

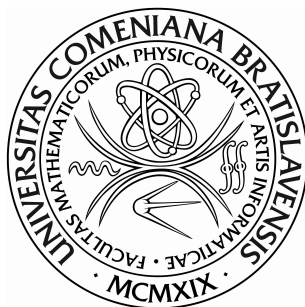
UNIVERZITA KOMENSKÉHO V BRATISLAVE
Fakulta matematiky, fyziky a informatiky

DIPLOMOVÁ PRÁCA

Katarína Türková

Bratislava 2011

UNIVERZITA KOMENSKÉHO V BRATISLAVE
Fakulta matematiky, fyziky a informatiky
Výpočtové centrum



**Využitie moderných metód spracovania údajov
pri finančnom riadení verejnej vysokej školy**

Diplomová práca

Študijný odbor:	9.1.9 Aplikovaná matematika
Študijný program:	Ekonomická a finančná matematika
Vedúci práce:	doc. RNDr. Peter Mederly, CSc.
Diplomant:	Katarína Türková
Kód:	8814814f-dd6f-4947-88df-97323743e7dd

Bratislava 2011



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Katarína Türková
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov : Využitie moderných metód spracovania údajov pri finančnom riadení verejnej vysokej školy

Cieľ : Prehľad moderných metód spracovania údajov a ich implementácia v jednotnom finančnom informačnom systéme slovenských verejných vysokých škôl SOFIA.

Vedúci : doc. RNDr. Peter Mederly, CSc.

Dátum zadania: 27.01.2010

Dátum schválenia: 14.04.2011

.....
prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Dátum potvrdenia finálnej verzie práce, súhlas s jej odovzdaním (vrátane spôsobu sprístupnenia)

.....
vedúci práce

Čestné prehlásenie

Týmto prehlasujem, že som diplomovú prácu na tému „Využitie moderných metód spracovania údajov pri finančnom riadení verejnej vysokej školy“ vypracovala samostatne s použitím uvedenej literatúry.

Bratislava, 20. apríla 2011

.....
Katarína Türková

Pod'akovanie

Rada by som poďakovala doc. RNDr. Petrovi Mederlymu, CSc. za jeho vedenie, čas, cenné rady a usmernenie pri písaní tejto práce.

Abstrakt

TÜRKOVÁ, Katarína: *Využitie moderných metód spracovania údajov pri finančnom riadení verejnej vysokej školy*. [Diplomová práca] - Univerzita Komenského v Bratislave. Fakulta matematiky, fyziky a informatiky: Výpočtové centrum. - Vedúci: doc. RNDr. Peter Mederly, CSc. Bratislava, 2011, 65 strán.

Pod pojmom moderné metódy spracovania údajov rozumieme v súčasnosti dynamicky sa rozvíjajúcu oblasť Business Intelligence. Možnosti, ktoré BI ponúka v oblasti spracovania dát, poskytujú podporu riadiacim pracovníkom v rozhodovacom procese.

Táto práca je voľným pokračovaním bakalárskej práce Business Intelligence a jej podpora v systéme SAP zaoberajúcej sa všeobecným popisom BI a nástrojov SAP. Zaoberá sa podrobnejšie pojmom data warehousing, teda tvorbou dátového skladu, a analytickými a reportovacími možnosťami nad takýmto skladom. Ďalej popisuje implementáciu týchto riešení v projekte jednotného finančného informačného systému slovenských VVŠ s názvom SOFIA, ktorý sa v súčasnosti realizuje v spolupráci Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky, S&T a združenia Siemens.

Kľúčové slová: Business Intelligence, Data Warehousing, OLAP, Reporting, SAP BI, projekt SOFIA

Abstract

TURKOVA, Katarina: *The application of modern methods of data processing in financial management of the universities*. [Diploma thesis] - Comenius University in Bratislava. Faculty of Mathematics, Physics and Informatics: Computing centre. - Tutor: doc. RNDr. Peter Mederly, CSc. Bratislava, 2011, 65 pages.

The term of modern methods of data processing represents currently dynamically developing field of Business Intelligence. Possibilities that BI offers in field of data processing support management in decision-making process.

This thesis is extension of bachelor thesis Business Intelligence and its SAP system's support which concerns general description of BI and SAP portfolio. It goes deeper and describes data warehousing, i.e. process of building data warehouse, analytical and reporting methods over this warehouse. It also describes the implementation of these solutions in project of unified financial information system of Slovak universities called SOFIA realized in cooperation of Ministry of Education, Science, Research and Sports, S&T and Siemens group.

Key words: Business Intelligence, Data Warehousing, OLAP, Reporting, SAP BI, project SOFIA

Obsah

Zoznam obrázkov	8
Zoznam skratiek	9
Slovník	10
Úvod	12
1 Úvod do Business Intelligence	13
1.1 Čo je to BI a prečo vznikla?	13
1.2 Ako sa buduje BI	14
1.3 Nástroje BI	14
2 Data Warehousing	16
2.1 Ako vznikol data warehousing	17
2.2 Ako sa vytvára DW	18
2.2.1 Normalizovaný a dimenzionálny prístup k dátam	19
2.2.2 Top-down a bottom-up metodológia	20
2.2.3 Budovanie dátových skladov v krokoch	23
2.3 ETL proces	25
3 OLAP	28
3.1 Relačná databáza	28
3.2 Multidimenzionálna databáza	30
3.2.1 Tabuľky faktov a tabuľky dimenzií	31
3.2.2 Usporiadanie tabuliek v multidim. databáze - schéma	34
3.3 Príklad pre OLAP analýzu	35
3.3.1 Pripojenie/vytvorenie dátového zdroja (Data Source)	35
3.3.2 Vytvorenie pohľadu na zdroj dát (Data Source View)	36
3.3.3 Vytvorenie dimenzií kocky (Dimensions)	37
3.3.4 Vytvorenie OLAP kocky (Cube)	39
3.3.5 Prezeranie OLAP kocky	43
4 Reporting	44

5 Implementácia BI v projekte SOFIA	46
5.1 Ciele implementácie systému SAP BI	46
5.2 Popis riešení SAP Business Intelligence	47
5.2.1 SAP Netweaver Business Warehouse	47
5.2.2 SAP Business Objects Business Intelligence	48
5.3 Predpokladané zdroje údajov	49
5.4 Systémová architektúra a toky dát	50
5.5 Funkcionalita	52
5.5.1 Postup implementácie funkcionality	53
5.6 Dátový model	55
5.7 Reporting	56
5.7.1 Zoznam preddefinovaných výstupov	57
5.7.2 Tvorba vybraných výstupov v programe XCelsius	59
Záver	63
Literatúra	64

Zoznam obrázkov

1.1	Schéma znázorňujúca spracovanie dát v BI, zdroj [6]	14
2.1	Schéma DSE	17
2.2	Procesná schéma dátového skladu	18
2.3	Front-end analytics, zdroj [12]	19
2.4	Základné rozdiely vo vlastnostiach operačnej databázy a DW, zdroj [4]	19
2.5	Top-down schéma	21
2.6	Bottom-up schéma	22
3.1	Relačná databáza	28
3.2	Vzťahy medzi tabuľkami v relačnej databáze	29
3.3	Príklad OLAP-kocky zloženej z troch dimenzií: dimenzie času, produktov a zákazníkov	30
3.4	Schéma tabuliek dimenzií a faktov, zdroj [22]	33
3.5	Snowflake schéma	34
3.6	Star schéma	34
4.1	Príklad reportu, zdroj [24]	45
5.1	Architektúra SAP BW, zdroj [25]	47
5.2	Zdroje dát, zdroj [2]	50
5.3	SAP BW/BPC, zdroj [2]	51
5.4	SAP BOBJ BI, zdroj [2]	51
5.5	Toky dát medzi systémami, zdroj [2]	52
5.6	Dashboard počtu študentov VVŠ SR	60

Zoznam skratiek

BI - Business Intelligence
BV - Bežné výdavky
DSE - Decision Support Environments
DSO - Data Store Object
DW - Data Warehouse
ERP - Enterprise Resource Planning
ETL - Extract Transform Load
FIS - Finančný informačný systém
IT - Informačné technológie (Information Technology)
KPI - Kľúčové indikátory výkonu (Key Performance Indicators)
KV - Kapitálové výdavky
OD - Operačná databáza
ODS - Operational Data Store
OLAP - Online Analytical Processing
OLTP - Online Transaction Processing
PSA - Persistent Staging Area
SAP - Systems Applications and Products
SQL - Structured Query Language
VVŠ - Verejná vysoká škola
3NF - Tretia normálna forma

Slovník

Definície označené [1], [2] sú citáciami týchto diel.

Dashboard („prístrojová doska“) - účelom dashboardov je automaticky zobrazovať užívateľovi informácie z rôznych komponentov (databázy, dokumenty, aplikácie, web, ...) potrebné pre jeho aktuálnu činnosť; dashboard organizuje a prezentuje informácie jednoduchým spôsobom a šetrí čas strávený vyhľadávaním.[1]

Datamarty (Data Marts) - zoskupená časť firemných dát, ktorá vzniká na základe požiadaviek pracovného tímu, oddelenia alebo za určitým cieľom.[1]

Data Store Object - objekty dátového modelu vo forme tabuľky.[2]

Dátový sklad (Data Warehouse) - rozsiahla štruktúrovaná databáza firemných dát, základný a najdôležitejší prvok BI.[1]

Data warehouse staging area - je dočasné úložisko dát, kam sa kopírujú dáta zo zdrojových systémov, aby mohli byť upravené a očistené ešte predtým ako sú uložené do dátového skladu.

Dolovanie dát (Data Mining) - slúži na odhaľovanie hodnotných a použiteľných informácií, ktoré boli doteraz skryté vo veľkých objemoch dát; používa štatistické metódy a ďalšie metódy hraničiace s oblasťou umelej inteligencie.[1]

Extract, transform, load (ETL) - je proces extrahovania dát zo zdrojových systémov, ich transformovania na požadovaný formát (zjednotenie štruktúry dát) a uloženia do databázy alebo dátového skladu.

Extraktor - program na strane zdrojového systému poskytujúci údaje pre systém SAP BW.[2]

FIS - jednotný finančný informačný systém pre slovenské verejné vysoké školy.

Kľúčové indikátory výkonu (Key Performance Indicators) - finančné a nefinančné indikátory výkonnosti a úspešnosti spoločnosti, slúžia na ohodnotenie súčasného stavu a vývoja firmy; sú meradlom napredovania firmy pri plnení daných cieľov; KPI musia byť merateľné (finančne, percentuálne, atď.) a väčšinou sú dlhodobé.[1]

Legacy system - je stará metóda, technológia, počítačový systém alebo aplikácia, ktorá sa používa, lebo spĺňa užívateľské potreby, napriek tomu, že už existuje novšia a efektívnejšia technológia.

Masterdáta - sú zjednotené súbory základných firemných dát (o zákazníkoch, produktoch, zamestnancoch, predajcoch, postupoch, aktivitách), ich atribútov a prepojení v aplikáciách spoločnosti, ktoré sa využívajú pri vykonávaní firemnej politiky.[1]

Metadáta - sú „dáta o iných dátach“, sú to teda dáta, ktoré opisujú vlastnosti (ná-

zov, veľkosť, typ), štruktúru (polia, stĺpce, dĺžku), lokalizáciu dát alebo iné údaje o dátach.[1]

On-line Analytical Processing (OLAP) - je technológia uloženia veľkého množstva dát v databáze tak, aby boli prístupné a zrozumiteľné pre analyzovanie výsledkov a trendov; v tejto technológii sú dáta usporiadané do multidimenzionálnych objektov, tzv. kociek (OLAP cubes), ktoré umožňujú takmer okamžitú analýzu a väčšinou sa na prácu s nimi využíva programovací jazyk SQL; v hierarchii spracovania dát stojí technológia OLAP nad technológiou OLTP.[1]

On-line Transaction Processing (OLTP) - je technológia uloženia dát v databáze, ktorá umožňuje ich jednoduchú a bezpečnú modifikáciu; spracováva dáta v reálnom čase.[1]

Reporting - hlavným cieľom reportingu je informovať formou reportov, ktoré poskytujú informácie o nejakej téme, oblasti alebo probléme spoločnosti, sú tvorené pre konkrétny účel a publikum; môžu obsahovať návrhy a odporúčania na začatie budúcich akcií; pre názornosť a zrozumiteľnosť obsahujú tabuľky, grafy, obrázky, zhrnutia, hypertextové odkazy, atď.; firemný reporting je jednou zo základných súčastí budovania firemnej BI.[1]

SAP Business Objects - skupina produktov spoločnosti SAP.[2]

SOFIA SAP BI (SOFIA 2) - označenie projektu technologického a funkčného upgrade finančného informačného systému VVŠ.[2]

SOFIA SAP ERP - základný transakčný systém pre spracovanie dokladov v rámci FIS.[2]

Transakčné dáta - zahŕňajú rutinné, každodenné účtovné operácie vo firme. Tieto operácie sa týkajú odberateľa, firmy, skladu a výroby, záväzkov, kontrol zásob a podobne.

Transakčné systémy - systémy využívajúce technológiu OLTP, pracujú s transakčnými dátami.

Slice & dice, roll up, drill down, drill up - sú metódy prehľadávania dát v databázach (smerom k detailnejšiemu alebo súhrnnejšiemu pohľadu).

3NF forma (tretia normálna forma) - je tretia úroveň normalizácie databázy, definovaná Edgarom Frankom Coddom. Existuje niekoľko normálnych foriem databáz (1NF,2NF,3NF,4NF,5NF), z ktorých každá musí spĺňať osobitné podmienky.

Úvod

„Aby bol podnik konkurencieschopný potrebuje prijímať strategické rozhodnutia založené na kvalitných informáciách o výkone a trhu. Firmy majú k dispozícii dostatočne veľa dát, avšak roztrúsených v rôznych zdrojoch a databázach. Hlavnou úlohou Business Intelligence je premeniť tieto dáta na vierohodné informácie, ktoré sú relevantné pre manažérov a rozhodovacích pracovníkov. BI predstavuje systém riešení potrebných na získavanie takýchto informácií.“[[1], str.8]

V mojej bakalárskej práci Business Intelligence a jej podpora v systéme SAP som sa zaoberala všeobecnou charakteristikou oblasti BI. Cieľom tejto práce je bližšie zoznámiť čitateľa s konkrétnymi riešeniami BI a demonštrovať ich použitie na reálnom príklade (projekt SOFIA).

Úvodná časť práce poskytuje stručný popis oblasti Business Intelligence a jej nástrojov tak, ako sme ju opísali v bakalárskej práci.

V druhej kapitole sa budeme podrobnejšie zaoberať tvorbou dátového skladu - data warehousingom, metodológiou jeho tvorby a procesom vyberania, transformácie a ukladania údajov do dátového skladu tzv. ETL.

Tretia kapitola popisuje rozdiely medzi relačnou a multidimenzionálnou databázou, výhody usporiadania dát v multidim. databáze vhodného pre multidimenzionálnu analýzu (OLAP) a ponúka konkrétny príklad.

Štvrtá kapitola sa krátko zaoberá pojmom reportingu.

Posledná piata kapitola popisuje implementáciu BI v projekte jednotného finančného informačného systému slovenských verejných vysokých škôl SOFIA. Hlavným zdrojom pri písaní tejto kapitoly bol Cieľový koncept projektu SOFIA, kapitola BI - Reporting. Zahŕňa ciele projektu, popis riešení, zdroje údajov, architektúru systémov, dátový model a pod. Na konci kapitoly uvádzame vzorové príklady výstupov zo systému (dashboardov) vytvorených v programe XCelsius Engage 2008.

Ďalšími významnými zdrojmi informácií pri písaní tejto práce bola bakalárska práca Business Intelligence a jej podpora v systéme SAP a Internet. Všetky zdroje, z ktorých som čerpala, sú uvedené na konci práce v Literatúre.

Kapitola 1

Úvod do Business Intelligence

Kapitola 1 je citáciou diela [1].

1.1 Čo je to BI a prečo vznikla?

1989 - Vznik termínu „Business Intelligence“ - Howard J. Dresner, zamestnanec analytickej spoločnosti Gartner Group, prvýkrát použil termín Business Intelligence, ktorý charakterizoval ako „súbor metód a konceptov určených na skvalitnenie rozhodovacích procesov firmy“.[[1], str.13]

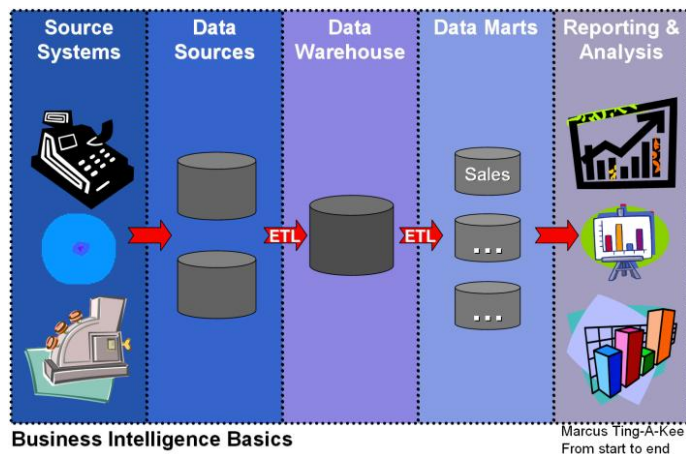
V osemdesiatych a deväťdesiatych rokoch 20. storočia nastal veľký rozmach informačných technológií a s tým súvisiaci nárast množstva dát, ktorými boli firmy zahltené. Klasický IT softvér dokázal tieto dáta technicky spracovať, ale neposkytoval informácie potrebné na podporu manažérskych úloh a rozhodovacích procesov. Vznikla teda potreba nového softvéru, ktorý je technicky vyspelý a tiež biznisovo orientovaný. Rozumie dátam, má technické schopnosti a využíva nové postupy a technológie. Časom sa tak vyvinuli nástroje a riešenia zastrešované spoločným pojmom Business Intelligence.

Doposiaľ nie je u nás zavedená oficiálna definícia termínu BI. Avšak existuje mnoho neformálnych definícií napr. [podľa serveru [5]]:

Business Intelligence predstavuje súhrn procesov a technológií slúžiacich na zber, skladovanie, analýzu a ďalšie poskytovanie informácií o organizácii najmä jej riadiacim pracovníkom a analytikom, s cieľom uľahčiť a zlepšiť rozhodovacie a riadiace procesy. Riešenia BI fungujú na princípe dodania správnych informácií (v podobe analýz, reportov, štatistík atď.) v správny čas (najčastejšie v reálnom čase prostredníctvom Internetu) do správnych rúk.[[1], str.9]

1.2 Ako sa buduje BI

Budovanie komplexnej firemnej BI sa realizuje v piatich základných fázach [spracované podľa [6]].



Obr. 1.1: Schéma znázorňujúca spracovanie dát v BI, zdroj [6]

1. Zdrojové systémy zhromažďujú údaje o spoločnosti.
2. Informácie zo zdrojových systémov sa ukladajú do databáz. Obyčajne sú tieto dáta vhodné na rýchle spracovanie technológiou OLTP, ale nie na tvorbu analýz.
3. Informácie z dátových zdrojov prejdú procesom nazývaným ETL, kde sa dáta vyberú, pretransformujú pre potreby spoločnosti a uložia do dátového skladu.
4. Z dátových skladov sa informácie ukladajú do datamartov (menších databáz), kde sú organizované, aby boli schopné odpovedať na špecifické otázky týkajúce sa danej oblasti (napr. predaja).
5. V poslednej fáze sa na spracovanie informácií z datamartov používajú analytické a reportovacie nástroje.

[1], str.15]

1.3 Nástroje BI

Nástroje BI predstavujú aplikačný softvér určený na spracovanie dát od zdrojových systémov až po koncového užívateľa. Implementácia riešení BI je kombináciou rozličných nástrojov, pričom ich podoba, rozsah a zložitosť sú dané potrebami a

podmienkami konkrétnej spoločnosti. Podľa hierarchie spracovania dát ich môžeme rozdeliť na:

1. **Nástroje dátovej transformácie**

ETL - systémy na transformáciu dát z vonkajších zdrojov a ich uloženie do dátových skladov.

EAI (Enterprise Application Integration) - systémy na integráciu primárnych podnikových systémov a redukciiu ich vzájomných rozhraní. Ide o dátovú integráciu (distribúcia dát) a aplikačnú integráciu (zdieľanie funkcií informačného systému).

2. **Databázové nástroje**

Dátové sklady, datamarty, operatívne dátové úložiská (Operational Data Store), dočasné úložiská dát (Data Staging Area), OLTP.

3. **Analytické nástroje**

DataMining, Reporting, OLAP.

4. **Nástroje pre koncových používateľov**

Mali by byť ľahko obsluhovateľné pre bežných užívateľov tak, aby pri ich využívaní nepotrebovali asistenciu IT špecialistov. Patria sem napr. dotazovacie nástroje, nástroje na tvorbu dashboardov a prezentácií.

[[1], str.17,18]

Kapitola 2

Data Warehousing

Čo je to data warehousing?

Data warehousing je proces zbierania a kombinovania dát z viacerých, obyčajne rozdielnych zdrojov, do jednej súhrnnej a jednoducho ovládateľnej databázy, teda je to proces vytvárania data warehouse. Data warehousing zahŕňa nástroje business intelligence, nástroje ETL, nástroje na správu metadát a iné.

Na čo slúži data warehousing?

„Aké sú náklady spoločnosti?“ „Aké sú príjmy?“ „Znížili naše posledné opatrenia náklady?“

Odpovede na tieto otázky musia byť komplexné a jasne vychádzajúce z firemných dát, no v praxi sa informačná štruktúra väčšiny spoločností skladá z viacerých rozdielnych systémov (systémy pre zákazníkov, zamestnancov, predaj, produkciu, financie, rozpočet). Tieto systémy sú zväčša málo integrované, a tak vznikajú hádky o tom, ktoré dáta sú tie správne. Data warehousing však systémy nepretvára, snaží sa dáta zobrazovať konzistentne, integrovane a konsolidovane bez ohľadu na zdrojové systémy.

Čo je to data warehouse?

Dátový sklad (data warehouse, DW) je rozsiahla štruktúrovaná databáza firemných dát, základný a najdôležitejší prvok BI zostrojený na uľahčenie reportingu a analýz. Existuje viacero definícií DW, no najznámejšie sú dve, ktoré zadefinovali priekopníci data warehousingu Bill Inmon a Ralph Kimball.

Definícia DW podľa Billa Inmona, 1990

DW je subjektovo-orientovaný, stály, integrovaný a časovo-rozlíšený súhrn dát dôležitých pre podporu riadiacich a rozhodovacích procesov.

Subjektovo-orientovaný: Dáta v DW sú organizované tak, že všetky dáta týkajúce sa jedného objektu v reálnom svete sú navzájom prepojené (napr. dáta o zamestnancoch, predaji atď.).

Stály: Dáta v DW sú neprepísateľné, nezmazateľné, statické, read-only, a udržiavané pre budúci reporting.

Integrovaný: DW obsahuje dáta z väčšiny alebo všetkých transakčných systémov organizácie a tieto dáta sú v DW konzistentné.

Časovo-rozlíšený: Dáta v DW sú identifikovateľné podľa jednotlivých časových období; v DW sú uložené aj historické dáta (transakčné systémy uchovávajú napríklad iba najnovšie adresy zákazníkov, ale v DW sú uložené všetky adresy spájajúce sa s jedným zákazníkom).

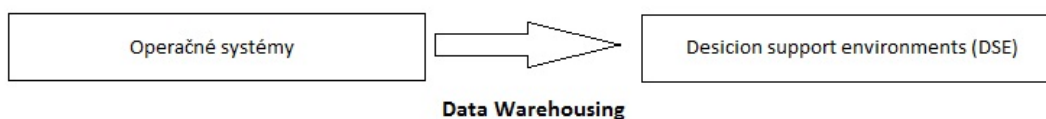
Definícia DW podľa Ralpa Kimballa

Ralph Kimball zaviedol oveľa jednoduchšiu definíciu DW vo svojej knihe *The Data Warehouse Toolkit*, kde píše, že DW je kópiou transakčných dát špeciálne štruktúrovaných pre dotazy a analýzy. Na rozdiel od Inmona, Kimball neuvádza ako by mal byť DW navrhnutý skôr sa sústreďuje na funkčnosť a tvrdí, že DW musia byť navrhnuté tak, aby boli zrozumiteľné a rýchle.

2.1 Ako vznikol data warehousing

História

Vznik data warehousingu sa datuje do 80. rokov 20. storočia, keď výskumníci firmy IBM Barry Devlin a Paul Murphy vytvorili „business data warehouse“. Základnou úlohou data warehousingu bolo poskytnúť model pre tok dát medzi transakčnými systémami a prostrediami na podporu rozhodovania (DSE). Vo väčších



Obr. 2.1: Schéma DSE

spoločnostiach bolo totiž bežné, že DSE pracovali nezávisle, a i keď každé z týchto prostredí používali rôzni užívatelia, často využívali rovnaké zdrojové dáta. A tak proces zbierania, očisťovania a integrovania dát, väčšinou z operačných systémov, bol opakovaný niekoľkokrát pre každé prostredie zvlášť. Data warehousing sa snažil riešiť tento problém a znižovať tak náklady firmy.

Vývoj data warehousingu [podľa [10]]

60. roky 20. stor. General Mills a Dartmouth College v spoločnom výskume zaviedli termíny „dimenzie“ (dimensions) a „fakty“ (facts).

70. roky 20. stor. ACNielsen a IRI poskytovali priestorové datamarty pre maloobchodníkov.

1983 Teradata predstavili databázový riadiaci systém špeciálne zostrojený pre podporu rozhodovania.

1988 Barry Devlin a Paul Murphy zverejnili článok *An architecture for a business and information systems* v IBM Systems Journal, v ktorom predstavili termín „business data warehouse“.

1990 Red Brick Systems predstavilo Red Brick Warehouse, špecializovaný systém pre správu warehouseov.

1991 Prism Solutions predstavilo Prism Warehouse Manager, softvér pre rozvoj dátových warehouseov.

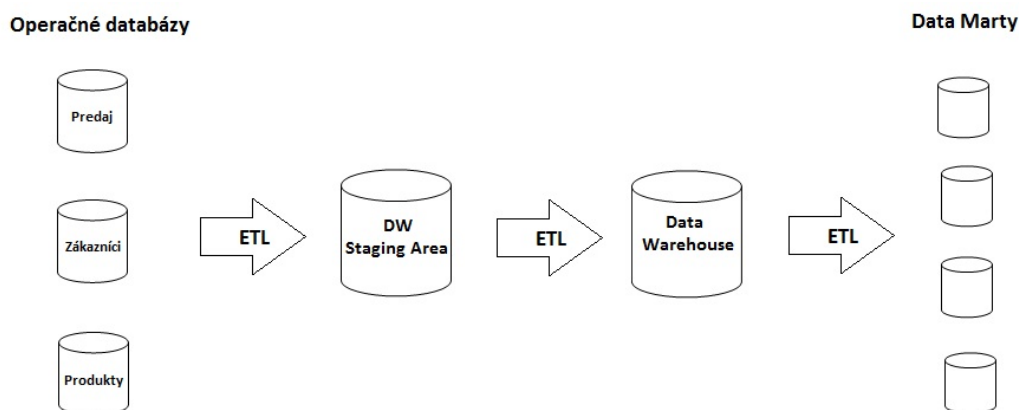
1991 Bill Inmon vydal knihu *Building the Data Warehouse*.

1995 Bol založený The Data Warehousing Institute (TDWI) - je prvou vzdelávacou inštitúciou v oblasti BI a data warehousingu, poskytuje vzdelanie, tréningy a výskum pre IT profesionálov.

1996 Ralph Kimball vydal knihu *The Data Warehouse Toolkit*.

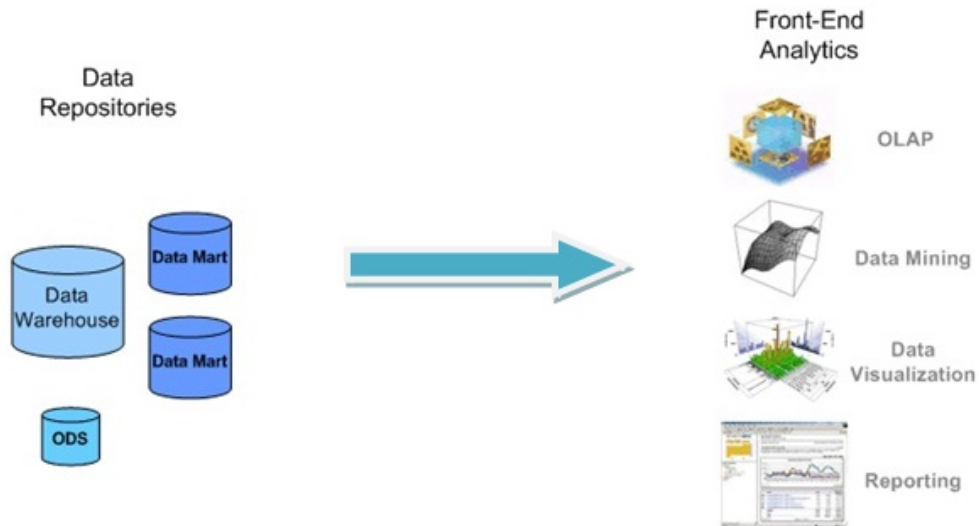
2000 Daniel Linstedt zostrojil Data Vault Modeling (Common Foundational Integration Modelling Architecture) - metódu tvorby databázy na uskladnenie historických dát prichádzajúcich z rôznych operačných systémov s kompletným trasovaním odkiaľ dáta v databáze pochádzajú.

2.2 Ako sa vytvára DW



Obr. 2.2: Procesná schéma dátového skladu

Základnou myšlienkou data warehousingu je očistiť a transformovať dáta z rôznych zdrojov pomocou ETL a urobiť ich jednoducho dostupné pre ďalšie spracovanie - datamining, OLAP, reporting atď.



Obr. 2.3: Front-end analytics, zdroj [12]

Vlastnosti	OD	DW
Čas odozvy	Zlomky sekundy-sekundy	Sekundy-hodiny
Operácie	Vkladanie, aktualizácia, mazanie	Čítanie
Pôvod dát	30-60 dní	Séria dát za čas. úsek
Organizácia	Podľa aplikácie	Podľa predmetu, času
Veľkosť	Malá až veľká	Veľká až veľmi veľká

Obr. 2.4: Základné rozdiely vo vlastnostiach operačnej databázy a DW, zdroj [4]

2.2.1 Normalizovaný a dimenzionálny prístup k dátam

Existujú dva základné prístupy k uskladneniu dát v data warehouseoch.

Normalizovaný prístup

Normalizácia je rozčlenenie dát za účelom vyhnúť sa duplikácii a prebytku dát premiestnením opakujúcich sa skupín dát do nových tabuliek. Normalizácia teda

zvyšuje počet tabuliek, ktoré musia byť navzájom prepojené, čím sa redukuje miesto potrebné na uskladnenie dát a tiež náročnosť aktualizácie dát.

Pri normalizovanom prístupe sú dáta v DW uložené v tabuľkách v normalizovanej forme, podľa normalizačných pravidiel. Tabuľky sú potom zoskupené podľa jednotlivých oblastí (napr. zákaznícke dáta, produkty, financie). Keď databáza používa model, ktorý je vysoko normalizovaný, dáta z jednej operácie sú často uskladnené vo väčšom množstve tabuliek. To umožňuje aby boli DW veľmi efektívne, pretože pri priebehu nejakej transakcie je iba malá časť dát ovplyvnená.

Hlavnou výhodou tohto prístupu je jednoduchosť priameho pridávania dát do databázy a flexibilita v prípade, že sa zmení spôsob organizácie spoločnosti alebo dátový model.

Nevýhodou je, že kvôli veľkému počtu tabuliek je náročné spojiť dáta z rôznych zdrojov tak, aby ponúkali zmysluplné informácie a boli vhodné na tvorbu analýz. Navyše, prístup k informáciám je zložitý, pokiaľ nerozumiete dobre zdrojom dát a dátovej štruktúre DW.

Dimenzionálny prístup

Pri dimenzionálnom prístupe sú dáta rozložené do *faktov* a *dimenzií*. Napríklad pri operáciách z predaja fakty predstavujú počet objednaných produktov a zaplatenú cenu, dimenzie sú dátum objednávky, meno zákazníka, číslo produktu, dodávacia adresa, predajca zodpovedný za vybavenie objednávky a pod.

Výhodou takéhoto prístupu je, že pre užívateľov je ľahšie mu porozumieť a používať ho, dáta sú ľahko prístupné, systém je jednoduchý na analyzovanie (slice & dice, roll up, drill down, drill up) a na správu a interpretáciu.

Nevýhodami sú dlhý čas a zložitost' nahratia dát do systému (za účelom zachovania integrity faktov a dimenzií) a obmedzená flexibilita v prípade, že sa zmení spôsob organizácie spoločnosti alebo sa menia dimenzie (je zložitá upravovať štruktúru DW).

Tieto dva prístupy sa vzájomne nevylučujú a existujú aj iné, kombinované prístupy (napr. dimenzionálne prístupy zahŕňajú normalizovanie dát do určitého stupňa). Podľa účelu databázy sa prispôsobuje aj jej štruktúra, a teda vo väčšine prípadov sú kombinované (hybridné) prístupy najvhodnejšie.

2.2.2 Top-down a bottom-up metodológia

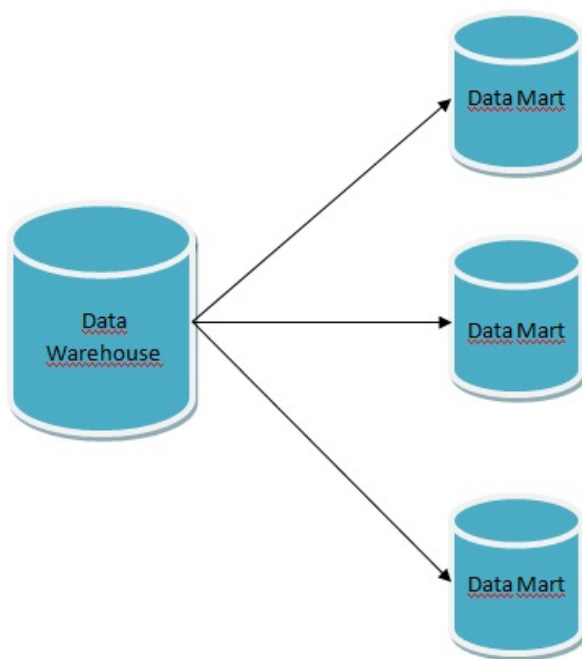
V tejto časti budeme hovoriť o metodológiách tvorby DW, ktorými sú top-down metodológia, bottom-up metodológia, hybridná metodológia a federatívna metodológia.

Top-down metodológia (The Dependent Data Mart structure)

Každé oddelenie firmy spracováva dáta iným spôsobom a DW by mal byť schopný odpovedať na požiadavky každého oddelenia. Toto je hlavnou myšlienkou top-down metodológie a jedným z jej vedúcich zástancov je Bill Inmon.

Dáta sú najskôr premiestnené z OLTP systémov do DW staging area, kde sú transformované a agregované. Následne sa prenášajú do data warehouse, kde sú zväčša uložené v tretej normalizovanej forme (3NF). Z dát sú potom vytvorené datamarty. Každý datamart je prispôbený pre individuálne potreby oddelenia a optimalizovaný na tvorbu analýz pre oddelenie, pre ktoré bol vytvorený. Top-down metodológia vytvára vysoko konzistentné dimenzionálne náhľady na dáta, pretože datamarty sú vytvárané z centrálného dátového skladu. DW teda slúži ako zdroj dát pre nové datamarty.

Top-down metodológia je robustná voči zmenám v štruktúre spoločnosti, pretože vygenerovanie nových dimenzionálnych datamartov je pomerne ľahká úloha. Jej hlavnou nevýhodou je zložitá implementácia.



Obr. 2.5: Top-down schéma

Bottom-up metodológia (The Data warehouse Bus Structure - BUS)

Takisto známa ako Kimballova metodológia vďaka jej zástancovi Ralphovi Kimballovi.

Kimball sa pozerá na dáta ako na kombináciu datamartov spojených do jedného data warehouseu tzv. busovou štruktúrou. V bottom-up dizajne sa vždy používa dimenzionálny prístup k uskladneniu dát, teda tabuľky obsahujú primárne dimenzie a fakty. Integrácia datamartov do DW prebieha pomocou *conformed dimenzií*, ktoré sú zdieľané medzi faktami a definujú možné integračné body medzi datamartami. Integrácia dvoch alebo viacerých datamartov je potom vykonaná procesom nazývaným *drill across* - sumarizuje dáta pomocou kľúčov zdieľaných conformed dimenzií každého faktu. V tomto prístupe nie je data warehouse fyzickým skladiskom dát ako v Inmonovom prístupe, ale virtuálnym skladiskom, pretože je iba kombináciou datamartov.

Považuje sa za výhodu Kimballovej metodológie, že DW sú rozdelené do množstva logických a konzistentných datamartov, a nie veľkého a často komplexného centrálného modelu. Ďalšou výhodou bottom-up prístupu je jeho rýchla implementácia.



Obr. 2.6: Bottom-up schéma

Hybridná metodológia

Hybridný prístup kombinuje výhody top-down metodológie (integrácia a celofiremná konzistencia dát) a bottom-up metodológie (rýchlosť vyhľadávania, jednoduchá užívateľská orientácia).

Federatívna metodológia

Táto architektúra sa označuje ako architektúra architektúr. Odporúča integráciu rozdielnych data warehouseov, datamartov a aplikácií, ktoré už existujú vo firme. Cieľom je integrovať existujúce analytické štruktúry, definovať dimenzie a kritéria na zdieľanie v rámci týchto štruktúr. Táto architektúra nie je veľmi elegantná, ale je jednoduchá na implementáciu v rámci firmy.

2.2.3 Budovanie dátových skladov v krokoch

Vytváranie dátového skladu prebieha vo 8-mich základných fázach [spracované podľa [20]]:

1. Vývoj realizačnej štúdie

Realizačná štúdia sleduje aktivity, náklady, výnosy a kritické faktory biznisu, tak aby bol implementovaný systém v budúcnosti úspešný. DW bude vystavaný v niekoľkých vývojových iteráciách podľa taktických a strategických požiadaviek biznisu. Sú tu definované dlhodobé aj krátkodobé ciele a identifikované náklady tak, aby sa splnil rozpočtový plán.

Realizačná štúdia sa ďalej predkladá manažérom, ktorí sú za projekt zodpovedný a zohrávajú kľúčovú úlohu v tejto fáze, lebo dobre poznajú biznis. Je výhodné spolupracovať s manažérmi z oblasti biznisu aj z oblasti IT. V tejto fáze sú takisto rozdelené úlohy a funkcie medzi ľuďmi, ktorí na projekte pracujú, čím sa jasne určia vzťahy medzi členmi tímu.

2. Analýza podnikateľského prostredia

V tomto bode sa sleduje dosiahnutie globálneho pohľadu na aktivitu organizácie a požiadavky užívateľov, určujú sa priority v požiadavkách užívateľov, obsah DW, identifikujú ciele a smery biznisu, určujú potrebné dáta na splnenie požiadaviek. Obsah DW sa určuje komunikáciou s koncovými užívateľmi a identifikáciou potrebných informácií, ktoré môžu pomôcť koncovým užívateľom. Zhodnocujú sa:

- funkcie oddelenia, KPI a rozhodujúce faktory úspechu
- vyrábaný produkt, dodávatelia a zákazníci
- informačné systémy a aplikácie používané v organizácii, externé zdroje informácií a potreba ručnej transformácie dát
- úroveň granularity dát, obsah a frekvencia užívateľských požiadaviek
- perióda v akej musia byť nové dáta ukladané do DW

Zároveň musí byť vykonaná analýza zdrojov dát a ich kvality, analýza technologickej architektúry organizácie, vzťahy medzi rôznymi systémami a ich vzájomná integrácia.

3. Návrh architektúry DW

Architektúra je logický základ, na ktorom je DW vystavaný. Sú identifikované komponenty, ktoré už v organizácii existujú a chýbajúce komponenty sú vybudované alebo prevzaté pre kompletizáciu DW. Architektúra DW musí byť navrhnutá tak, aby bolo možné ju v budúcnosti ďalej rozvíjať bez väčších zásahov do existujúceho modelu. Musí prezentovať relačné a multidimenzionálne modely využité v DW, ich umiestnenie, prepojenie a prístupové nástroje. Je definovaná logická štruktúra: centrálny sklad dát celej organizácie, voliteľný operačný data store, datamarty a úložiská metadát.

Ďalej sa definujú 4 nasledovné architektúry:

Dátová architektúra má za úlohu organizovať dátové zdroje a definovať kvalitu dát aj metadát.

Aplikačná architektúra prezentuje softvérové súčasti DW.

Technická architektúra poskytuje infraštruktúru pre dátovú a aplikačnú architektúru. Zahŕňa serverové, sieťové, hardvérové a softvérové komponenty na pripojenie a komunikáciu.

Podporná architektúra zahŕňa nástroje na back-up, archiváciu a monitorovanie výkonu.

4. Výber technologických riešení

V tejto fáze sa vyberajú najvhodnejšie nástroje pre požiadavky biznisu definované v architektúre DW, pričom sa berie ohľad na technickú infraštruktúru organizácie.

Pri výbere hardvéru sa berie do úvahy dátový objem určený na uloženie do DW. Zároveň sú vyberané softvérové komponenty: operačné systémy, databázové manažment systémy, vývojové a analytické nástroje.

Podľa funkcie môžu byť nástroje rozdelené do nasledovných kategórií: extrakcia a transformácia dát, čistenie dát, ukladanie a obnovenie dát, prístup k dátam, ochrana dát, konfiguračný manažment, dátový back-up a obnova dát, monitorovanie výkonu, modelovanie dát a metadáta manažment.

5. Plánovanie iterácií projektu

V tejto fáze sa vypracováva detailná analýza. Sú identifikované a zdokumentované všetky obmedzenia v zdrojových systémoch. Tieto obmedzenia ovplyvňujú spôsob akým bude DW ďalej vyvíjaný. Vykonáva sa podrobný audit zdrojových systémov. Hlavnú úlohu pri ňom zohrávajú administrátori a IT

špecialisti, ktorí sa oboznamujú s kvalitou dát a rozhodujú, ktoré informácie budú posunuté do procesu ETL.

6. Modelovanie DW

V tomto kroku sa navrhuje fyzický model DW a sú definované metadáta. Schéma DW môže byť vytvorená na relačnom modeli s dátovou normalizáciou alebo na multidimenzionálnom modeli s denormalizáciou. Zároveň sa dáta zo zdrojových systémov prenesú do cieľových polí v DW pričom prechádzajú procesom ETL pred alebo počas prenosu. V tejto fáze sú vybudované aj aplikácie na podporu analýz, nástroje na tvorbu OLAP kociek, data mining, reporting, tvorbu web stránok a dashboardov.

7. Implementácia a testovanie DW

Keď sú návrhy hotové môže začať implementácia. Zavedú sa vývojové a testovacie rozhrania, inštalujú hardvérové a softvérové komponenty, rozhodne sa o delení DW (normalizovaný DW sa ľahšie spravuje, ale je menej výkonný), zrealizuje sa fyzický dizajn tabuliek faktov a dimenzií. Zároveň sú vyvinuté a nakonfigurované programy na extrakciu, čistenie, transformáciu, ukladanie a obnovu dát.

Uložia sa počiatočné dáta a technické a obchodné metadáta. Navrhne sa užívateľské rozhranie a reporty, spustia sa jednoduché ad-hoc dotazy na otestovanie databázy a preverí sa správnosť výsledkov. Zaškolia sa užívatelia a definuje sa prístup užívateľov k dátam.

Po prvých testoch urobených tímom vývojárov sú zapojení koneční užívatelia. Používajú systém tak, aby našli a opravili chyby, identifikovali potreby pre zlepšenie výkonu a zoznámili sa s novým systémom.

8. Spustenie

V tejto fáze sú programy pre extrakciu, čistenie, transformáciu a ukladanie dát spustené smerom k dátovým zdrojom. Definuje sa aj údržba a vývoj DW, čo zahŕňa periodické obnovovanie dát v DW, výpočet štatistických indikátorov za cieľom sledovania výkonu a zaťaženia DW (počet dotazov vykonaných za deň, priemerný čas odpovede systému, počet užívateľov za deň atď.), hodnotenie veľkosti DW, pričom veľkosť sa môže redukovať napríklad agregáciou dát.

2.3 ETL proces

ETL proces je nevyhnutný pri migrácii dát a zavádzaní DW a pre získavanie vysoko kvalitných dát v DW. Nasadzuje sa na dáta z heterogénnych zdrojov, ktoré

treba pred uložením do DW extrahovať, očistiť a upraviť.

Má tri etapy:

1. **Extrakcia (Extract)** - predstavuje výber dát zo zdrojových systémov,
2. **Transformácia (Transform)** - predstavuje overovanie, čistenie, integrovanie a časové označenie dát,
3. **Uloženie dát do DW (Loading)**

Existujú 2 modely prenášania dát zo zdrojov do DW:

- **Model lokálneho prenášania**, kde sa dáta spracovávajú priamo v operačných systémoch, a potom sú prenesené do DW (môže značne zaťažovať operačné systémy),
- **Model vzdialeného prenášania**, kde sa dáta vynesú do DW Staging area alebo priamo DW a tam sa spracujú.

Extrakcia

Úlohou extrakcie je získať dáta z rôznych zdrojov (operačných systémov, transakčných systémov OLTP, databázových systémov, archívnych systémov, atď.). Okrem interných firemných dát sa do DW ukladajú aj externé dáta napr. dáta o zákazníkoch, o konkurencii, cenové kurzy a indexy z Internetu. K dispozícii na extrakciu sú rôzne nástroje, programy a aplikácie.

Transformácia

Transformácia je súbor úkonov, ktoré vedú ku zvýšeniu kvality dát. Potreba transformácie je z dôvodu obmien technológií (napr. operačných systémov) alebo dôsledkom ľudského faktora (pravopisné chyby). Pri transformácii sa zjednocuje formátovanie dát, dátové typy, jednotky a meny. Príklady transformácie dát:

Potreba rozloženia údajov na atomické hodnoty

<i>Adresa</i>	<i>Ulica a číslo</i>	<i>Mesto</i>	<i>Smer. číslo</i>
<i>Staré Grunty 6 Bratislava 845 45</i>	<i>Staré Grunty 6</i>	<i>Bratislava</i>	<i>845 45</i>

Nejednoznačnosť dát

<i>Meno</i>	<i>Štát</i>	<i>Meno</i>	<i>Štát</i>
<i>Ševčík</i>	<i>SR</i>	<i>Ševčík</i>	<i>Slovenská republika</i>
<i>Valko</i>	<i>Slovenská republika</i>	<i>Valko</i>	<i>Slovenská republika</i>
<i>Malý</i>	<i>Slovensko</i>	<i>Malý</i>	<i>Slovenská republika</i>

Chýbajúce a duplicitné hodnoty

Duplicitné údaje sa dajú odstrániť, ale niekedy je ťažké ich odhaliť. Chýbajúce údaje sa dajú doplniť z iných zdrojov.

Nejednotnosť názvov pojmov a objektov

Napríklad: Krajina <-> Štát, Zamestnanie <-> Profesia

Nejednotnosť peňažných mien

Treba rozlišovať, v akej mene sú údaje uložené v databáze.

Nejednotnosť formátov čísel a textových reťazcov

Problém nastáva napríklad, keď je rodné číslo v zdrojovom systéme uložené ako textový reťazec vo formáte 875331/5234 a do databázy zavedené ako číslo, čo by systém bral ako matematickú operáciu $875331/5234 = 167,24$.

Chýbajúci dátum

Dátum je v DW dôležitá hodnota, musí byť prítomný v dátach pred ich zavedením do DW alebo sa musí pridať pri ukladaní dát do DW.

Uloženie dát do DW (Loading)

Predstavuje prenos dát a uloženie do databázových tabuliek. Mal by byť plánovaný, automatizovaný a maximálne optimalizovaný. Na začiatku ide o veľký objem dát a neskôr sa nové dáta ukladajú periodicky (každý deň, týždeň), prípadne sa aktualizujú. Po uložení sa dáta indexujú pre ich jednoznačnú identifikáciu.

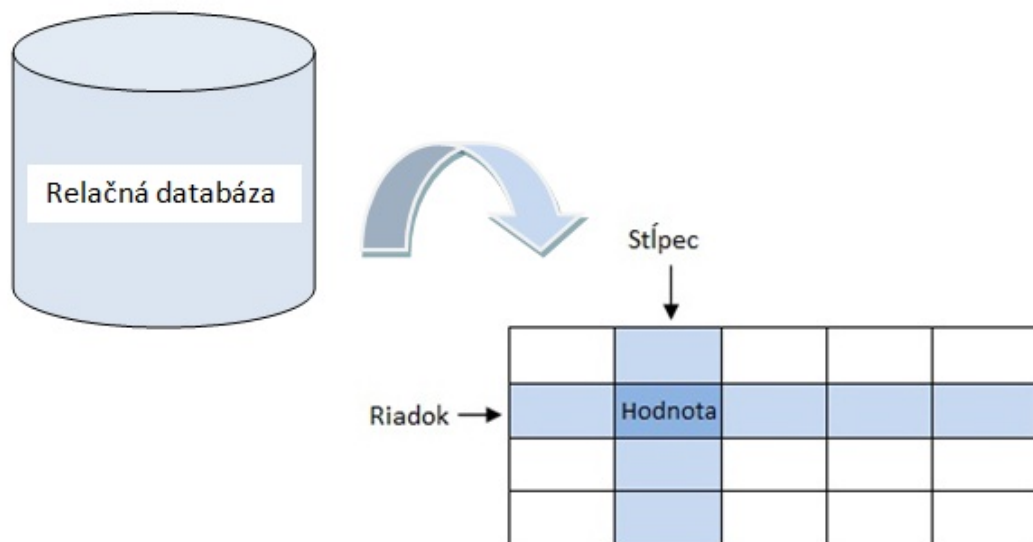
Vo finálnej fáze je potrebné etapu ETL dôkladne otestovať na simulovaných a neskôr na reálnych dátach, pretože tvorí dôležitú súčasť zavádzania kvalitných dát do DW.

Kapitola 3

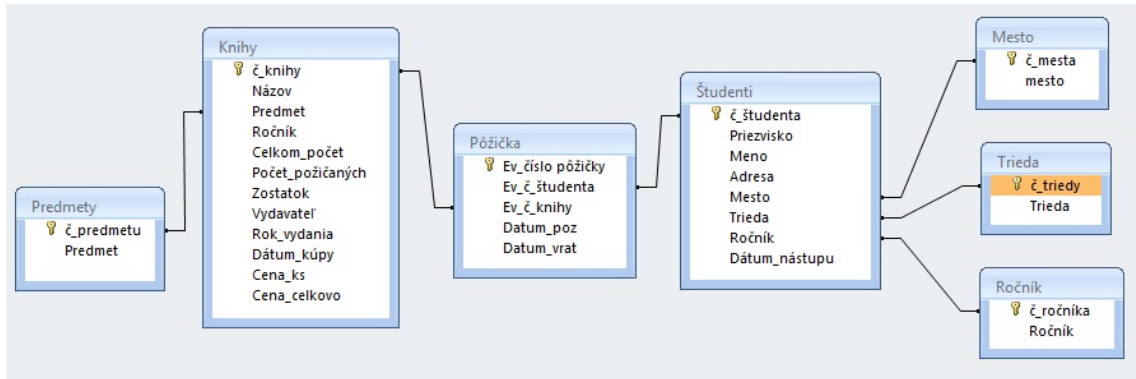
OLAP

3.1 Relačná databáza

Relačné databázy sa nazývajú aj transakčné databázy alebo operačné databázy. Využívajú technológiu OLTP (Online Transaction Processing), kde sú dáta uložené v databázových tabuľkách, medzi ktorými sú definované relačné vzťahy. Základy teórie relačných databáz položil E.F.Codd, ktorý definoval 3 podmienky minimálnej relačnosti a 12 pravidiel pre relačné databázové systémy.



Obr. 3.1: Relačná databáza



Obr. 3.2: Vzťahy medzi tabuľkami v relačnej databáze

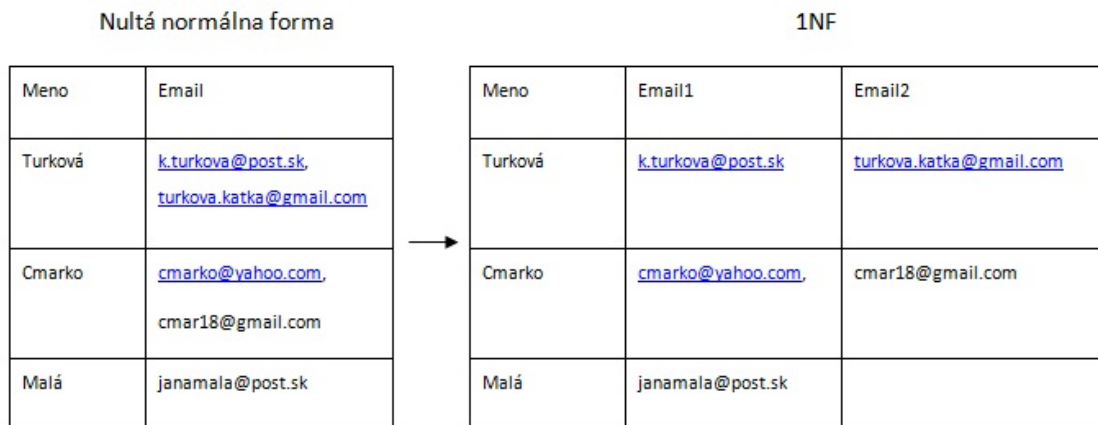
Položky označené malým kľúčom na obrázku hore predstavujú primárny kľúč a slúžia na jednoznačnú identifikáciu každého záznamu. Položka, s ktorou je primárny kľúč prepojený v inej tabuľke obsahuje cudzí kľúč. Bez primárneho kľúča nie je možné definovať relácie medzi tabuľkami.

Normalizácia databázy

Normalizácia predstavuje rozloženie tabuliek podľa určitých pravidiel. Znižuje prebytok dát a zefektívňuje prácu s databázami. Využíva sa najmä v relačných databázach, pretože v multidimenzionálnej databáze je prístup k normalizovaným dátam príliš zložitý.

Prvá normálna forma (1NF)

Tabuľka je v prvej normálnej forme, ak sú všetky jej stĺpce ďalej nedeliteľné. Napríklad:



Druhá normálna forma (2NF)

Tabuľka je v druhej normálnej forme, keď spĺňa podmienku 1NF a každý atribút

okrem primárneho kľúča je úplne závislý na celom primárnom kľúči.

Tretia normálna forma (3NF)

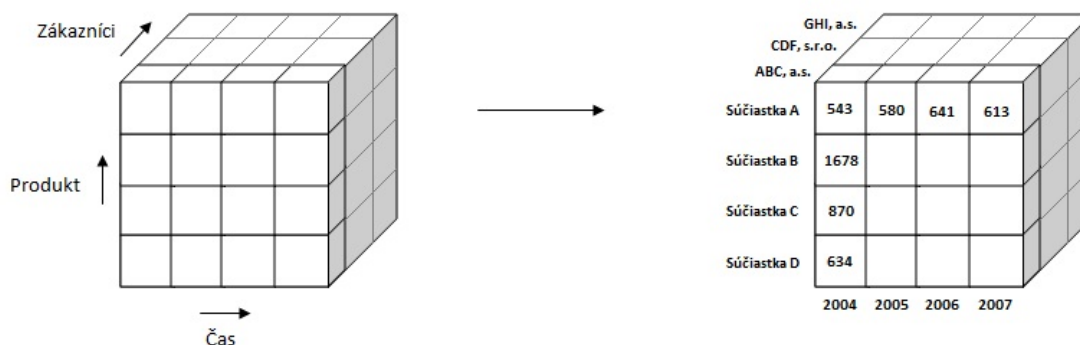
Tabuľka je v tretej normálnej forme, keď spĺňa podmienku 2NF a zároveň neexistujú závislosti neklúčových stĺpcov tabuľky.

3.2 Multidimenzionálna databáza

Multidimenzionálna databáza slúži na analýzu veľkého množstva dát, ktorej výsledkom sú reporty poskytujúce súhrnné i detailné informácie o behu organizácie, umožňuje multidimenzionálny pohľad na dáta. Hlavnou výhodou multidimenzionálnej databázy je, že dokáže odpovedať na dotazy používateľov oveľa rýchlejšie ako relačné databázy.

Na vytváranie týchto databáz sa využíva technológia OLAP (Online Analytical Processing). Základným princípom je vytvorenie multidimenzionálnej dátovej štruktúry tzv. OLAP kocky. Na výpočet OLAP kociek je potrebné vykonať veľké množstvo výpočtov a agregácií.

Fyzicky sa vlastne nejedná o kocku, pretože každá OLAP kocka môže mať ľubovoľne veľa dimenzií. „Fyzicky sú dáta uložené v tabuľkách relačnej databázy a multidimenzionálna štruktúra je dosiahnutá relačnými väzbami medzi tabuľkami faktov a tabuľkami dimenzií.“ [[3], str. 173]



Obr. 3.3: Príklad OLAP-kocky zloženej z troch dimenzií: dimenzie času, produktov a zákazníkov

Štandardne býva jedným z ukazovateľov v OLAP kocke čas a ostatné sú ekonomické premenné, geografické premenné, zákazníci, dodávatelia a pod. Premietnutie všetkých dimenzií do jedného bodu tvorí prvok multidimenzionálnej databázy.

Hlavným priekopníkom technológie OLAP bol takisto E.F.Codd, ktorý zaviedol aj nasledovných dvanásť pravidiel OLAP:

1. Multidimenzionálny pohľad na dáta vhodný pre tvorbu analýz,
2. Transparentnosť,
3. Dostupnosť; teda prístup ku všetkým dátam potrebným na tvorbu analýz nezávisle na tom, z akého zdroja pochádzajú,
4. Konzistentný výkon; aj keď sa počet dát v DW zväčšuje každý deň, užívateľ by nemal pociťovať zníženie výkonu systému,
5. Klient-server architektúra,
6. Generická dimenzionalita; všetky dimenzie musia byť ekvivalentné v štruktúre a operačných schopnostiach,
7. Ošetrovanie riedkych matic,
8. Multiužívateľská podpora,
9. Neobmedzené cezdimenzionálne operácie; musí vedieť rozoznať hierarchie dimenzií a vykonať kumulované výpočty v rámci dimenzií aj medzi nimi,
10. Intuitívna manipulácia s dátami; musí umožňovať drill-down a drill-up pre užívateľa jednoduchým spôsobom „ukázať a kliknúť“ alebo „zachytiť a premiestniť“,
11. Flexibilný reporting; musí umožňovať analýzu vizuálnou prezentáciou analytických zostáv,
12. Neobmedzené dimenzie.

3.2.1 Tabuľky faktov a tabuľky dimenzií

Multidimenzionálna databáza sa skladá z OLAP kociek, na ktorých vytvorenie potrebujeme 2 druhy dát - fakty a dimenzie.

Tabuľky dimenzií obsahujú logicky a hierarchicky usporiadané dáta. Sú zvyčajne menšie ako tabuľky faktov a používajú sa na výber a agregáciu dát na požadovanej úrovni detailu. Príklady hierarchickej stromovej štruktúry jednotlivých dimenzií:

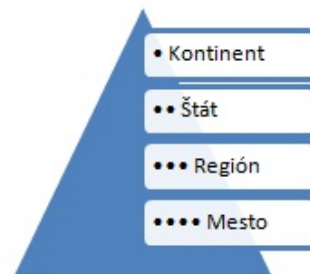
Časová dimenzia



Produktová dimenzia



Regionálna dimenzia



Tabuľka faktov býva najväčšou tabuľkou databázy a obsahuje numerické merné jednotky obchodovania. Vo väčšine modelov sú fakty definované kombináciou hodnôt dimenzií a obsahujú veľké množstvo agregovaných dát.

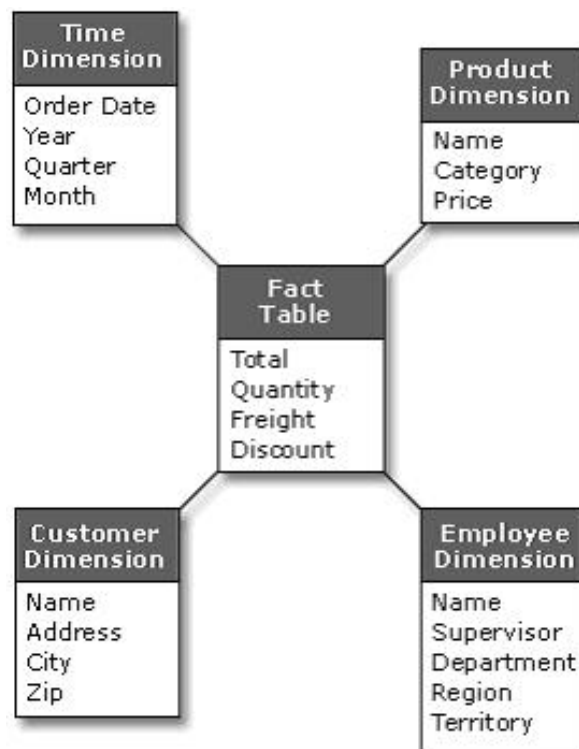
Granularita

Pod granularitou rozumieme úroveň podrobnosti dát uložených v tabuľkách faktov a dimenzií. Fáza vytvárania tabuľky faktov zahŕňa 2 kroky:

- 1. výber dimenzií, na základe ktorých bude tabuľka faktov vytvorená;**
ich výber závisí najmä od požiadaviek biznisu a užívateľov, časová a regionálna dimenzia bývajú zväčša zahrnuté.

2. určenie granularity dimenzií;

ak napríklad biznis vyžaduje robiť analýzy hodinového predaja tovaru, tak je najvhodnejšie za najnižšiu formu granularity zvoliť hodinu. Ak postačujú denné analýzy, tak dni. Čím je však väčšia granularita (dáta sú podrobnejšie), tým je väčšie množstvo prepočítaných dát v tabuľke faktov. Z praktického hľadiska je vhodné zvoliť vyššiu úroveň granularity, aby sa v budúcnosti, ak sa zmenia podmienky biznisu nemusel prerábať celý model.

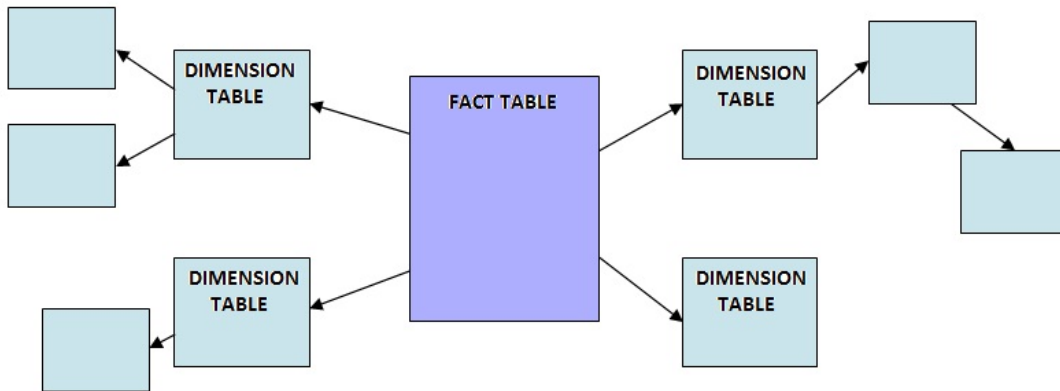


Obr. 3.4: Schéma tabuliek dimenzií a faktov, zdroj [22]

Na hornom obrázku je príklad usporiadania tabuliek multidimenzionálnej databázy. Táto kocka obsahuje 4 dimenzie: Čas, Zákazníci, Zamestnanci a Produkt. Dimenzia Zamestnanca v tomto prípade predstavuje aj geografickú dimenziu. Tabuľka faktov obsahuje agregované údaje - celkovú sumu, množstvo, náklady na dopravu alebo zľavy. Užívateľ môže v takto usporiadanej databáze rýchlo analyzovať napríklad počet predaných produktov podľa kategórii, za určité časové obdobie (rok, štvrtrok, mesiac), podľa jednotlivých zákazníkov alebo zamestnancov, regiónov alebo oblastí.

3.2.2 Usporiadanie tabuliek v multidim. databáze - schéma

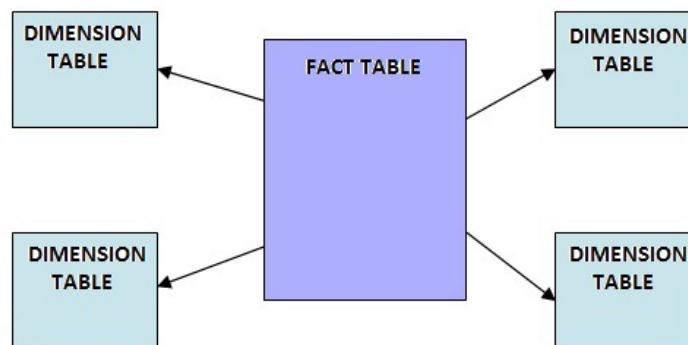
Snowflake schéma



Obr. 3.5: Snowflake schéma

Snowflake schéma je reprezentovaná centralizovanou tabuľkou faktov, ktorá je prepojená s tabuľkami dimenzií v normalizovanej forme.

Star schéma



Obr. 3.6: Star schéma

Je špeciálny prípad snowflake schémy a najjednoduchšia schéma tvorby warehouseov. Skladá sa z jednej alebo viacerých tabuliek faktov a prislúchajúcich tabuliek dimenzií. Na rozdiel od snowflake schémy, kde sú dimenzie normalizované do

viacerých prepojených tabuliek, tabuľky v star schéme sú denormalizované na jednu dimenziu reprezentovanú iba jednou tabuľkou.

3.3 Príklad pre OLAP analýzu

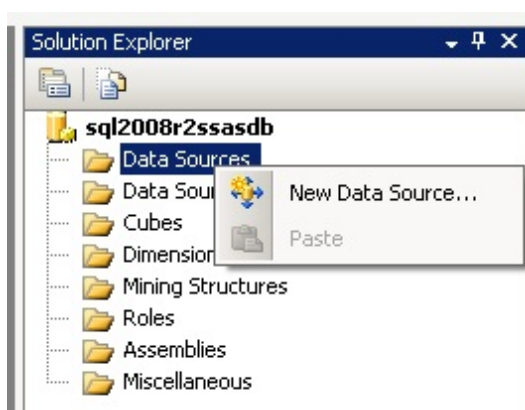
V tejto časti budeme ilustrovať príklad tvorby OLAP kocky zo stránky <http://www.kodyaz.com/articles/how-to-create-olap-cube-in-business-intelligence-development-studio.aspx>, pomocou služieb Microsoft SQL Server 2008 R2 použitím aplikácie SQL Server Business Intelligence Development Studio (BIDS). Nebudeme zachádzať do úplných podrobností, celú obrázkovú dokumentáciu a podrobný návod je možné nájsť na stránke.

Proces tvorby OLAP kocky prebieha v nasledovných štyroch krokoch:

1. Pripojenie/vytvorenie dátového zdroja (Data Source)
2. Vytvorenie pohľadu na zdroj dát (Data Source View)
3. Vytvorenie dimenzií kocky (Dimensions)
4. Vytvorenie OLAP kocky (Cube)

3.3.1 Pripojenie/vytvorenie dátového zdroja (Data Source)

V tomto kroku sa treba pripojiť na zdroj dát, teda databázu, z ktorej budeme vytvárať OLAP kocku.

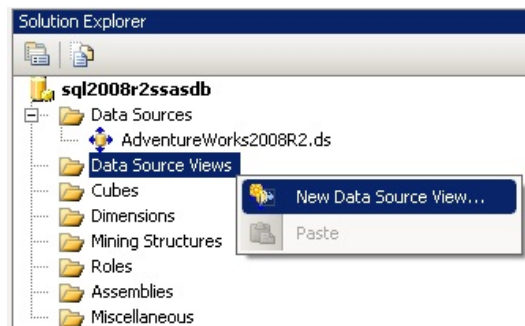


V našom prípade pôjde o vzorovú databázu Microsoft SQL Server 2008 R2 AdventureWorks. Pomocou sprievodcu nastavíme parametre a po ukončení môžeme dané pripojenie otestovať.

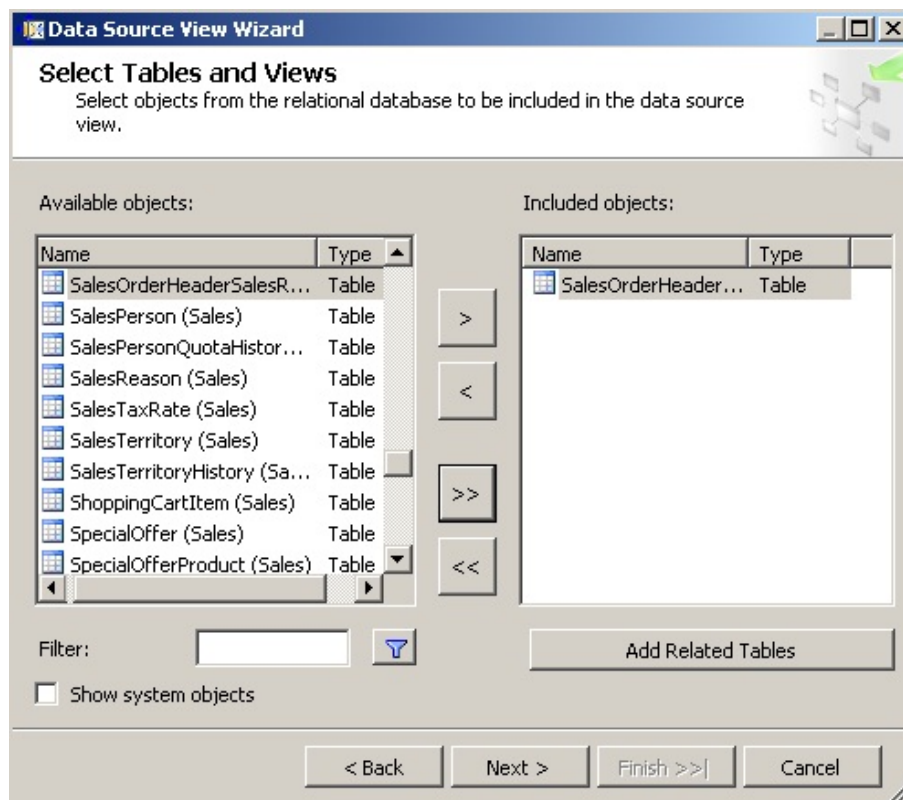


3.3.2 Vytvorenie pohľadu na zdroj dát (Data Source View)

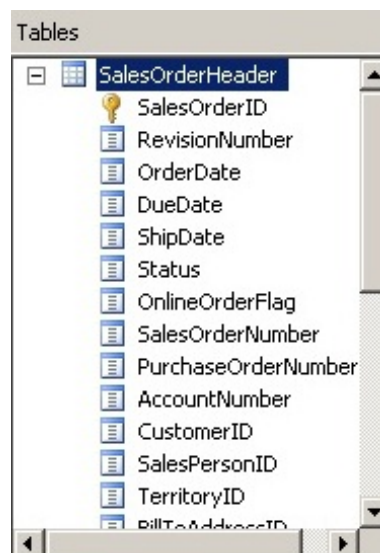
Po pripojení dát nastavíme pohľad na dáta.



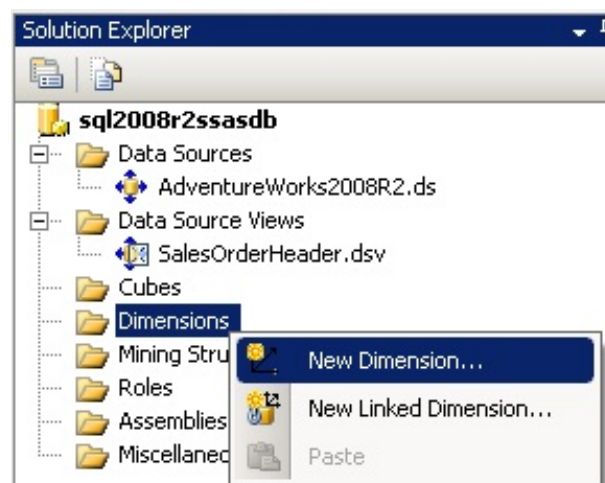
Spríevodca nám ponúkne skupinu prepojených tabuliek zdrojovej databázy.



Pre jednoduchosť budeme pracovať iba s jednou tabuľkou *SalesOrderHeader*, ktorá obsahuje informácie o predaji a budeme chcieť zobraziť obnosy predaja v jednotlivých časových obdobiach. Takže v tomto prípade bude tabuľka *SalesOrderHeader* tabuľkou faktov a jedinou tabuľkou dimenzií, ktorú budeme potrebovať bude časová dimenzia, ktorú vytvoríme v ďalšom kroku.
Položky tabuľky *SalesOrderHeader*:



3.3.3 Vytvorenie dimenzií kocky (Dimensions)



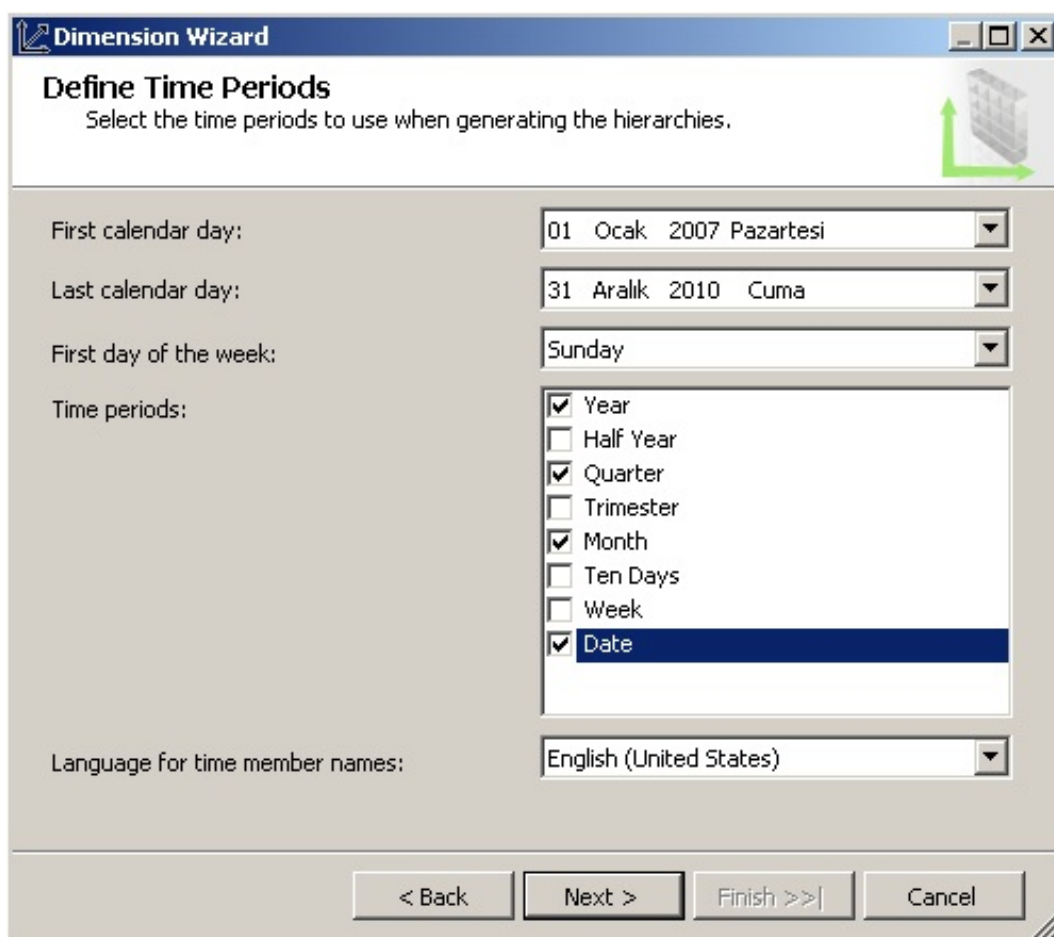
Po tom, ako sme pridali tabuľky faktov a dimenzií z dátového zdroja, môžeme vytvoriť nové dimenzie, ktoré budeme v našom modeli potrebovať. Je dôležité vytvárať dimenzie nezávisle od OLAP kociek, pretože jedna dimenzia môže byť zdieľaná vo viacerých OLAP kockách. Takto sa pri spracovaní dimenzie obnovia dáta vo všetkých

súvisiacich OLAP kockách. Zdieľané dimenzie teda urýchľujú proces spracovania dát.

Pomocou pomocníka máme na výber metódu akou bude dimenzia vytvorená:

- vytvorením novej tabuľky,
- vygenerovaním časovej tabuľky v dátovom zdroji,
- vygenerovaním časovej tabuľky na serveri,
- vygenerovaním nečasovej tabuľky v dátovom zdroji.

V našom prípade vyberieme možnosť vygenerovania časovej tabuľky na serveri, čím sa vytvorí časová dimenzia na serveri bez využitia zdrojových dát. Nasleduje výber časových období, čo predstavuje vytvorenie časovej hierarchie, na základe ktorej sa budú dáta zobrazovať v OLAP kocke.



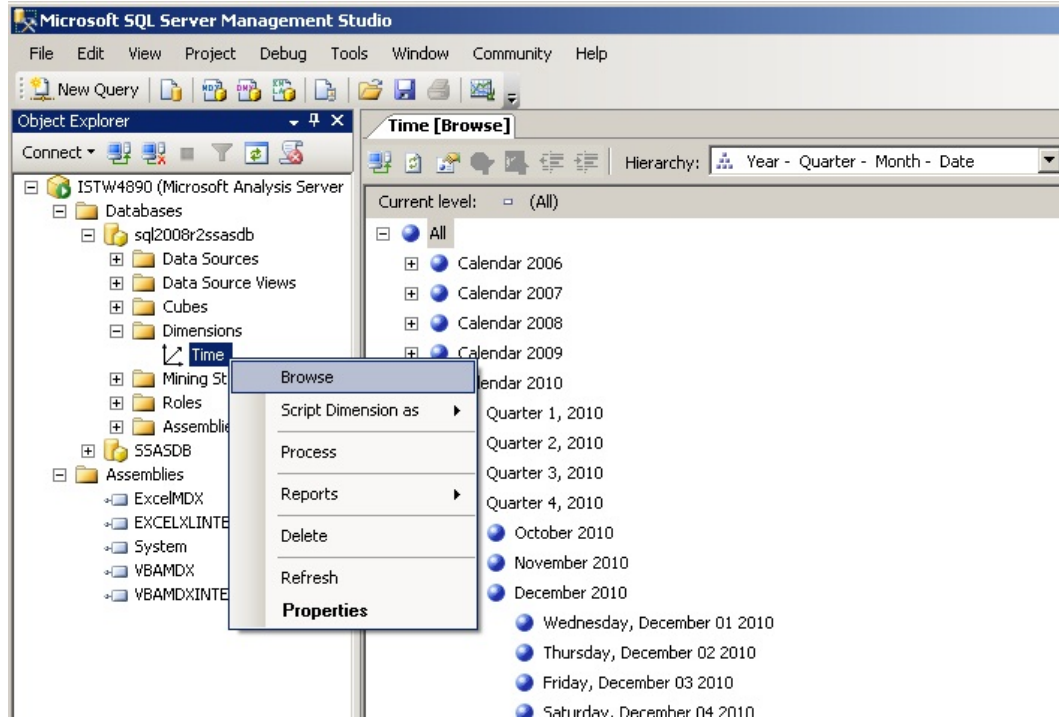
The screenshot shows the 'Dimension Wizard' dialog box, specifically the 'Define Time Periods' step. The title bar reads 'Dimension Wizard'. Below the title bar, the main heading is 'Define Time Periods' with the instruction 'Select the time periods to use when generating the hierarchies.' To the right of the instruction is a small 3D cube icon with a green arrow pointing upwards. The dialog contains several fields and a list:

- 'First calendar day:' dropdown menu showing '01 Ocak 2007 Pazartesi'.
- 'Last calendar day:' dropdown menu showing '31 Aralık 2010 Cuma'.
- 'First day of the week:' dropdown menu showing 'Sunday'.
- 'Time periods:' list with checkboxes:
 - Year
 - Half Year
 - Quarter
 - Trimester
 - Month
 - Ten Days
 - Week
 - Date (highlighted)
- 'Language for time member names:' dropdown menu showing 'English (United States)'.

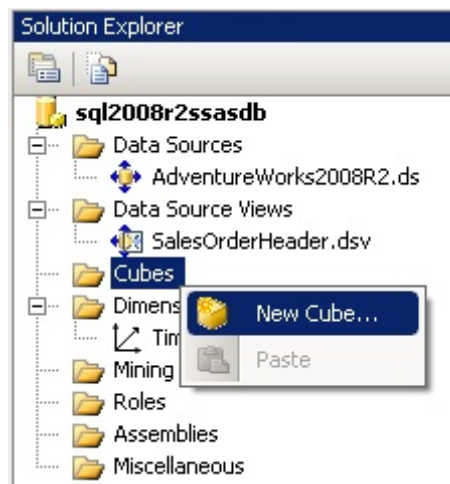
At the bottom of the dialog are four buttons: '< Back', 'Next >', 'Finish >>|', and 'Cancel'.

Zvolíme počiatočný a konečný deň, prvý deň týždňa, jazyk a samotnú hierarchiu časových období (Rok, Štvrťrok, Mesiac, Dátum). V nasledujúcom kroku sprievodcu je možné vybrať druh kalendára, pre ktorý sa má hierarchia vytvoriť:

bežný kalendár, fiškálny kalendár, marketingový kalendár, výrobný kalendár atď. Zaškrtneme možnosť bežného kalendára, pomenujeme dimenziu a ukončíme sprievodcu. Dimenzia je vytvorená, ďalej treba spustiť proces na jej naplnenie dátami a po jeho ukončení si môžeme časovú dimenziu prezerať podľa hierarchie.



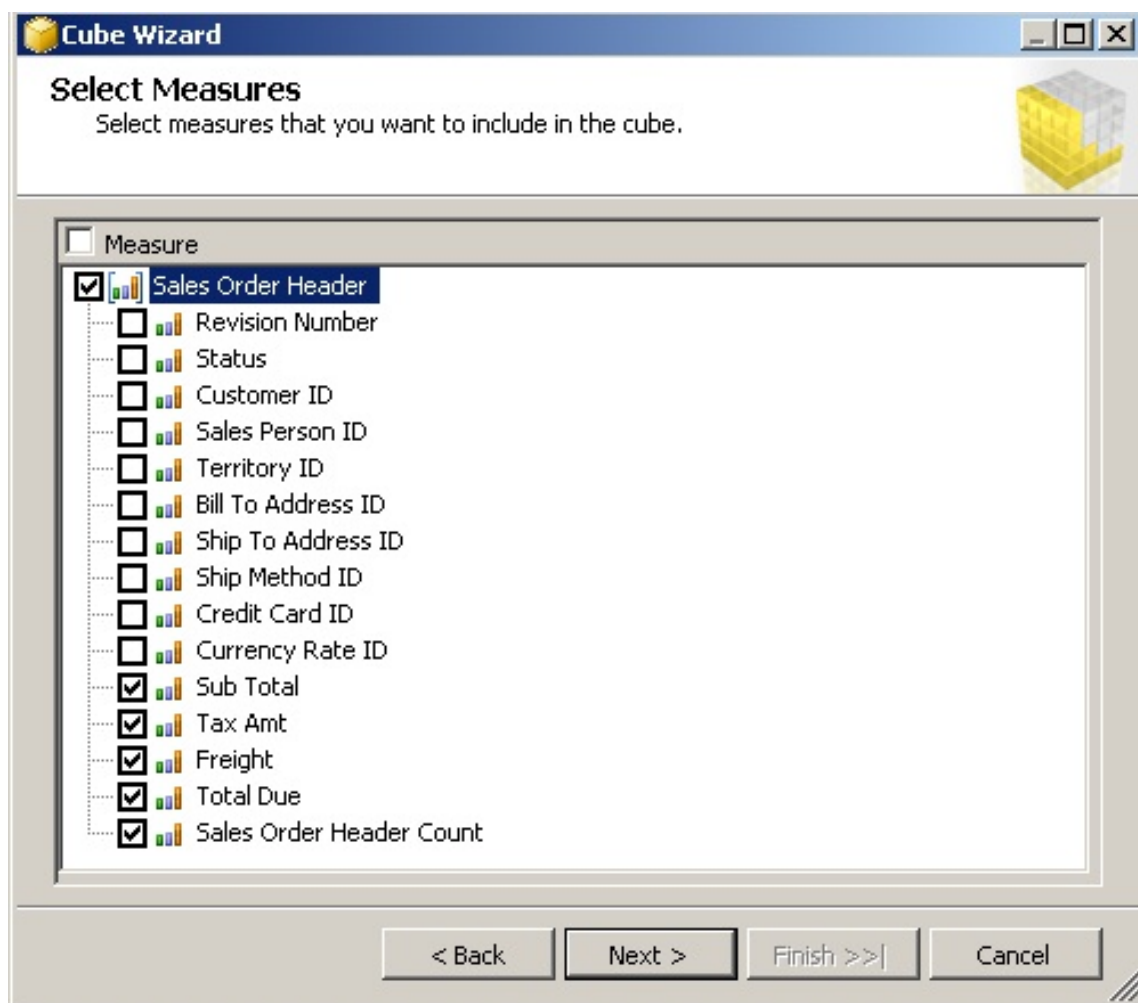
3.3.4 Vytvorenie OLAP kocky (Cube)



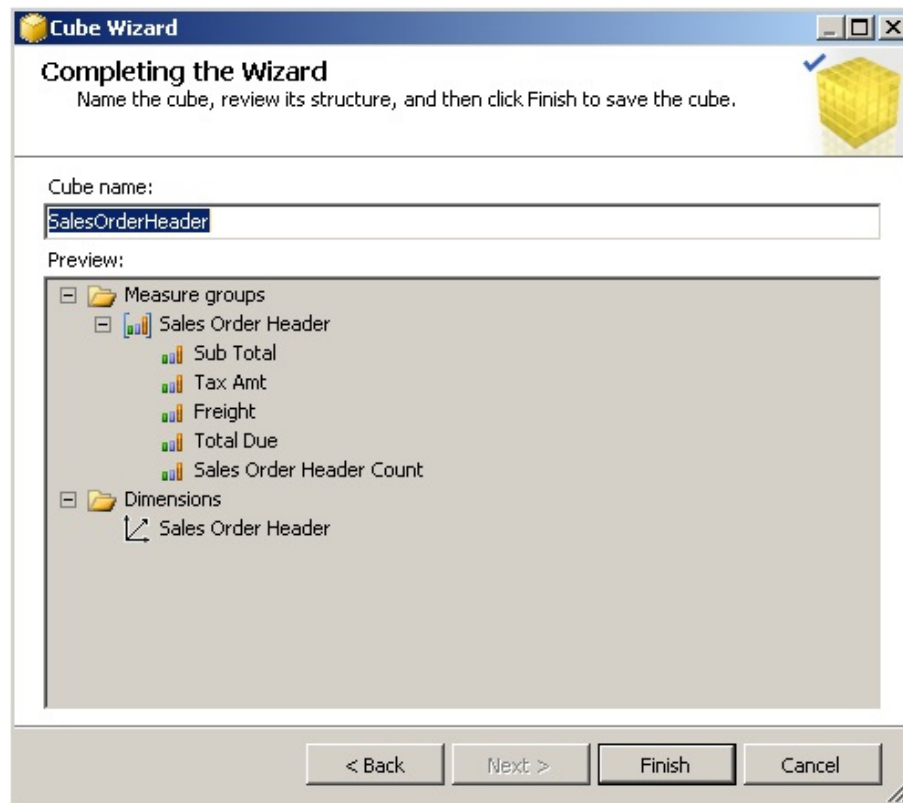
Pri tvorbe samotnej OLAP kocky budeme taktiež využívať sprievodcu, ktorý nám ponúka 3 možnosti tvorby OLAP kocky:

- vytvoriť kocku použitím existujúcich tabuliek,
- vytvoriť prázdnu kocku,
- vygenerovať tabuľky v dátovom zdroji.

Vyberieme prvú možnosť a v ďalšom kroku vyberieme pohľad na zdroj dát *Sales-OrderHeader* (z tých, ktoré sme už vytvorili), z ktorého sa budú počítať hodnoty OLAP kocky. Ďalej vyberieme, ktoré hodnoty z tohto pohľadu chceme zahrnúť do kocky:

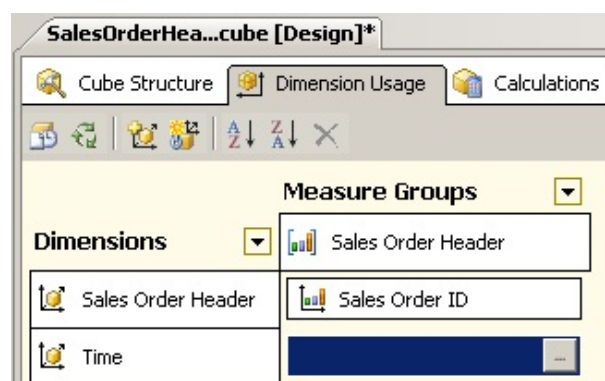


Určíme dimenzie, ktoré majú byť zahrnuté do modelu na základe existujúcich tabuliek. Potom ako sme definovali hodnoty a dimenzie, máme možnosť prezrieť si štruktúru OLAP kocky a pomenovať ju:



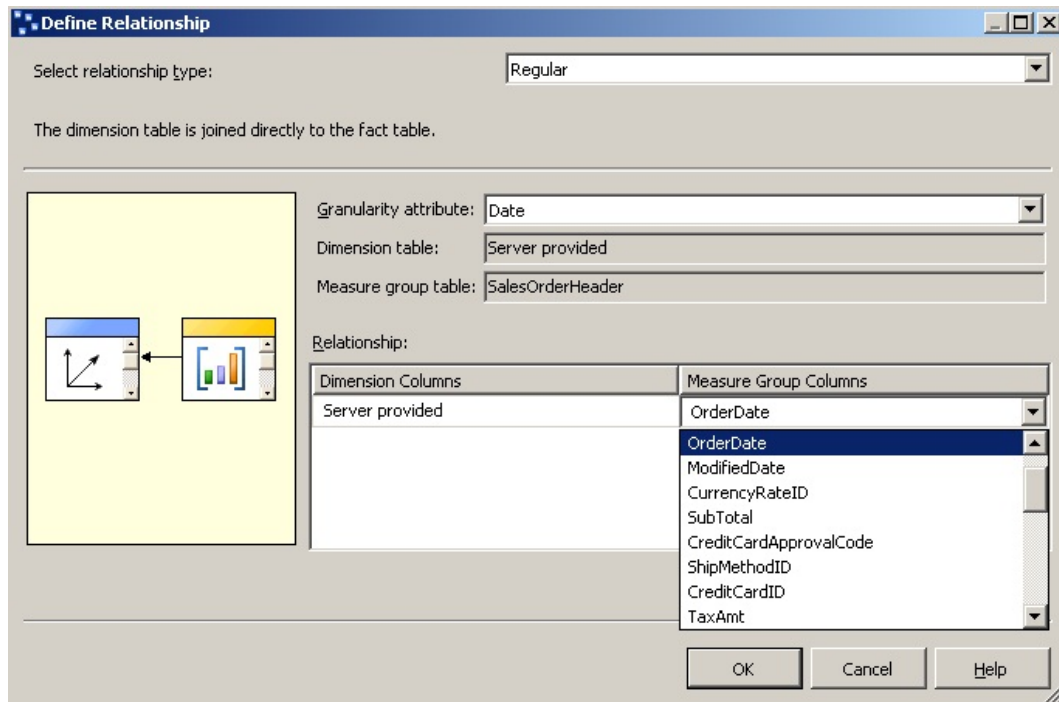
Po vytvorení kocky musíme ešte pridať časovú dimenziu do modelu, zo *SalesOrderHeader* však nebudeme využívať žiadne dimenzie.

Najdôležitejšou časťou tvorby OLAP kocky je konfigurácia prepojenia tabuliek faktov s tabuľkami dimenzií definovaním stĺpcov prepojenia a typov relácií v záložke Dimension Usage. Možné typy relácií: Regular, Fact, Referenced, Many-to-Many, No Relationship.



Zvolíme Regular a vyberieme stĺpce, ktoré budú vytvárať vzťah. V časovej dimenzii je už prednastavený stĺpec a z tabuľky faktov vyberieme stĺpec OrderDate

pre zobrazenie dátumu objednávky v časovej dimenzii OLAP kocky.



Takže prepojenie medzi dimenziami a faktami je zrealizované pomocou dátumu. Pre vytvorenie dát OLAP kocky musíme spustiť proces, pričom môžeme meniť nastavenia procesu.

3.3.5 Prezeranie OLAP kocky

Výslednú OLAP kocku si môžeme prezerať v záložke Browser.

Year	Quarter	Month	Date	Sub Total	Sales Order Header Cou
Calendar 2006				11311.8478	7
Calendar 2007				42011037.1641981	12443
Calendar 2008	Quarter 1, 2008			11398376.2784007	6087
	Quarter 2, 2008	April 2008		3820583.4921001	2128
		May 2008		5194121.52290014	2386
		June 2008		5364840.17930016	2374
		Total		14379545.194301	6888
	Quarter 3, 2008			50840.6300000001	976
	Total			25828762.1026999	13951
Grand Total				67851111.1146947	26401

Môžeme vidieť časovú dimenziu hierarchicky ako sme ju vytvorili podľa roku, štvrťroku, mesiaca a dňa. Fakty predstavujú sumárne hodnoty predaja (Sales). Agregácie sú tvorené z tabuľky faktov agregovaním dát pozdĺž dimenzií. Počet možných agregácií je určený všetkými prípustnými kombináciami granularity dimenzií. Fakty sú prepočítavané položky Sub Total, Sales Order Header Count, Freight, Tax Amount a Total Due, ktoré sme určili.

Prehľadávanie OLAP kocky slúži ako reportovacie riešenie a nástroj pre analýzy.

Kapitola 4

Reporting

Reporting je proces tvorby reportov. Umožňuje organizácii jednoduchý prístup, formátovanie a distribúciu informácií svojim zamestnancom, zákazníkom alebo obchodným partnerom. Jeho úlohou je vo vhodnej forme a včas poskytnúť podklady pre podporu rozhodovania.

Reporty môžeme rozdeliť na:

- Statické - podobné ako papierové,
- Interaktívne - pomocou ovládacích prvkov sa dajú ľahko prispôsobiť pre potreby konkrétneho užívateľa.

Reporting rozlišujeme:

- Firemný (Enterprise) - prezentácia informácií v rámci spoločnosti, koncoví užívatelia prístupujú k reportom prostredníctvom intranetu; reporty by mali čo najvernejšie zobrazovať skutočnosť,
- B2B (Business to business) - generovanie reportov pre obchodných partnerov a zákazníkov; reporty sa líšia v zobrazovaní "citlivých informácií".

Reportovacie služby umožňujú generovať zostavy a reporty z relačných aj analytických databáz.

Životný cyklus reportu:

- Návrh reportu - určuje k akým dátam a akým spôsobom sa bude pristupovať,
- Správa reportu - údržba a prípadný ďalší rozvoj reportu,
- Doručenie reportu - vhodným spôsobom, v požadovanom čase a forme klientovi.

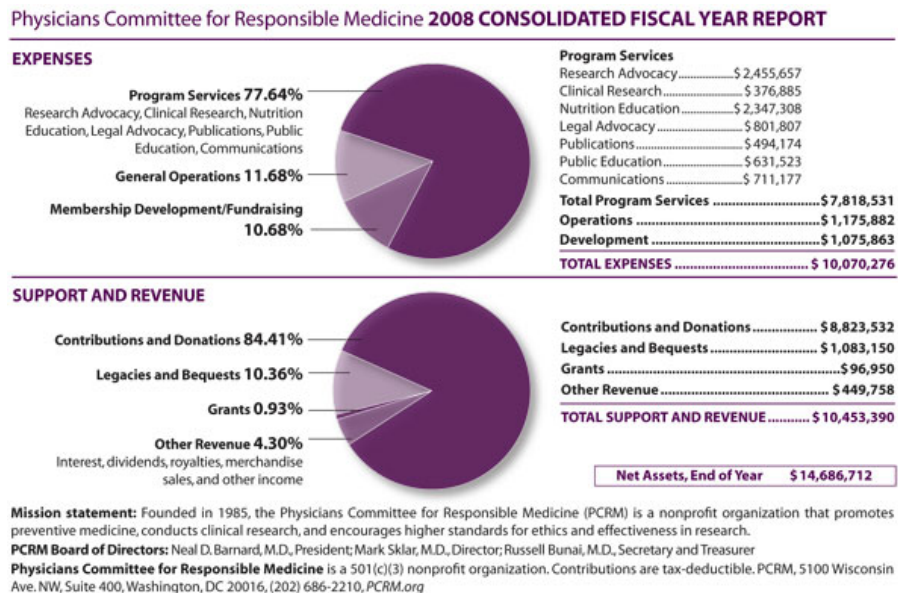
Návrh reportu

Vývojár sa musí najskôr zoznámiť so štruktúrou dát a rozhodnúť, na základe akých dát sa bude report vytvárať a aké bude mať návrhové usporiadanie. Úprava reportu býva často vopred špecifikovaná podľa firemnej úpravy alebo podľa požiadaviek oddelenia. Možnosti zobrazenia dát: tabuľky, matice, grafy, dynamické a hierarchické parametre, možnosti filtrovania, zoskupovania, škálovateľné merítka, gauges, agregčné funkcie.

Doručenie reportu

Doručenie sa môže realizovať na požiadanie alebo na základe časového plánu. Existujú dva základné spôsoby prístupu klienta k reportom:

- Od klienta k reportu - prístup priamo na stránku s reportom,
- Od reportu ku klientovi - posielanie reportov elektronickou poštou.



Obr. 4.1: Príklad reportu, zdroj [24]

Kapitola 5

Implementácia BI v projekte SOFIA

Cieľom tejto kapitoly je opísať postup implementácie systému SAP Business Intelligence v informačnom systéme slovenských verejných vysokých škôl, v projekte nazvanom SOFIA 2 združenia Siemens, S&T a Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky. Kapitola je kompletne spracovaná podľa Cieľového konceptu projektu [2].

Podrobnejšie o spoločnosti SAP AG a jej produktoch sa možno dočítať takisto v diele [1].

5.1 Ciele implementácie systému SAP BI

Základné ciele projektu:

- poskytnúť nástroje na efektívnu prípravu a vytváranie reportov potrebných pre riadenie VVŠ prostredníctvom jednotnej konsolidovanej informačnej platformy, teda dátového skladu, schopného dlhodobo ukladať a analyzovať údaje verejných vysokých škôl,
- vytvoriť konsolidovanú evidenciu vybraných ukazovateľov a reportov pre všetky VVŠ,
- pokryť požiadavky na výkazníctvo - porovnávanie informácií v čase, multidimenzionálna analýza, možnosť meniť podrobnosti informácií,
- poskytnúť kompletný otvorený nástroj, ktorý si bude môcť zákazník ďalej spravovať a rozširovať bez nutnosti programovania.

5.2 Popis riešení SAP Business Intelligence

Systém SAP BI poskytuje efektívne nástroje pre tvorbu požadovaných výkazov a jeho súčasťou je prezentačná vrstva postavená na báze MS Excel. Kompletnú funkčnosť BI pokrývajú riešenia

- SAP Netweaver Business Warehouse (SAP BW)
- SAP BusinessObjects Business Intelligence (BOBJ BI)

5.2.1 SAP Netweaver Business Warehouse

Je technológia umožňujúca extrakciu, spracovanie a skladovanie požadovaných údajov vo forme vhodnej pre multidimenzionálnu analýzu - dátový sklad.



Obr. 5.1: Architektúra SAP BW, zdroj [25]

Systém SAP BW tvoria 3 základné bloky:

Business Information Warehouse Server - umožňuje spracovanie veľkého množstva operatívnych a historických údajov.

Data Warehousing Workbench - nástroj pre riadenie, monitorovanie a údržbu všetkých procesov spojených s prípravou dát v rámci systému SAP BW. Umožňuje tvorbu multidimenzionálneho dátového modelu, tvorbu a nastavenie objektov, tabuliek, prenosových pravidiel a programov potrebných pre kompletnú realizáciu dátového skladu, ďalej monitorovanie, plánovanie a automatizáciu činností v rámci prevádzky systému SAP BW a dokumentáciu objektov SAP BW v HTML formáte (Metadata Repository).

Business Explorer (BEx) - umožňuje analýzu dát uložených v SAP BW pomocou dotazov a prezentáciu výsledkov analýz formou výkazov, tabuliek a grafov. Má 2 hlavné súčasti:

- **Business Explorer Analyzer** umožňuje používateľom prehľadávať dáta z rôznych pohľadov a v rôznych kombináciách.
- **Business Explorer Browser** poskytuje rozhranie pre prístup užívateľa k dátam a prácu s nimi prostredníctvom web browsera pri plnom zachovaní analytických funkcií (drilldown, filtre, atď.) a možnosti grafickej prezentácie.

5.2.2 SAP Business Objects Business Intelligence

Rozšírená prezentačná vrstva SAP BOBJ BI poskytuje nástroje pre analytický reporting a prístup všetkých užívateľov k informáciám s minimálnou závislosťou na zdrojoch IT a vývojárach.

Crystal Reports

Umožňuje ľahkú tvorbu formátovaných výkazov a zostáv. Obsahuje veľa možností formátovania výsledného výstupu, nastavenia medzisúčtov a triedení. Finálne výstupy sú zväčša statické reporty, môžu obsahovať grafy.

Xcelsius SAP

Zvyšuje zrozumiteľnosť dát prostredníctvom konsolidovaných prehľadov kľúčových ukazovateľov. Umožňuje vytvárať interaktívne dashboards zo súborov Microsoft Excel (Xcelsius Engage) alebo interaktívne prezentácie (Xcelsius Present).

SAP Business Objects Explorer

Tento nástroj umožňuje jednoduchú a veľmi rýchlu analýzu dát v tabuľkovej a zároveň aj grafickej podobe.

SAP Business Objects Web Intelligence

Komplexné funkcie pre výkazníctvo a reporting, dotazy a analýzy, dashboardy a vizualizácie, správu dát, plánovanie a zostavovanie rozpočtu, riadenie stratégií.

5.3 Predpokladané zdroje údajov

Systémy BI sú úzko prepojené na zdrojové systémy a moduly a preberajú z nich transakčné aj kmeňové dáta. Kmeňové a transakčné dáta budú udržiavané alebo vytvárané v transakčných systémoch a do SAP BW sa budú iba automatizovane alebo manuálne načítavať. V SAP BW sa nepredpokladá zmena alebo údržba týchto údajov, za kvalitu údajov zodpovedajú zdrojové transakčné systémy.

SAP BW bude komunikovať so zdrojovými systémami pomocou štandardných rozhraní pre extrakciu údajov alebo budú prostriedkami SAP vyvinuté nové extraktory v systéme. Počas analýzy sa identifikovali nasledovné zdroje údajov do BI systému:

1. FIS - Finančný informačný systém (SAP ERP)

Je primárny zdroj údajov pre BI systém. Údaje do SAP BW preberá z oblastí:

- FI (Financials) - modul Finančné účtovníctvo,
- PSM (Public Sector Management) - modul Rozpočet,
- CO (Controlling) - modul Controlling,
- HR (Human Resources) - modul Personalistika.

2. CRZP - Centrálny register záverečných prác

Slúži na vyhodnotenie počtov záverečných prác podľa kategórií.

3. CREPČ - Centrálny register publikačnej činnosti

Slúži na vyhodnotenie počtov publikačnej činnosti.

4. CRŠ2 - Centrálny register študentov verzia 2

Poskytuje informácie o študentoch, štúdiu, štipendiách a ubytovaní študentov, informácie o internátoch.

5. KEGA, VEGA, APVV, MVTS, atď. - Grantové agentúry

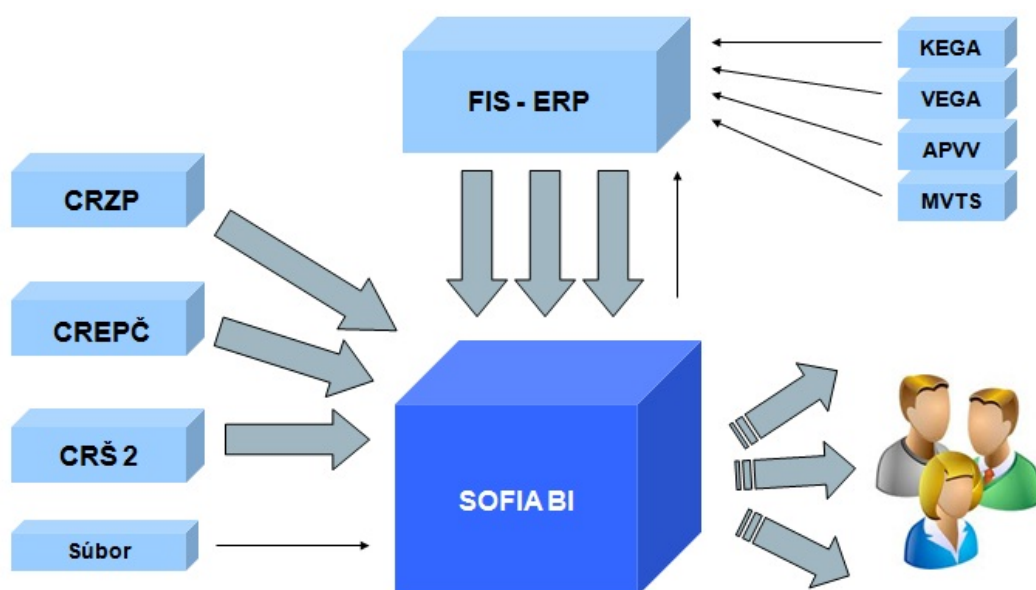
Všetky informácie o grantoch budú čerpané z FIS systému (externé číslo grantu, zoznam riešiteľov, plán pridelených KV a BV po jednotlivých rokoch, a pod.).

6. AIS - Akademické informačné systémy

V projekte sa nebudú realizovať samostatné rozhrania medzi akademickými informačnými systémami a SAP BI, ale namiesto toho bolo navrhnuté aby sa rozšírili už existujúce rozhrania na CRŠ2 (informácie o ubytovaní a internátoch, ID a meno školiteľa).

7. Súbor - externé číselníky, informácie o študentských jedálňach

Pre údaje, pre ktoré nebolo možné identifikovať vhodný univerzálny zdroj údajov sa ako vstupné rozhranie použije nahrávanie údajov z CSV súborov.



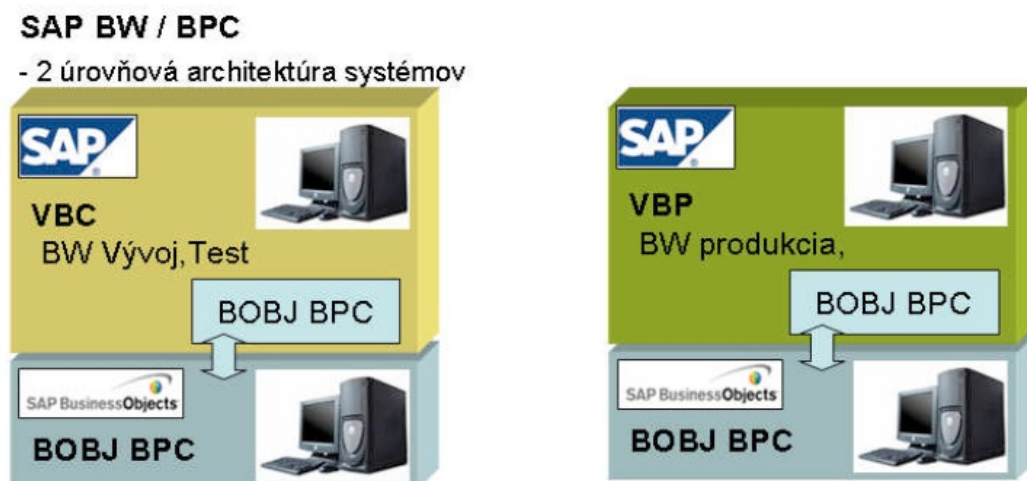
Obr. 5.2: Zdroje dát, zdroj [2]

5.4 Systémová architektúra a toky dát

Systémová architektúra pozostáva okrem nezávislého ERP systému z minimálne dvoch ďalších samostatných BW systémov - vývojového a produktívneho. Systémy SAP BW budú založené na dvojúrovňovej architektúre:

- Systém VBC - slúžiaci ako vývojový a testovací systém,
- Systém VBP - slúžiaci na produktívnu prevádzku.

Súčasne oba BW systémy slúžia ako aplikačný server a dátová základňa pre systémy **SAP BOBJ BPC (Business Planning and Consolidation)**, ktoré budú prevádzkované na princípoch dvojúrovňovej architektúry - Vývoj/Test a Produkcia.



Obr. 5.3: SAP BW/BPC, zdroj [2]

Systém **SAP BOBJ BI** bude mať iba jednoúrovňovú architektúru, ktorá sa bude využívať pre vývojové aj produktívne prostredie. Rozlíšenie zdrojov dát (VBC, VBP) bude realizované rôznym mapovaním dátových univerz.

Jednoúrovňovú architektúru bude mať aj **SAP BOBJ Dataservices**, nástroj paralelný k SAP BW ETL nástrojom, slúžiaci na extrakciu, transformáciu a nahrávanie dát do BI systému.

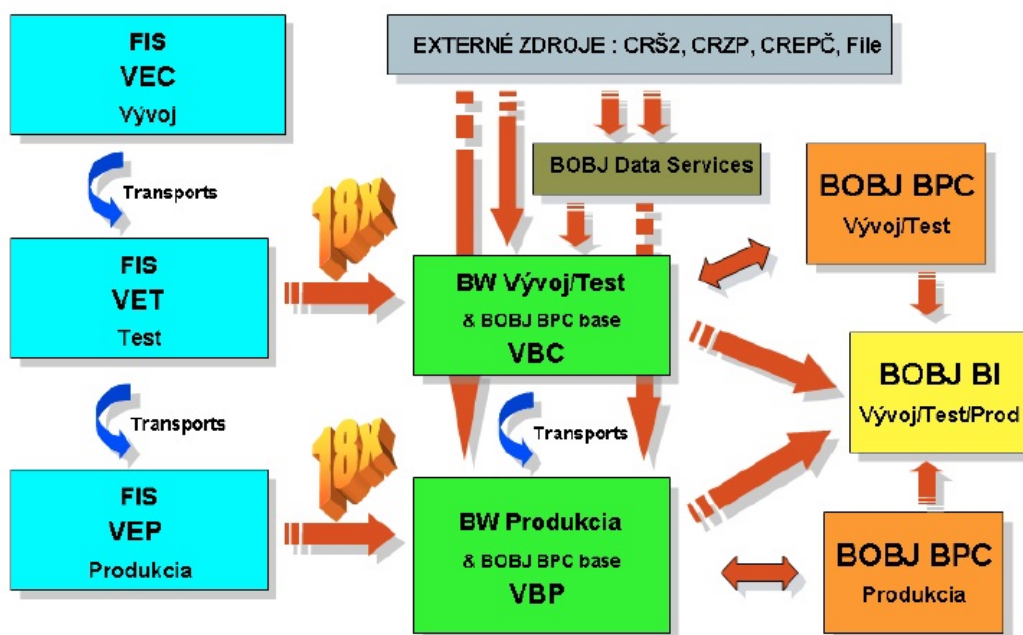


Obr. 5.4: SAP BOBJ BI, zdroj [2]

Systém SAP BW slúži ako hlavné úložisko dát z rôznych zdrojových systémov pre účely multidimenzionálnej analýzy. Zo zdrojových systémov sa do SAP DW prenášajú nastavenia systémov (transportný systém) ako aj aplikačné dáta.

Vývojový a testovací systém BW bude napojený na testovacie dáta z testovacieho FIS systému a produktívny BW systém bude napojený na produktívny FIS systém. V rámci systému FIS sa budú musieť transportovať aj extraktory a nastavenia v zdrojovom systéme. Dáta sa transportujú oddelene pre každú VVŠ (18x).

Riešenie SAP BOBJ BI bude čerpať údaje výlučne zo systémov SAP BW.



Obr. 5.5: Toky dát medzi systémami, zdroj [2]

5.5 Funkcionalita

Vďaka zlúčeniu informácií z mnohých systémov do jednotného modelu SAP BI sa otvárajú nové, komplexnejšie možnosti pre reporting a analýzu údajov VVŠ ako celkov, nielen rozdrobených samostatných systémov, z ktorých každý dokáže narábať so svojimi čiastočnými údajmi so samostatnými prístupmi a používateľskými rozhraniami.

Systém SAP BI bude pokrývať nasledovné požiadavky:

- Extrakcia transakčných údajov z FIS,

- Možnosť extrakcie transakčných a kmeňových údajov z iných systémov,
- Extrakcia zodpovedajúcich kmeňových údajov a štruktúr (hierarchií),
- Načítanie extrahovaných údajov do SAP BW,
- Načítanie doplňujúcich manuálne vytvorených údajov do SAP BW,
- Dlhodobé uskladnenie údajov v SAP BW,
- Vytvorenie výkazov nad údajmi v SAP BW,
- Prezentácia/spúšťanie výkazov v prostredí MS Excel a SAP BOBJ nástrojoch,
- Možnosť tvorby nových výkazov nad údajmi SAP BW,
- Riadený prístup užívateľov k systému.

5.5.1 Postup implementácie funkcionality

EXTRAKCIA

Predstavuje výber a prenos údajov zo zdrojového systému do SAP BW. Funkcie na vstupe SAP BW zabezpečujú uloženie vstupujúcich údajov vo forme 1:1 vo vstupnej oblasti (Persistent Staging Area), kde sú dáta ukladané zvlášť pre každý zdrojový systém. Tu sú uchovávané do doby, kedy budú konzistentne uložené prípadne agregované v dátových cieľoch SAP BW.

Rozsah implementácie:

Aktivácia dátových zdrojov a extraktorov v SAP ERP (FIS),

Replikácia dátových zdrojov do SAP BW,

Vytvorenie Infopaketu pre načítanie údajov,

Načítanie údajov do SAP BW,

Vytvorenie dátových zdrojov v SAP BW pre načítanie údajov z externých systémov alebo súborov.

TRANSFORMÁCIA A ZAVÁDZANIE ÚDAJOV

Ďalším krokom je čistenie a transformácia údajov zo vstupnej oblasti PSA ešte pred ich uložením do dátových cieľov SAP BW. Programy zabezpečujúce čistenie a transformáciu dát sa nazývajú Transformations. Je možné ich používateľsky definovať a meniť.

Rozsah implementácie:

Definícia, úprava a aktivácia transformácií.

ÚLOŽENIE DÁT PRE ON-LINE ANALÝZU

Predstavuje uloženie dát vo forme multidimenzionálneho dátového modelu. Jeho reprezentáciou v SAP BW je **Infokocka**, pri ktorej modelovaní sa využíva star schéma. Uchovávanie údajov je založené na inteligentnej kombinácii Infokocky (informačných dátových modelov) a kmeňových údajov. Kmeňové údaje obsahujú trvalejšie informácie o atribútoch jednotlivých charakteristík (adresy, regióny a pod.).

Dalšími z tzv. dátových cieľov môžu byť objekty **Data Store Objects (DSO)**.

Rozsah implementácie:

Uloženie údajov v infokockách, objektoch DSO a tabuľkách kmeňových údajov, Externé hierarchie.

MONITOROVANIE PROCESOV V SAP BW

Celý proces ETL je monitorovaný pomocou nástroja Monitor, ktorý je súčasťou Data Warehousing Workbench. O všetkých súvisiacich udalostiach sú zaznamenávané protokoly, aby bolo možné identifikovať prípadné chybové stavy a odstrániť ich príčiny.

Rozsah implementácie:

Monitor procesov načítavania údajov, Protokolovanie aktivácie objektov pri modelovaní.

ANALÝZY

Využíva sa technológia OLAP, nástroje analýzy umožňujú meniť parametre, filtre a nastavenia výkazov.

Rozsah implementácie:

Funkcie multidimenzionálnej analýzy (drill-down, drill-up, filtre, premenné, atď.).

PREZENTÁCIA

Možnosť prezerať a upravovať informácie podľa požiadaviek umožňuje nástroj Business Explorer. Výkazy a zostavy možno zoskupovať do preferovaných skupín, a tiež je možné prechádzať sklad údajov. Prezentačná vrstva umožňuje umiestnenie výkazov na intranete alebo internete a prístup k nim pomocou webového prehliadača.

Rozsah implementácie:

Prepojenie BOBJ BI so SAP BW, Bex Analyzer - na báze MS Excel, Vytvorenie BOBJ Univerz pre vybrané oblasti, Vytvorenie príkladových BOBJ výstupov (reportov).

SPRÁVA DÁT

Konzola správcu SAP BW podporuje riadenie a údržbu kľúčových aspektov skladovania údajov v grafickom prostredí.

Rozsah implementácie:

Jednotný archív metadát (Metadata Repository) - zabezpečuje konzistentnosť údajov celého systému,

Modelovanie - Dataprovidery, Transformácie, Infozdroje (komunikačné štruktúry), Dátové zdroje (prenosové štruktúry).

SPRÁVA SYSTÉMU

SAP BW obsahuje funkcie na udržanie najvyššieho možného výkonu systému. Napríklad funkciu monitorovania zaťaženia a prístupu, plánovania kapacity a podporu zadefinovania prístupových práv na mnohých úrovniach (možnosť zadefinovať kto má mať prístup k akým údajom a na akej úrovni podrobnosti údajov).

Rozsah implementácie:

Bázová administrácia systému,

Správa rolí a oprávnení.

5.6 Dátový model

Objekty nad ktorými je možné vytvárať reportingové dotazy označujeme spoločným pojmom **Infoprovideri** (infokocky, DSO objekty, infoobjekty a ich kombinácie). V tomto projekte bude nastavený nasledovný základný dátový model:

Oblasť Human Resources

- Infokocka : ZPAPA_C02 - Stav zamestnancov
- Infokocka : ZPY_C02 - Presné dáta zúčtovania

Oblasť Financials

- Infokocka : ZFIGL_C01 - Hlavná kniha : Obraty na účte
- Infokocka : ZFIGL_VC01 - Hlavná kniha: Súvaha a výsledovka

Oblasť Public Sector Management

- Infokocka : ZPU_C02 - Obligo/skut. a rozpočet v riadení rozpočtu (BCS)
- Infokocka : ZPU_C03 - Obligo/skut. a rozpočet s priradením účtov RR (BCS)

Oblasť Controlling

- Infokocka : ZCCA_C11 - CO-OM-CCA: Náklady a zúčtovania
- Infokocka : ZPCA_C01 - Profit Center Accounting: Transakčné Dáta

Oblasť Travel Management

- Infokocka : ZFITV_C01 - Zúčtovanie cestovných nákladov: Súčty a doklady

Oblasť : Externé zdroje - Publikácie

- Infokocka : ZPUB_C01 - Dáta z CREPČ
- Infokocka : ZPUB_C02 - Dáta z CRZP

Oblasť : Externé zdroje - Informácie o študentoch

- Infoobjekt : ZSTUDENT - Študent
- Infokocka : ZSTUD_C01 - Štúdium
- Infokocka : ZSTUD_C02 - Štipendiá
- Infokocka : ZSTUD_C03 - Ubytovanie

Oblasť : Externé zdroje - ostatné

- Infoobjekt : ZSTPROG - Študijný program
- DSO : ZEXT_C01 - Študentské jedálne

Dané infokocky obsahujú **charakteristiky (dimenzie)** ako napríklad čas, klienta (všetky transakčné údaje budú uložené v systéme SAP BI s identifikáciou zdrojového systému - VVŠ) alebo druh rozpočtu a **klúčové číselné údaje (ukazovatele/fakty)** ako napríklad schválený rozpočet alebo čerpanie rozpočtu.

Podrobný popis dátového modelu jednotlivých kociek (ukazateľov a charakteristík) je možné nájsť v cieľovom koncepte. Popísaný dátový model je možné v budúcnosti ďalej rozširovať a kombinovať.

5.7 Reporting

Súčasťou funkcionality je aj vytvorenie vyšpecifikovaných hotových príkladových výstupov, ktoré môžu slúžiť ako vzory pre tvorbu ďalších výstupov nad existujúcim dátovým modelom. Dátový model SAP BW bude obsahovať všetky dáta potrebné pre vytvorenie všetkých schválených výkazov podľa ich zadaných vzorov.

Skupiny výstupov

- Výkazy výročnej správy,
- Akreditačné výstupy,
- Manažérske HR výstupy,
- Manažérske reporty,
- Kontrolingové reporty,
- Agentúrne vykazovanie,
- Reporty nad CRZP,
- Reporty nad CRPČ,
- Reporty nad študentmi.

5.7.1 Zoznam preddefinovaných výstupov

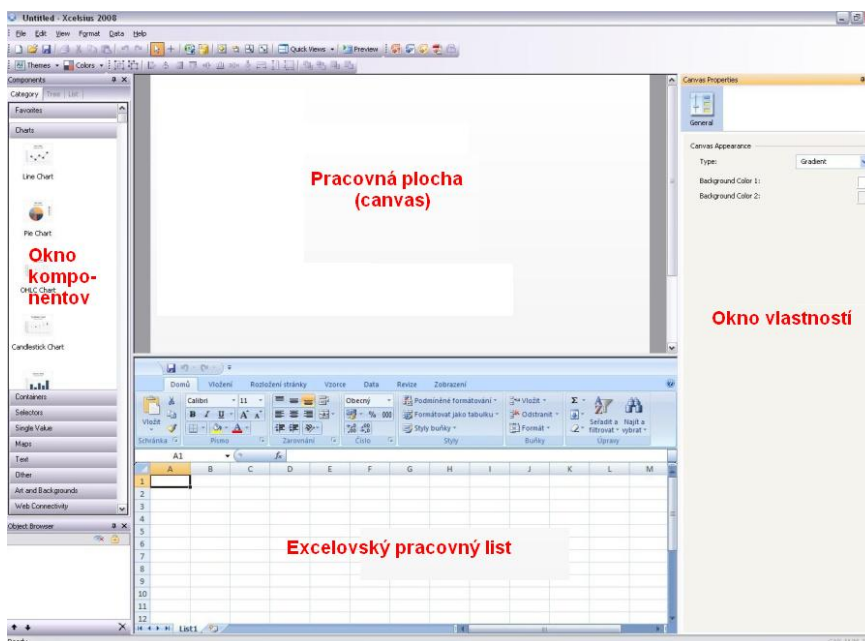
Preddefinované výstupy zo systému sú:

- Tabuľkové výstupy (napr. Výnosy verejnej vysokej školy v rokoch XXXX-1 a XXXX, Náklady verejnej vysokej školy v rokoch XXXX-1 a XXXX, Súvaha za rok XXXX, atď.),
- Prepočítané počty publikácií podľa kategórií zamestnancov,
- Počet študentov v študijných programoch prvého a druhého stupňa na prepočítaný evidenčný počet vysokoškolských učiteľov v jednotlivých rokoch hodnoteného obdobia,
- Počet doktorandov denného doktorandského štúdia a externého doktorandského štúdia, ktorí sú zároveň zamestnancami školy, na prepočítaný evidenčný počet docentov a profesorov,
- Absolventi doktorandského štúdia a počet profesorov za kalendárne roky,
- Sumár grantov podľa typu po rokoch,
- Objem financií na vedu - Objem financií na výskum prepočítaný na docenta a profesora,
- Vyhodnotenie zákaziek UNIKANU - prehľad štruktúry nákladov, prípadne porovnanie plánu so skutočnosťou,

- Sumárny prehľad nákladov a výnosov dotačných, nedotačných a PC úloh, porovnanie na plán,
- Sumárny prehľad nákladov a výnosov organizačnej štruktúry, porovnanie na plán (ziskové strediská/nákladové strediská),
- Priemerný vek zamestnancov podľa kvalifikačnej štruktúry,
- Počet titulov učiteľov v pomere pracovného úväzku,
- Vývoj miezd podľa období podľa ŠPP prvkov,
- Výkonové ukazovatele,
- Prehľad projektov a finančných prostriedkov,
- Výkaz pre Grantistu - Výkaz porovnania plánu a kvartálneho čerpania projektu,
- Výkaz pre Grantistu - Výkaz čerpania projektu za viac rokov,
- Výkaz čerpania podľa prvkov ŠPP pre agentúry,
- Výkaz čerpania grantu v členení na fázy,
- Detailný prehľad čerpania grantu,
- Prehľad služobných ciest na grant,
- Report nad základnými informáciami o študentoch,
- Report nad Info z CRZP,
- Report nad Info z CREPČ.

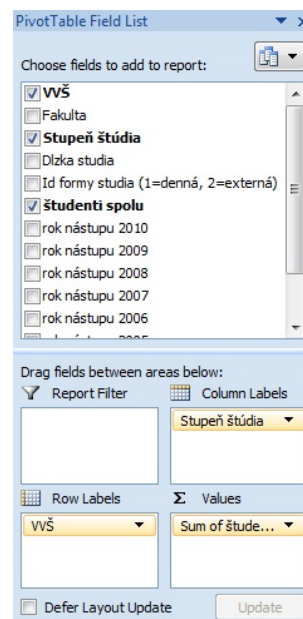
5.7.2 Tvorba vybraných výstupov v programe XCelsius

Podrobnejší popis prostredia a vybraných komponentov programu XCelsius Engage 2008 ako aj postup inštalácie 30-dňovej trial verzie je popísaný v bakalárskej práci [1]. V tejto časti budeme demonštrovať tvorbu vybraných preddefinovaných výstupov projektu SOFIA 2 v programe XCelsius.



Dáta, z ktorých sme vychádzali, sme čerpali z databázy študentov obsahujúcej 7877 podrobných záznamov. Agregované údaje sme získali pomocou kontingenčnej tabuľky v programe MsExcel.

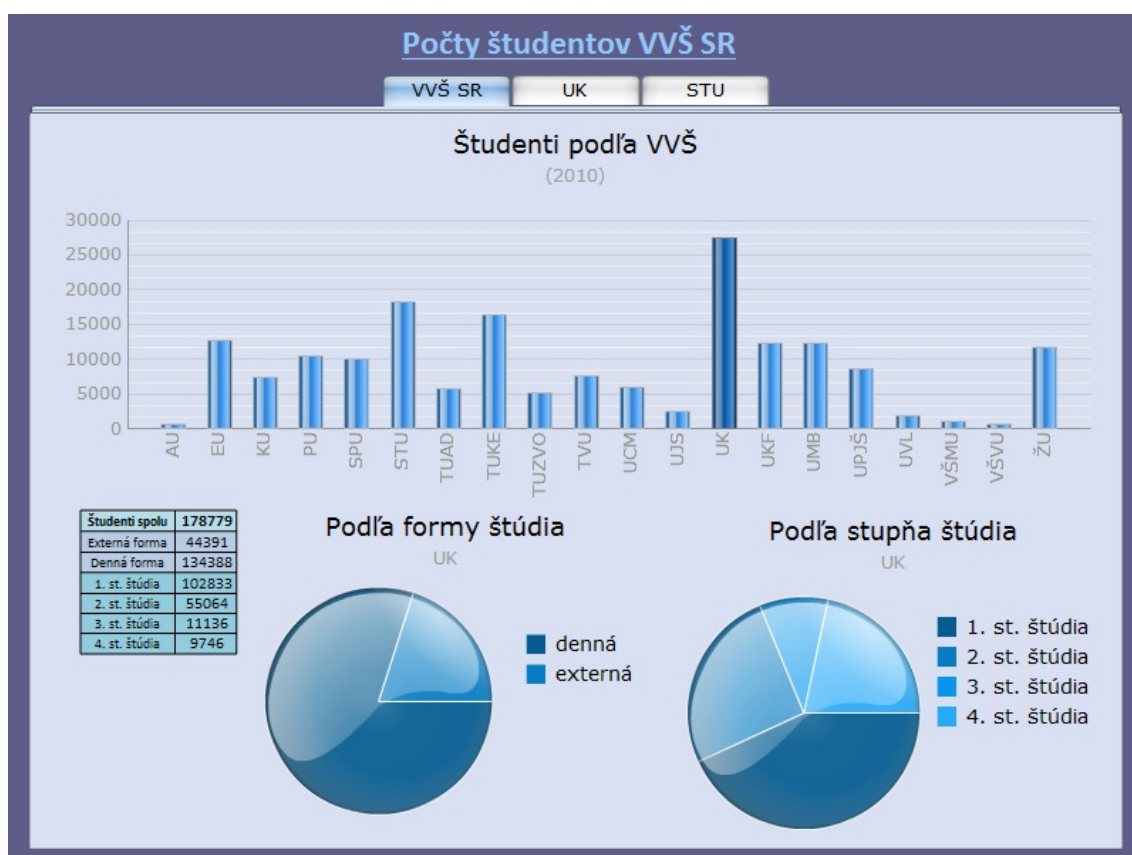
Sum of študenti spolu	Stupeň štúdia				Spolu
	1	2	3	4 (blank)	
VVŠ					
AU	374	186	69		629
EU	6977	5195	584	3	12759
KU	4312	2405	422	214	7353
PU	6492	3314	571	194	10571
SPU	5884	3623	449	8	9964
STU	10973	5426	1790		18189
TUAD	3558	2074	111		5743
TUKE	10167	5128	1023	1	16319
TUZVO	3280	1664	291		5235
TVU	4671	2360	478	7	7516
UCM	4056	1751	138	2	5947
UJS	1607	722	15	59	2403
UK	11818	7060	2620	5924	27422
UKF	7896	3935	520	3	12354
UMB	6956	4921	447	25	12349
UPJŠ	4636	1596	584	1883	8699
UVL	281	11	175	1406	1873
VŠMU	608	318	116	10	1052
VŠVU	379	224	84		687
ZU	7908	3151	649	7	11715
(blank)				178779	178779
Spolu	102833	55064	11136	9746	178779
					357558



Report nad základnými informáciami o študentoch

Príklad interaktívnych dashboardov zobrazujúcich sumarizované údaje o počte študentov všetkých ako aj jednotlivých VVŠ. Pre ukážku dashboardov jednotlivých VVŠ sme vybrali dve vzorové univerzity, a to UK a STU.

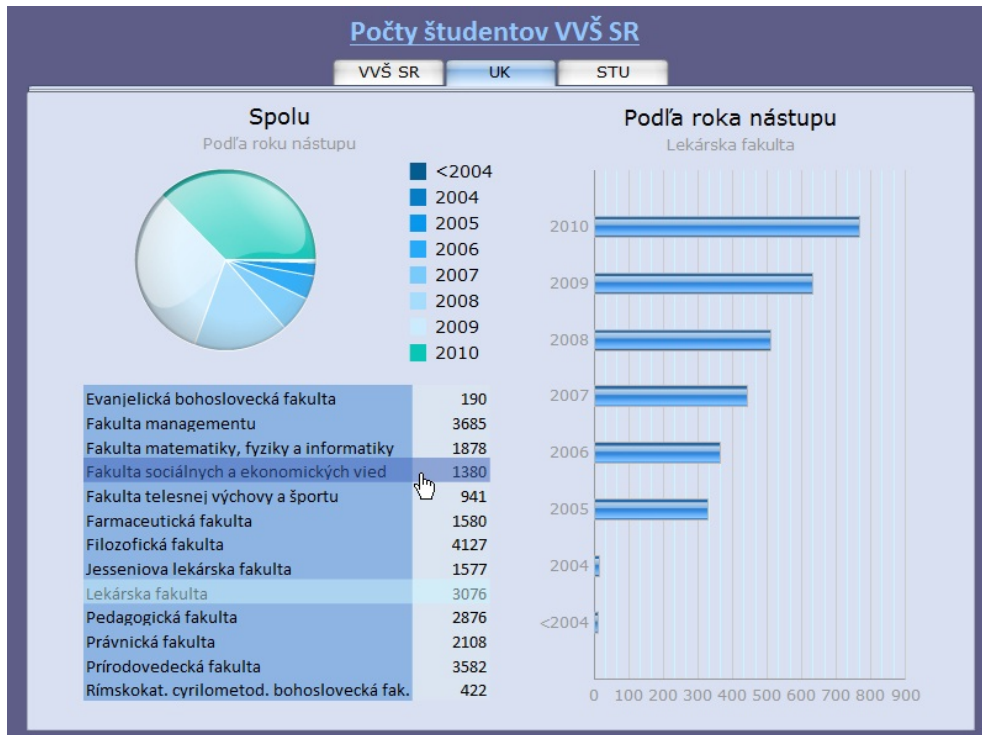
Pri tvorbe využívame nasledovné komponenty: skupina tabulátorov (Tab Set), stĺpcový graf (Column Chart, Bar Chart), koláčový graf (Pie Chart), tabuľka programu MsExcel (Spreadsheet Table), Zoznam (List Box) a zároveň funkciu Drill Down, ktorá umožňuje prepojiť jednotlivé komponenty a zobrazovať podrobnosti interaktívnym spôsobom.



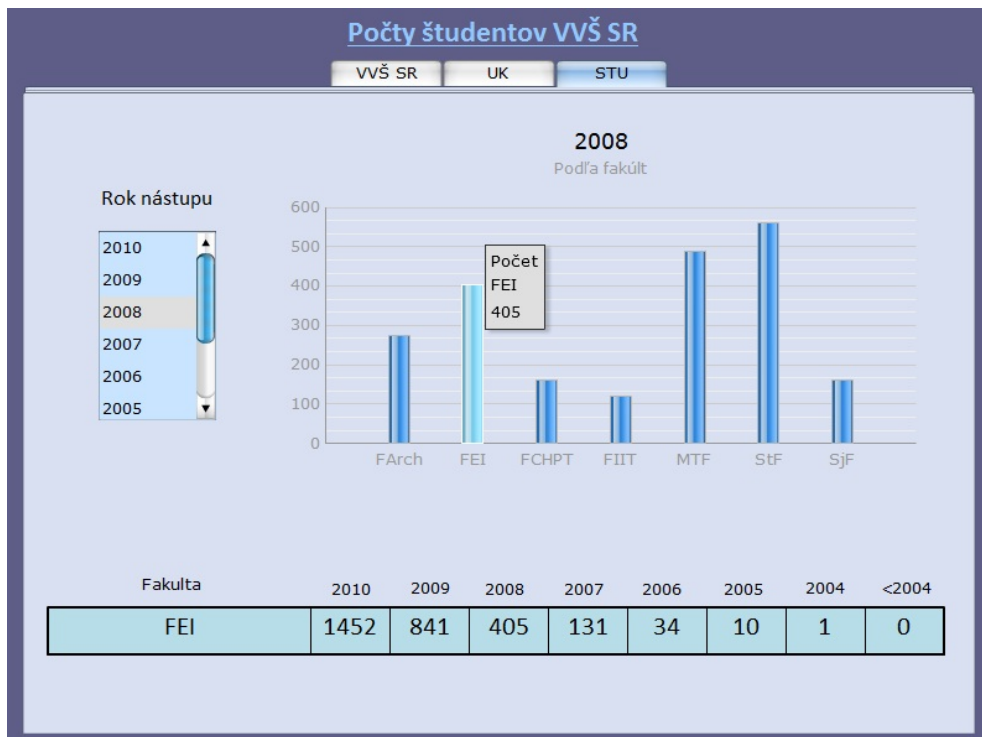
Obr. 5.6: Dashboard počtu študentov VVŠ SR

Uvedený dashboard je interaktívny, to znamená, že po kliknutí myšou na jednotlivé položky stĺpcového grafu sa nám zobrazia údaje o jednotlivých univerzitách na koláčových grafoch v dolnej časti dashboardu. Dashboard obsahuje vo forme tabuľky aj prehľad sumárneho počtu študentov všetkých univerzít podľa jednotlivých kritérií.

Univerzita Komenského:



Slovenská technická univerzita:



Na predchádzajúcej strane sa nachádzajú ukážky ďalších interaktívnych dashboardov. Ilustrujú rôzne možnosti zobrazenia požadovaných dát a ich prepojenia. V prvom dashboarde po výbere fakulty vidíme podrobný prehľad študentov podľa roku nástupu. V druhom prípade určíme najskôr rok nástupu a na grafe vidíme počet študentov, ktorí nastúpili v danom roku, podľa fakúlt. Po kliknutí na konkrétnu fakultu sa nám v dolnej tabuľke zobrazia dáta pre danú fakultu za všetky roky.

Takýmto spôsobom získavame rýchly prehľad dát z rôznych uhľov pohľadu.

Záver

V tejto diplomovej práci sme sa venovali téme Business Intelligence, konkrétne oblastiam tvorby dátového skladu, analytickým možnostiam nad dátami uloženými v dátovom sklade (technológia OLAP) a reportingu.

Prvá časť práce predstavuje teoretický prehľad týchto metód pričom pre ľahšie pochopenie danej problematiky uvádzame aj konkrétne príklady. V druhej časti prezentujeme konkrétny príklad využitia BI riešení v praxi v projekte funkčného upgradu finančného informačného systému slovenských verejných vysokých škôl. Tento projekt je momentálne v realizačnej fáze a jeho prechod do produktívnej prevádzky sa plánuje od 1.1.2012.

V oblasti samotnej BI je pravdepodobné, že so stále rastúcim množstvom dát sa bude BI naďalej rozvíjať a poskytovať stále lepšie riešenia pre podporu rozhodovacích procesov.

Literatúra

- [1] TÜRKOVÁ, Katarína: Business Intelligence a jej podpora v systéme SAP (bakalárska práca), 2009, dostupné online na <http://stella.uniba.sk/zkp-storage/dzb/dostupne/FM/2009/2009-FM-yPGLfA/>
- [2] KAPALLA, Pavol, Ing.: Cieľový koncept projektu SOFIA 2 - Kapitola 4.4.Reporting, Združenie Siemens a S&T, 11.4.2011
- [3] LACKO, Luboslav: Business Intelligence v SQL Serveru 2008, Computer Press, 2009, ISBN 978-80-251-2887-9
- [4] LACKO, Luboslav: ORACLE - Správa, programování a použití databázového systému, Computer Press, 2003, ISBN 80-7226-699-3
- [5] [online] [3.6.2009]
http://searchdatamanagement.techtarget.com/sDefinition/0,,sid91_gci213571,00.html
- [6] [online] [3.6.2009]
<http://rationalizedthoughts.blogspot.com/2007/01/business-intelligence-basics.html>
- [7] [online] [23.3.2011]
<http://www.wisegeek.com/what-is-data-warehousing.htm>
- [8] [online] [23.3.2011]
<http://www.1keydata.com/datawarehousing/data-warehouse-definition.html>
- [9] [online] [23.3.2011]
<http://www.intranetjournal.com/features/datawarehousing.html>
- [10] [online] [23.3.2011]
http://en.wikipedia.org/wiki/Data_warehouse#History
- [11] [online] [23.3.2011]
<http://tdwi.org/>

- [12] [online] [23.3.2011]
http://www.dwreview.com/DW_Overview.html
- [13] [online] [23.3.2011]
<http://www.exforsys.com/tutorials/data-warehousing/how-data-is-stored-within-a-data-warehouse.html>
- [14] [online] [23.3.2011]
<http://it.toolbox.com/blogs/bi-telecom/data-warehouse-dimensional-model-vs-normalized-model-40253>
- [15] [online] [23.3.2011]
<http://www.exforsys.com/tutorials/msas/data-warehouse-design-kimball-vs-inmon.html>
- [16] [online] [23.3.2011]
<http://mydbaworld.wordpress.com/2009/07/23/bill-inmon-vs-ralph-kimball/>
- [17] [online] [23.3.2011]
<http://www.itnirvanas.com/2009/02/kimball-vs-inmon.html>
- [18] [online] [23.3.2011]
http://it.toolbox.com/wiki/index.php/Data_Warehouse_Architecture
- [19] [online] [23.3.2011]
<http://www.mysqlperformanceblog.com/2010/07/15/data-mart-or-data-warehouse/>
- [20] [online] [23.3.2011]
<http://www.inmoncif.com/home/>
- [21] [online] [23.3.2011]
<http://revistaie.ase.ro/content/42/velicanu.pdf>
- [22] [online] [23.3.2011]
<http://www.aspfree.com/c/a/MS-SQL-Server/Accessing-OLAP-using-ASP-dot-NET>
- [23] [online] [23.3.2011]
<http://www.kodyaz.com/articles/how-to-create-olap-cube-in-business-intelligence-development-studio.aspx>
- [24] [online] [23.3.2011]
http://www.pcrm.org/magazine/gm09winter/yir_report.html
- [25] [online] [23.3.2011]
<http://wiki.sdn.sap.com/wiki/display/ERPFI/Overview+BW+Delta+Extraction>