnostech.

### 2.3 Postavení BI v IS/ICT

BI lze chápat jako široký komplex nástrojů, které jak vstupně tak i výstupně spolupracují s aplikacemi svého okolí.

Aplikace BI jsou určeny k využití dat, která vznikají na jiném místě (obrázek 1), ze kterých následně čerpá. V některých případech současnosti se stává, že aplikace čerpá data z IS/ICT a poté je zpět do nich vrací. Proto můžeme říct, že kvalita řešení BI je úzce spojena s kvalitou ostatních aplikací a především na kvalitě jejich dat, resp. databází.



Obrázek 1: Postavení BI v aplikační architektuře v IS/ICT

Zdroj:[8]

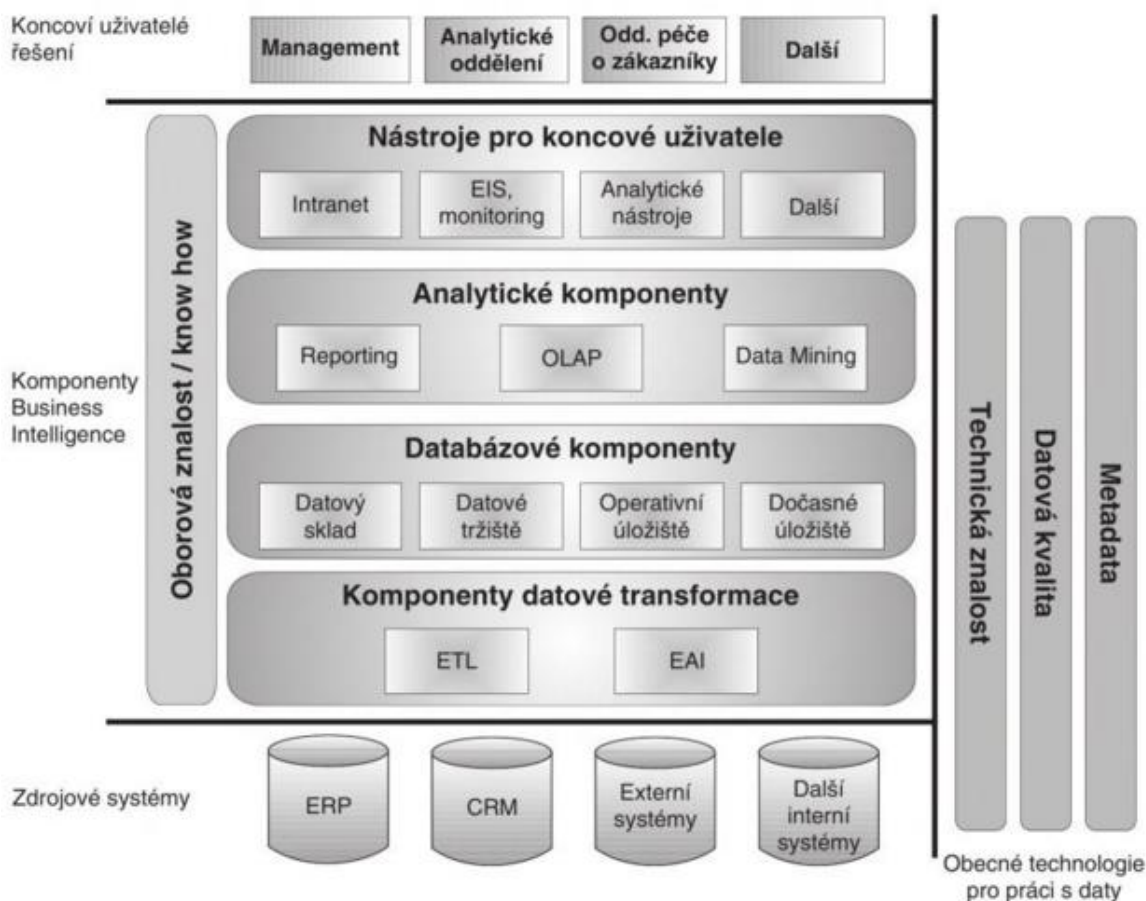
## 2.4 Nástroje a aplikace BI

Do nástrojů a aplikací BI se zahrnují [8]

- Zdrojové systémy
- Dočasná uložení dat (DSA – Data Staging Area)
- Transformační nástroje (ETL – Extraction Transformation Loading)
- Integrované nástroje (EAI)
- Datové sklady (DWH – Data Warehouse)
- Datová tržiště (DMA – Data Marts)
- OLAP (On Line Analytical Processing)
- Reporting
- Manažerské aplikace (EIS – Executive Information Systems)
- Dolování dat (Data Mining)
- Nástroje pro zajištění kvality dat

Některé z výše uvedených nástrojů, z hlediska své důležitosti pro architekturu datových skladů, budou více rozebrány v kapitole 2.2.

Pro lepší představu, jak tyto komponenty spolu souvisí, uvedu zde jednoduché schéma, které ukáže jednotlivé vrstvy.



**Obrázek 2:** Komponenty Business Intelligence

*Zdroj: [8]*

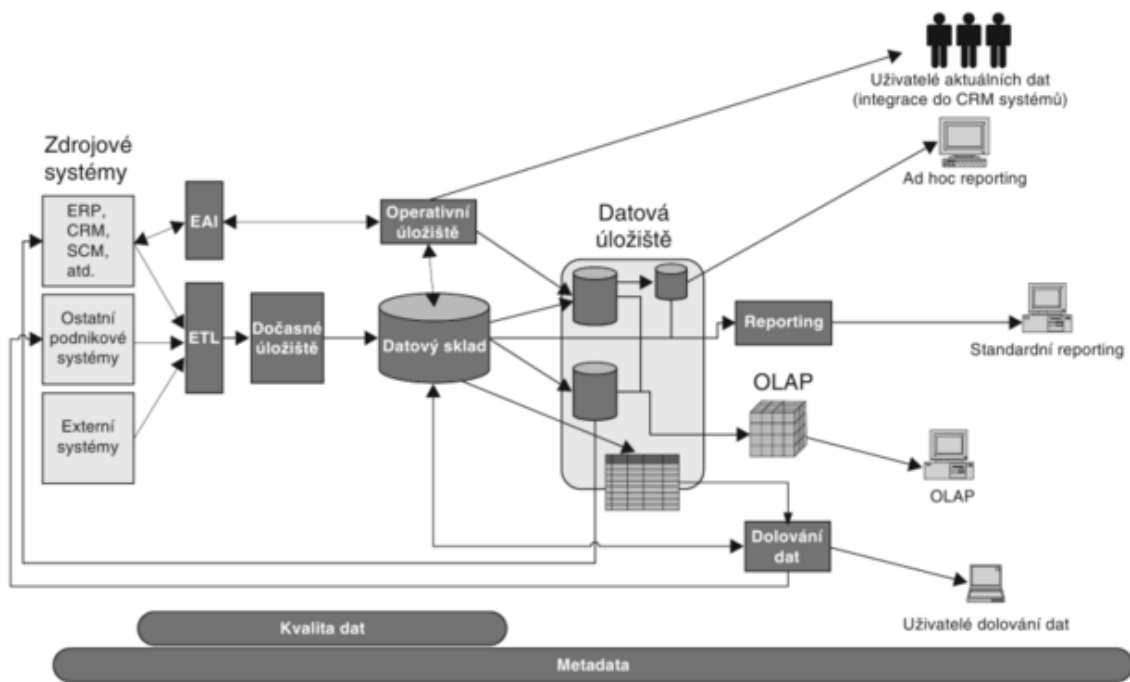
Z obrázku 2 tedy jasně vyplývá, že nástroje BI jsou rozděleny do 4 hlavních na sebe navazujících vrstev. Jednotlivé vrstvy jsou i zároveň jednotlivé typy procesů, na jehož základě jsou data zpracována. Pořadí je pevně dané a proto by se v praxi nemělo stát, aby data byla zpracována dříve, než úspěšně projde předešlou vrstvou.

První fáze představuje v podstatě filtr, který extrahuje a transformuje potřebná data tak, aby odpovídala požadovanému formátu, do jejich nového úložiště, kam jsou nahrána a následně se s nimi dále pracuje.

Další fáze je určena na analýzu dat, které jsou připraveny z předešlé fáze. Tato vrstva tedy slouží, aby se z připravených dat získaly potřebné a cenné strategické informace, které mohou být použity v rozhodování společnosti.

Konečná fáze je jednoduše řečeno jistá prezentace výsledků analýz koncovým uživatelům. Ti pak mohou díky těmto výsledkům lépe rozhodovat.

Na dalším obrázku bude pohled na hlavní vazby BI.



Obrázek 3: Komponenty BI a jejich vazby

Zdroj: [8]

### 3 Datové sklady (DWH – Data Warehouse)

Datové sklady jsou hlavním dodavatelem dat pro aplikace BI. Proto se v této kapitole podíváme, jak jsou datové sklady definovány, jaké mají podoby a základní principy jejich architektury.

V podniku je mnoho dat a většinou bývají roztroušena po celé společnosti v podnikových systémech a management nemá šanci všechna tato data řádně zpracovat, a proto nemůže učinit zcela správné rozhodnutí. S nárůstem dat se tento problém ještě více prohluboval a získání přehledu nad daty je složitější.

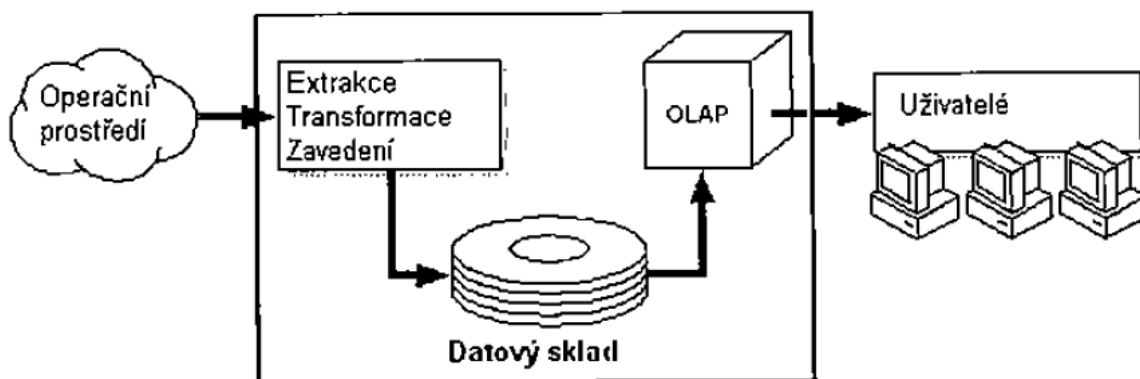
#### 3.1 Definice

Existuje mnoho definic, avšak žádná není zcela přesná. Jedna z definic datového skladu zní: „*Datový sklad je subjektivě orientovaná, integrovaná, časově proměnná a stálá kolekce dat pro podporu rozhodování managementu.*“ – W . H . Inmon

Pro lepší pochopení této definice jsou vysvětleny jednotlivé části.



- **Subjektově orientovaná:**
  - je organizovaná kolem majoritních subjektů, jako například zákazník, produkt, prodej,
  - zaměřuje se na modelování a analýzu dat, na denní zpracování operací a transakcí,
  - zajišťuje jednoduchý a stručný pohled na různé části subjektů přičemž vynechává data, která nejsou užitečná v rozhodovacím procesu.
  
- **Integrovaná:**
  - je sestavená integrací několika heterogenních zdrojů dat (relační databáze, soubory, transakce),
  - techniky na čištění a integraci dat, které zaručují konzistenci v pojmenování, attributech apod. při zpracování různých zdrojů dat (cena, měna) když se data přesouvají do datového skladu, jsou konvertována.
  
- **Časově proměnná:**
  - oproti operační databázi, která má aktuální data, datový sklad uchovává historická data (např. 5 let) resp. časový horizont je v případě datového skladu delší než u operační databáze,
  - každá klíčová struktura v datovém skladu obsahuje časový element.
  
- **Stálá:**
  - fyzicky oddělený datový sklad transformovaný z operačního prostředí,
  - operační update dat se v prostředí datových skladů nevyskytuje.



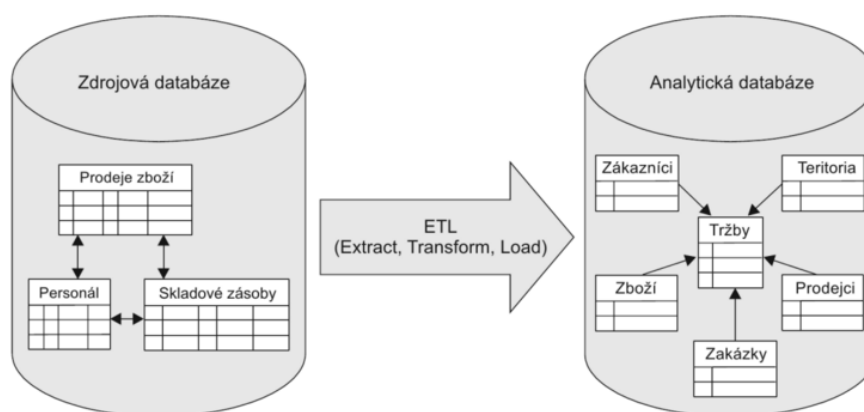
Obrázek 4: Datový sklad

Zdroj: [5]

### 3.2 Důležité pojmy spojené s datovými sklady

Nejprve je potřeba vymezit nejdůležitější pojmy spojené s datovými sklady.

- **ETL (Extaction Transformation Loading)** – jde o tzv. „datovou pumpu“. Z názvu lze jednoduše zjistit, o co jde. Nejprve slovo „Extraction“ neboli česky „získání“ má za úkol vybrat data ze zdrojového systému. Následně slovo „Transformation“ česky transformace, upraví data do požadované formy a poslední slovo „Loading“ je nahrává do datových skladů nebo datových tržišť.. Mezi nejznámější nástroje, které se používají, jsou například MS SQL Server Integration Services.



Obrázek 5: Vztah transakčních a analytických databází v PIS

Zdroj: [6]

- **Dočasné uložení dat (DSA - Data Staging Area)** – jedná se o speciální typ relační databáze, jde o prozatímní uložení extrahovaných dat z produkčních databází s cílem následně je upravit a transformovat. Jedná se pouze o krátkodobé uložení dat, většinou v rozmezí jednoho dne až měsíce. Používá se zejména u produkčních systémů, kde tento systém umožňuje vybírat a přenášet data aniž by nějak tyto systémy zatěžoval. Data zde uložená nejsou zkontrolována, nejsou historická resp. nenesou historickou hodnotu a „*jsou uložena v přesně stejné struktuře, v jaké jsou ve zdrojových systémech*“. [10]
- **Operativní uložení dat (ODS - Operation Data Store)** – opět jako DSA neobsahuje historická data. Rozdíl je v datech, která díky transformačním operacím, jsou konzistentní a v některých případech doplněné o agregaci. To ovšem není ten nejhlavnější rozdíl oproti DSA. Základní rozdíl je ten, že DSA je jakási mezizastávka, než data jsou zpracována a nahrána do datového skladu a poté jsou vymazána. ODS oproti tomu slouží jako databáze, kde přímo podporuje analytické procesy, tedy *ODS je budována právě s cílem zpřístupnit uživatelům nebo ostatním systémům data pro analýzy či dotazy s minimálním zpožděním oproti jejich pořízení*. [10]
- **Datové tržiště (DMA – Data Marts)** – jsou velmi podobné datovým skladům, rozdíl je ovšem v jejich určení a to pouze pro omezený okruh uživatelů resp. pouze pro jeden podnikový útvar. Ve firmách se občas stává, že se tvoří tyto decentralizované datové sklady a až později se vytvoří jeden celopodnikový datový sklad. Datová tržiště mohou nadále fungovat i za přítomnosti již vybudovaného datového skladu. V tomto případě datové tržiště pokrývá konkrétní problematiku dané části podniku a slouží jako mezistupeň při transformaci dat. [10]

### 3.3 Návrh datového skladu, hardware a software

„*Největší procento z hodnoty datového skladu se skrývá v jeho návrhu a koncepci.*“ [5]

51 Pro společnost, která se chystá implementovat datový sklad do svého podniku, je důležitý návrh datového skladu. Tento návrh může firma sama vytvořit nebo využije služeb třetích stran.

**Hardware** je nedílnou složkou datového skladu ale nahraditelnou. Záleží na společnosti, jestli se rozhodne investovat větší finanční zdroje do hardwaru, aby byl přístup k datům rychlejší a bezpečnější.

V dnešní době se lze více setkat s analytickými nástroji přímo v databázových serverech. To se samozřejmě podepíše na ceně. Za tyto analytické služby se někdy platí licenční poplatky.

### **3.3.1 Metody budování datového skladu**

Pro úspěšnost a správnou funkčnost datového skladu, je třeba postupovat dle vhodné strategie. Je tedy nutné správně vybrat metodu implementace datového skladu. Nejznámější metody jsou **metoda „velkého třesku“** a **přírůstková metoda**.

#### **Metoda „velkého třesku“**

Princip **metody „velkého třesku“** spočívá v implementaci datového skladu během jediného projektu. Je zde ovšem velká časová náročnost a to přináší riziko, protože během doby realizace projektu, mohou vznikat nové technologie a jiné požadavky klientů a uživatelů. Metoda se skládá ze tří etap:

1. Analýza požadavků podniku
2. Vytvoření podnikového datového skladu
3. Vytvoření přístupu buď přímo, nebo přes datové trhy

**Výhodou** je vypracování celého projektu datového skladu ještě před samotnou realizací. **Nevýhodou** je menší flexibilita při změnách požadavků a zdoluhavý proces vývoje, který se odvozuje od velikosti datového skladu.

#### **Přírůstková metoda**

**Přírůstková metoda** nazývaná také jako evoluční, vzhledem k budování skladu po jednotlivých etapách. Oproti výše zmíněnému se místo vybudování jednoho celého datového skladu, tvoří postupně jednotlivá řešení. Touto metodou se udržuje spojitost mezi řešením a požadavky klienta

**Výhodou** je zachování kontinuity budovaného projektu s požadavky a potřebami uživatele. Umožňuje implementovat škálovatelnou, tedy rozšířitelnou architekturu a zabezpečí

rychlejší zisk a tím i rychlejší návratnost investice. **Nevýhodou** je častější komunikace a jednání se zadavatelem.

### 3.4 Architektura

Při budování datového skladu se jako jejich architektura nejvíce prosadily dva základní koncepty datového skladu a to dvouvrstvá a třívrstvá architektura.

#### 3.4.1 Dvouvrstvá architektura

Budování datového skladu vyžaduje často vysoké náklady a to jak finanční tak i časové a proto se v praxi často vyskytuje dvouvrstvá architektura. S touto architekturou přišel Ralph Kimball a na rozdíl od třívrstvé architektury, její výstavba je lacinější a časově kratší. Neobsahuje centrální uložení dat, ale pouze samotná datová tržiště. Ale při použití této architektury se stává, že některá data se objevují duplicitně, takže to si žádá i větší paměťový prostor a tím se zvyšuje počet tržišť a klesá tím konsolidovanost a rostou náklady na údržbu.

#### 3.4.2 Třívrstvá architektura

Tvůrcem třívrstvé architektury je Bill Inmon. Koncepce obsahuje dočasné uložení, centrální uložení a datové tržiště. Tato koncepce má počáteční náklady větší a doba realizace je delší. Oproti předchozí architektuře má však řadu výhod. Například je flexibilní a integrovaná pro podporu náročných analytických úloh anebo díky komplexnosti uložených dat, lze budovat takřka neomezené množství datových tržišť. [8]

## 4 OLAP

Dr. E. F. Codd byl počítačový vědec, který pracoval pro společnost IBM a měl značný přínos v teorii relačních databází a jeho příspěvky k této problematice zůstaly, i přes jeho další příspěvky v informatice, nejvíce známým úspěchem.

Řeč je o termínu OLAP, který zavedl již zmiňovaný Dr. E. F. Codd a tímto termínem popisuje technologii, která má pomoci překlenout mezeru mezi využitím osobních počítačů a řízením podnikových dat.

Kromě definic je známé tzv. „dvanáctero“ pravidel OLAP od Dr. E. F. Codd. [1]

- **Pravidlo první:** Multidimenzionální konceptuální pohled - OLAP by měl uživateli poskytovat multidimenzionální model odpovídající jeho

podnikatelským záměrům tak, aby tento model mohl využívat pro analýzu shromážděných údajů.

- **Pravidlo druhé:** Transparentnost - systém OLAP s podřízenou databází a architektura výpočtů by měly být natolik transparentní, aby se dala naplno uplatnit produktivita a odbornost pracovníka. Klientský program se připojuje k OLAP serveru. Důležitá je i heterogenost vstupních dat, kterou zajišťuje procesu ETL.
- **Pravidlo třetí:** Dostupnost - systém OLAP by měl přistupovat pouze k těm údajům, které jsou potřebné pro analýzu. Údaje by měly být přístupné z kteréhokoliv podnikového zdroje. Tuto činnost umožňuje etapa ETL, ve které se údaje očistí, zahustí a přetransformují.
- **Pravidlo čtvrté:** Konzistentní vykazování – i při postupném zvětšování objemu záznamů by uživatel neměl pocítit žádné zásadní snížení výkonu.
- **Pravidlo páté:** Architektura klient-server - Systém OLAP musí odpovídat principu architektury klient-server. Přihlíží se také k maximální ceně a výkonu, flexibilitě a interoperabilitě<sup>1</sup>.
- **Pravidlo šesté:** Generická dimenzionalita – Je třeba, aby všechny dimenze údajů byly ekvivalentní ve struktuře i operačních schopnostech.
- **Pravidlo sedmé:** Dynamické ošetření řídkých matic - Systém OLAP by měl být schopen přizpůsobit své fyzické schéma na konkrétní analytický model, čímž se ošetří řídké matice a dosáhne se požadované úrovně výkonu.
- **Pravidlo osmé:** Podpora pro více uživatelů - Systém OLAP musí být schopen podporovat celou skupinu uživatelů pracujících současně z více míst na jednom modelu.
- **Pravidlo deváté:** Neomezené křížové dimenzionální operace - Systém OLAP musí umět určit dimenzionální hierarchie a sám provádět asociované kumulované kalkulace v rámci dimenzí i mezi nimi.

---

<sup>1</sup> Schopnost různých systémů vzájemně a efektivně spolupracovat a poskytovat si služby.

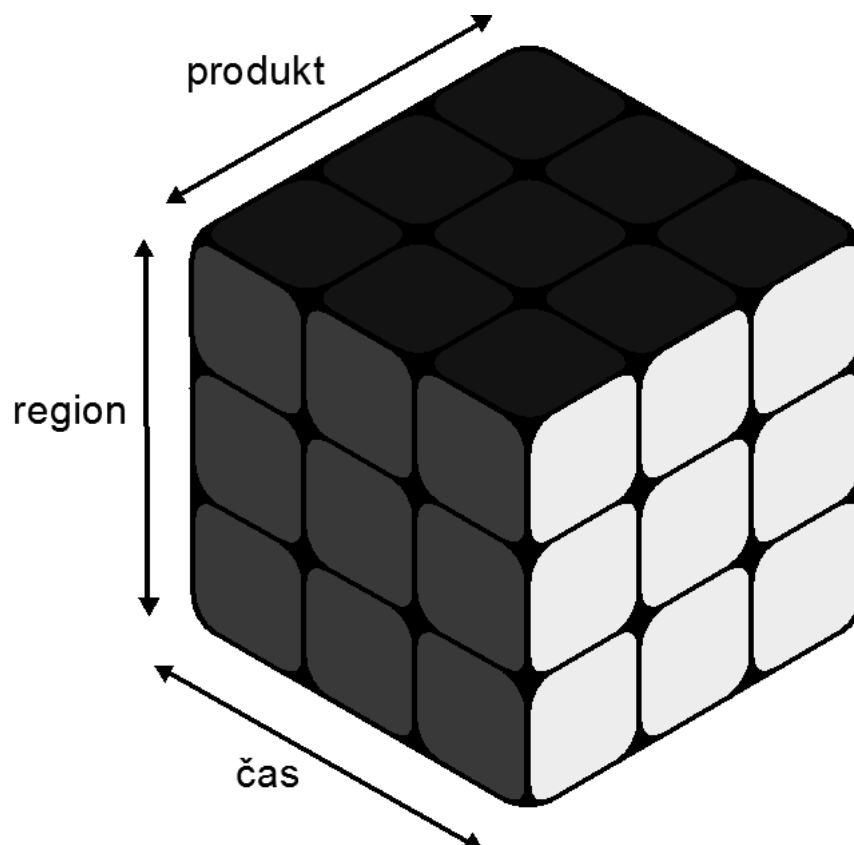
- **Pravidlo desáté:** Intuitivní manipulace s údaji - Pravidlo definuje konsolidované přeorientování cest na detailní úroveň a zpět (drill down, drill up). Uživatelské rozhraní by mělo intuitivním způsobem umožňovat manipulace s buňkami krychle.
- **Pravidlo jedenácté:** Flexibilní vykazování – Zajištění schopnosti řazení řádků, sloupců a buněk pro vznik sestav. Takovéto pohledy jsou pak využity jako podpůrné prostředky při práci analytiků.
- **Pravidlo dvanácté:** Neomezené dimenze a úrovně agregace - V závislosti na požadavcích podnikání nabývá analytický model libovolného množství dimenzí, přičemž každá z nich může mít vícenásobné hierarchie. Nemělo by docházet k omezení počtu dimenzí.

#### 4.1 Fakta a dimenze

Jsou dva, kterými je tvořena každá krychle OLAP, a to fakty a dimenzemi.

„**Fakta** jsou numerické měrné jednotky obchodování.“ [5], a proto není divu, že tabulka faktů je největší tabulka v databázi a obsahuje velký objem dat. Spolu s dimenzemi potom mohou tvořit určitá schémata, například hvězdicové schéma nebo schéma sněhové vločky. Tato schémata budou popsána v následující podkapitole.

„**Dimenze** obsahují logicky nebo organizačně hierarchicky uspořádané údaje. Jsou to vlastně textové popisy obchodování.“ [5] Oproti faktům jsou tabulky dimenzí menší a data v nich jsou více neměnné.



**Obrázek 6:** OLAP kostka, příklad dimenzí

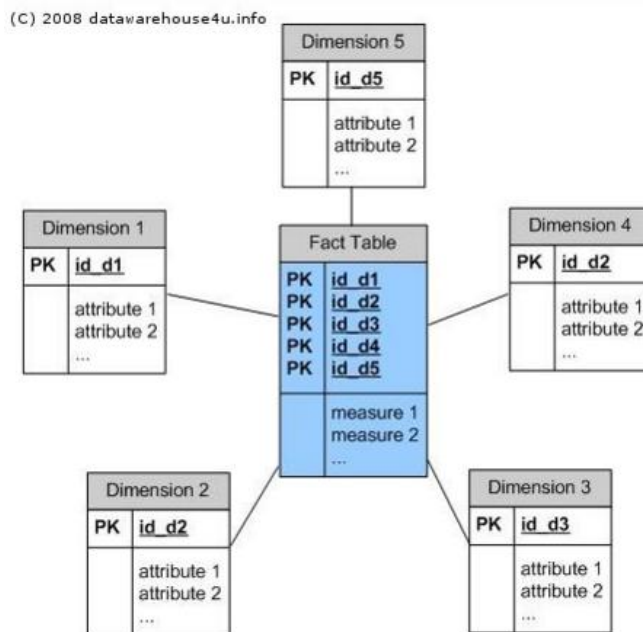
*Zdroj: vlastní zpracování*

## 4.2 Schémata databází

Nejvíce používaná schémata databází jsou hvězdicové schéma a schéma sněhové vločky, neboli anglicky „Star schema“ a „Snowflake schema“.

**Hvězdicové schéma** je jednoduché na pochopení, ale relativně pomalé na vytvoření takového modelu. Princip spočívá v tom, že tabulka faktů obsahuje cizí klíče, které se vztahují k primárním klíčům v tabulkách dimenzí. Tento typ databáze disponuje vysokým dotazovacím výkonem. Uvedu zde příklad tohoto schématu.

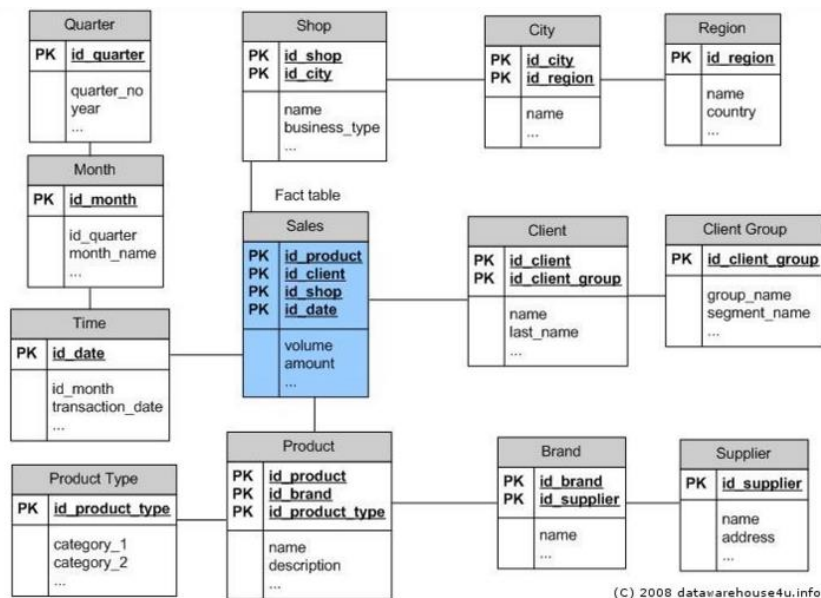




Obrázek 7: Příklad hvězdicového schéma

Zdroj: [12]

**Schéma sněhové vločky**, oproti předchozímu schématu, obsahuje některé dimenze sestavené z dalších tabulek. Sice tento model umožňuje rychlejší zavedení dat do tabulek, ale má nižší dotazovací výkon, kvůli množství spojení tabulek. Více nám prozradí jednoduché schéma pro představu.



Obrázek 8: Příklad schéma sněhové vločky

Zdroj: [11]

## 4.3 MOLAP, ROLAP, HOLAP

Nejprve si zde uvedeme rozdíl mezi relační a multidimenzionální databází neboli OLAP.

**Relační databázový model**, jehož údaje jsou v dvoudimenzionální tabulce, tedy vlastně dvourozměrné struktury. Sloupce jsou atributy a řádky jsou záznamy.

**Multidimenzionální databázový model** jak z názvu vypovídá, jde o vícerozměrnou strukturu a nejčastěji se zobrazuje jako vícerozměrná krychle, která je ekvivalentem tabulky v relační databázi. Každá krychle má několik dimenzí. Pro člověka je nejjednodušší představit si klasickou krychli, která je trojrozměrná, tedy jde o tři dimenze, ale počet dimenzí v multidimenzionálních databázích je větší.

Ovšem s větším počtem dimenzí rostou i požadavky na úložný prostor a proto se v praxi používají různé technologie na kompresi použitého místa na disku.

### 4.3.1 MOLAP (Multidimensional Online Analytical Processing)

Jedná se o tradiční způsob analýzy OLAP, kde jsou data uložena v multidimenzionální databázi. Je vysoce optimalizován pro maximální výkon vzhledem k dotazům uživatelů. Další výhodou je možnost provádění složitých výpočtů, protože všechny výpočty byly předem vygenerované při vytvoření krychle. Nevýhodou je omezené množství údajů, které lze provést.

### 4.3.2 ROLAP (Relational Online Analytical Processing)

Tato varianta poskytuje uživateli multidimenzionální náhled dat, aniž by došlo k duplikování dat, tedy zůstávají uložena ve své původní databázi. Výhodou je možnost zpracování velkého objemu dat, není zde v podstatě žádné omezení. Nevýhodou tedy je výkon, a to vede k prodlevě odpovědi na SQL dotaz.

### 4.3.3 HOLAP (Hybrid Online Analytical Processing)

Zde se dostáváme ke kombinaci výše zmíněných variant, které se pokouší spojit výhody a eliminovat nevýhody obou a dát je dohromady. Data jsou ponechána v původních relačních tabulkách. Toto řešení nabízí propojení mezi velkým objemem dat v relačních tabulkách a poskytuje rychlejší zpracování multidimenzionálních agregací. [14]

## 4.4 Reporting

Každý si reporting představuje trochu jinak. Lidé z IT oddělení si budou představovat reporting jako OLAP databázi. Pracovníci z účtárny je budou mít spojené s účetními výkazy. A manažery obchodu si představí barevné grafy, které jim umožní rozhodování.

A každý z nich má pravdu neboť není stanovená přesná a jasná definice, jakou by měl mít reporting podobu. Ovšem princip je známý.

Na začátku máme neuspořádaná data, ve kterých se málokdo vyzná, a proto existuje reporting, díky kterému lze tyto data uspořádat například do tabulek nebo grafů, kterým už daný manažer rozumí a je pro ně jednodušší učinit rozhodnutí.

Máme zde několik typů reportingu, které se liší způsobem vytvořením, využitím a svým přínosem:

- **Statistický reporting** je vhodný zejména pro vizualizaci informací standardní struktury a vzhledu s takřka neměnnými vstupními parametry. Proto se hodí například pro účetní výkaznictví (výkazy za určité období), přehledy prodejů (v MS Excel), atd. Hlavní podmínka je, aby byl „čitelný“ všemi, kdo ho budou využívat.
- **Dynamický reporting** má podobnost se statickým reportingem. Rozdílný je v tom, že si uživatel může ovlivňovat obsah a formu reportu zadáváním vstupních parametrů. Velmi vhodné pro přehledy vztahující se k předem neznámým časovým obdobím, kategoriím produktů a zákazníků, v případě potřeby lze částečně ovlivnit i design samotného reportu. Výhoda toho typu reportu je, že se dá upravit k potřebám konkrétního uživatele, který bude report využívat.
- **Ad hoc reporting** se používá tehdy, pokud uživateli nevyhovuje ani jeden z výše uvedených reportů. Tento report bude odpovídat přesně jeho konkrétním potřebám. Tyto reporty jsou také velmi vhodné, pokud není dopředu známo, jakou formu a jaký obsah má daný report splňovat. [3]