

**Univerzita Hradec Králové**  
**Fakulta informatiky a managementu**  
**Katedra informačních technologií**

**Datový sklad a Business Intelligence pro použití ve veřejné  
správě**

Diplomová práce

Autor: Miroslav Sajvera  
Studijní obor: Aplikovaná informatika

Vedoucí práce: Ing. Zuzana Němcová, Ph.D.

Odborný konzultant: Ing. Petr Němeček, Karel Lidmila, oba GIST s.r.o.

Prohlášení:

Prohlašuji, že jsem diplomovou práci na téma „Datový sklad a Business Intelligence pro použití ve veřejné správě“ zpracoval samostatně a s použitím uvedené literatury.

V Hradci Králové dne 18.4.2018

Miroslav Sajvera

#### Poděkování:

Tímto bych chtěl poděkovat vedoucí diplomové práce Ing. Zuzaně Němcové, Ph.D. za odborné vedení práce a udělené rady. Rovněž bych rád touto cestou poděkoval pracovníkům ze společnosti GIST s.r.o., díky nimž bylo možné vypracovat reálně použitelné řešení zkoumané oblasti a kteří cennými radami přispěli k jeho úspěšné realizaci.

## **Anotace**

Tato diplomová práce se zabývá návrhem a realizací pilotního projektu komplexního BI/DWH řešení pro oblast školství v Hradci Králové. V teoretické části práce je představena problematika Business Intelligence ve spojení s datovým skladem, která zahrnuje popis architektury a hlavních komponent obou oblastí. Rovněž je popsáno dimenzionální modelování, tvorba ETL procesů a analýza OLAP. V praktické části práce je nejprve představena oblast školství a provedena analýza současného stavu společně s identifikací požadavků na nové řešení. Následuje sestavení návrhu BI/DWH řešení, od konceptuálního přes logický až po fyzický model a jeho pilotní implementace. Ta zahrnuje tvorbu datového skladu, ETL procesů, OLAP kostky a uživatelských reportů pomocí nástrojů od společnosti Microsoft.

## **Annotation**

**Title: Data warehouse and Business Intelligence in public administration**

This diploma thesis deals with the design and implementation of a pilot project of a complex BI/DWH solution for education area in Hradec Králové. In the theoretical part of the thesis is presented the issue of Business Intelligence in connection with data warehouse, which includes a description of an architecture and main components of both areas. Dimensional modeling, ETL creation and OLAP analysis are also described in the theoretical part of the thesis. The practical part of the thesis firstly introduces the area of education and analyzes the current state together with the identification of the requirements for the new solution. Secondly the practical part also deals with the design of BI/DWH solution, from conceptual and logical models to physical model and its pilot implementation. This includes creation of data warehouse, ETL processes, OLAP cubes, and user reports using Microsoft tools.

## Obsah

1	Úvod.....	1
2	Cíl a struktura práce.....	2
3	Business Intelligence .....	3
3.1	Oblasti využití BI .....	3
3.2	Důvody zavedení BI .....	4
3.3	Architektura BI .....	5
4	Aktuální trendy a analýza trhu BI.....	10
4.1	Aktuální trendy.....	10
4.2	Analýza trhu.....	15
5	Dimenzionální modelování.....	19
5.1	Fakta.....	20
5.2	Dimenze.....	21
5.3	Schéma propojení tabulek faktů a dimenzí.....	21
6	Datový sklad.....	24
6.1	Životní cyklus datového skladu.....	24
6.2	System datového skladu.....	26
6.3	Architektura datového skladu .....	27
6.4	Metody budování datového skladu.....	36
7	ETL.....	38
7.1	Extrakce dat .....	39
7.2	Transformace a čištění dat.....	41
7.3	Nahrání dat.....	43
8	OLAP.....	45
8.1	OLAP kostka.....	45
8.2	Varianty OLAP .....	47

9	Praktická část.....	50
9.1	Představení oblasti školství.....	50
9.2	Současný stav .....	51
9.3	Požadavky na nové řešení .....	52
9.4	Návrh řešení.....	54
9.5	Pilotní implementace.....	63
9.6	Prezentace dat.....	79
9.7	Přínosy navrženého řešení.....	81
10	Závěry a doporučení .....	82
11	Seznam použité literatury .....	84
12	Přílohy .....	87

## Seznam obrázků

Obr. 1 Obecná architektura BI .....	5
Obr. 2 Výsledky studie BI Trend Monitor 2018 .....	11
Obr. 3 Analýza trhu 2017 dle Gartner.....	17
Obr. 4 Schéma hvězdy .....	22
Obr. 5 Schéma sněhové vločky .....	23
Obr. 6 Životní cyklus datového skladu .....	25
Obr. 7 Systém datového skladu .....	26
Obr. 8 Architektura s jednou vrstvou.....	28
Obr. 9 Architektura se dvěma vrstvami .....	29
Obr. 10 Architektura se třemi vrstvami .....	30
Obr. 11 Architektura samostatných datových trhů.....	31
Obr. 12 Architektura sběrnice .....	32
Obr. 13 Architektura centrálního úložiště .....	34
Obr. 14 Federativní architektura .....	35
Obr. 15 Architektura ETL procesu .....	38
Obr. 16 OLAP kostka .....	46
Obr. 17 Architektura navrženého BI/DWH řešení .....	58
Obr. 18 Připojení k databázovému serveru KHK.....	65
Obr. 19 Ukázka rozšířených vlastností tabulky F_VykazV01_L1.....	67
Obr. 20 Výsledná struktura datového skladu .....	68
Obr. 21 Vytvoření nového SSIS projektu .....	69
Obr. 22 Panely SSIS projektu.....	70
Obr. 23 OLE DB připojení do databáze .....	71
Obr. 24 OLE DB připojení do databáze .....	71
Obr. 25 Control Flow .....	72
Obr. 26 Data Flow .....	73
Obr. 27 Truncate tabulky ve vrstvě L0 .....	74
Obr. 28 Vytvoření nového SSAS projektu.....	76
Obr. 29 Průvodce datových zdrojů .....	76
Obr. 30 Průvodce datových zdrojů .....	77

Obr. 31 Vytvoření pohledu na zdrojová data.....	78
Obr. 32 Výběr tabulek do OLAP kostky .....	78
Obr. 33 Připojení k OLAP kostce v MS Excel .....	79
Obr. 34 Průvodce připojení k OLAP kostce v MS Excel .....	80

## **Seznam tabulek**

Tabulka 1 Číselníky ve školství .....	55
Tabulka 2 Výkazy škol a školských zařízení.....	56



# 1 Úvod

Intenzivnější využívání výpočetní techniky ve 21. století přineslo změnu ve způsobu podnikání a v procesu rozhodování, které nepochybně provází nejednu společnost každý den. Konkurenceschopnost každé společnosti je přitom podmíněna správným rozhodováním v důležitých situacích, v nichž jsou na management kladeny stále větší nároky. Velká pozornost se tak v posledních letech soustřeďuje na systémy a nástroje datových skladů a Business Intelligence, jejichž cílem je získat a využít poznatky založené na znalostech, které mají zlepšit taktickou a strategickou výhodu společnosti na trhu.

Datový sklad a Business Intelligence jsou velmi dynamické a perspektivní oblasti nacházející se na pomezí mezi informatikou, řízením společnosti a statistikou. Velké množství společností po celém světě již těží z možností, které Business Intelligence ve spojení s datovým skladem nabízí v oblasti analýzy, predikce a prezentace informací. Trend budování a nasazování těchto řešení má rostoucí charakter a společnosti, které je efektivně a účelně nasazují, tak získávají ve svém oboru konkurenční výhodu.

Společnosti dnes produkují velké množství dat a informace v nich obsažené se bezpochyby stávají jedním ze strategických zdrojů, který může být využit v konkurenčním boji na daném trhu. Avšak stejně jako narůstá objem dat ve společnosti, ukazuje se potřeba tato data efektivně zpracovávat, analyzovat a přehledně prezentovat. A přesně to je doménou Business Intelligence. Dokáže totiž z velkého objemu uložených a mnohdy dále nevyužitých dat účinně získat informace a výrazně tak zkvalitnit rozhodování a řízení celé společnosti.

## 2 Cíl a struktura práce

Cílem diplomové práce je prozkoumat a popsat problematiku Business Intelligence se zaměřením na datové sklady, navrhnout BI/DWH řešení pro oblast školství v Hradci Králové a provést pilotní implementaci navrženého řešení. Tomu odpovídá i členění práce, která je rozdělena na dvě části, a to část teoretickou a část praktickou.

V rámci teoretické části je nejprve v kapitole 3 popsána oblast Business Intelligence s uvedením její architektury a popisem jednotlivých komponent, respektive vrstev. Rovněž jsou popsány oblasti využití a důvody, proč se vyplatí tuto technologii ve firemním prostředí zavádět. Kapitola 4 je věnována aktuálním trendům a analýze trhu Business Intelligence. V následující kapitole 5 je popsáno dimenzionální modelování, které je ve spojení s Business Intelligence a datovými sklady hojně využíváno. Nejdůležitější a nejobsáhlejší je však kapitola 6, která je věnována datovým skladům. V rámci této kapitoly je uveden popis životního cyklu datového skladu, popis datového skladu jako systému a metody jeho budování. Podrobně je popsána rovněž architektura datových skladů, co se vrstev a komponent týče. Poslední dvě kapitoly teoretické části práce jsou věnovány vrstvě ETL (kapitola 7) a analýze OLAP (kapitola 8), které shrnují možnosti, jak datový sklad naplnit daty a následně nad nimi vykonávat analytické dotazy.

Obsahem kapitoly 9 je celá praktická část práce, která se zabývá především návrhem a pilotní implementací BI/DWH řešení pro oblast školství v Hradci Králové. K tomu však bylo zapotřebí provést analýzu současného stavu, datových zdrojů a určit požadavky na nově vzniklé řešení. Po návrhu řešení následuje vlastní implementace pilotního projektu, která zahrnuje tvorbu samotného datového skladu, ETL procesů a OLAP kostky v prostředí nástrojů MS SQL Server Management Studio a MS Visual Studio. Na závěr jsou také představeny možnosti reportingu pomocí známého nástroje MS Excel a shrnuty přínosy navrženého BI/DWH řešení.

### 3 Business Intelligence

Business Intelligence (dále jen BI) je komplexní sada procesů, konceptů, metodik, know-how a technologií, jejichž smyslem je nejen účelně, ale rovněž účinně podporovat a zlepšovat rozhodovací a řídicí procesy ve společnosti [1].

Účelem BI je konvertovat velké objemy dat na poznatky, které jsou významné pro koncového uživatele. Tyto poznatky potom mohou být efektivně využity právě v procesu rozhodování, jehož výsledkem může mnohdy být vytvoření konkurenční výhody. [2] Síla BI tedy spočívá zejména ve schopnosti poskytnout pohled na skutečnost, získat znalosti, umožnit odůvodnit možný problém a podpořit plánování či inovace v budoucnosti [3].

Výsledné BI systémy pak můžeme považovat za efektivní, vyznačují-li se tím, že jsou aktuální, přesné, jejich přínos je vysoký a podporují již zmíněné rozhodování. Pojmem aktuální se myslí, že data jsou v rámci daného BI řešení dostupná podle předem vytvořeného rozvrhu, vysoký přínos spočívá v užitečnosti pro uživatele podnikových informací a podpora rozhodování souvisí s uplatněním získaných informací v podnikovém rozhodovacím procesu. [4]

#### 3.1 Oblasti využití BI

Podle [5] je možné Business Intelligence nástroje aplikovat prakticky kdekoliv, kde je potřeba nějakým způsobem analyzovat a vyhodnocovat zdrojová data. Z toho důvodu v dnešní době existuje celá škála aplikací a BI řešení, která se specializují na určitou oblast. Můžeme se však často setkat i s případy, kdy se jednotlivé oblasti svojí funkcionalitou částečně překrývají. Proto vznikají univerzální řešení, která pro konkrétní nasazení stačí mírně přizpůsobit. V každém případě však aplikace BI pokrývají plánovací, analytické a rozhodovací funkce většiny z oblastí podnikového řízení.

Mezi nejčastější oblasti využití BI nástrojů můžeme zařadit následující [5]:

- Obchod – BI zde nachází široké uplatnění v analýze prodeje dle různých kritérií (např. sezónnost, oblast, kategorie produktu atd.).

- Finance – BI zde slouží ke kontrole finančního hospodaření společnosti, analýze nákladů a výnosů, porovnání plánu vůči skutečnosti apod.
- Výroba – BI v této oblasti poskytuje komplexní přehled nejen o historii, ale i stavu výroby včetně struktury zařízení spojených s kontrolou požadované kvality. Dále aplikace pro plánování a monitorování výrobních procesů, obrátů zásob, průchodností linek atp.
- Logistika – v rámci logistiky BI slouží především k analýze dopravců, dopravních nákladů, reklamací, kapacitnímu plánování a doby dodávky.
- Marketing – BI zde pomáhá analyzovat a plánovat portfolio produktů a služeb, spravovat, sledovat a vyhodnocovat marketingové kampaně a segmentovat zákazníky dle reakcí na dané kampaně.
- Personalistika – v této oblasti BI umožňuje analyzovat pracovní síly, vytížení, výběr a motivaci zaměstnanců a náklady s nimi spojené.

### **3.2 Důvody zavedení BI**

Důvodů, proč je v současnosti dobré zavádět, případně klást důraz na používání a udržování Business Intelligence řešení ve společnosti, je hned několik, a jsou jimi zejména tyto [6]:

- Snižování nákladů – pomocí BI je možné identifikovat, kde přesně dochází k plýtvání potřebnými zdroji a reagovat na to přijetím nezbytného opatření založeného na znalostech získaných z dat nacházejících se v různých zdrojových systémech.
- Zvýšení ziskovosti – BI nástroje mohou poskytnout jednotlivým oddělením společnosti jak ucelený, tak i detailní pohled na problematiku daného oddělení. BI může být nápomocné při určování hodnoty například zákazníků, přičemž je takto možné lépe a přesněji rozpoznat ziskové klienty od těch neziskových. Obdobným způsobem BI umožňuje rozpoznat ziskové či neziskové produkty a služby.
- Snižování rizika – společnosti, které dokážou rozpoznat a rozumět datům uloženým v datových skladech, využíváje přitom nástrojů BI, jsou schopny v pravý čas identifikovat potenciální problém. Na rozdíl od konkurence, kte-

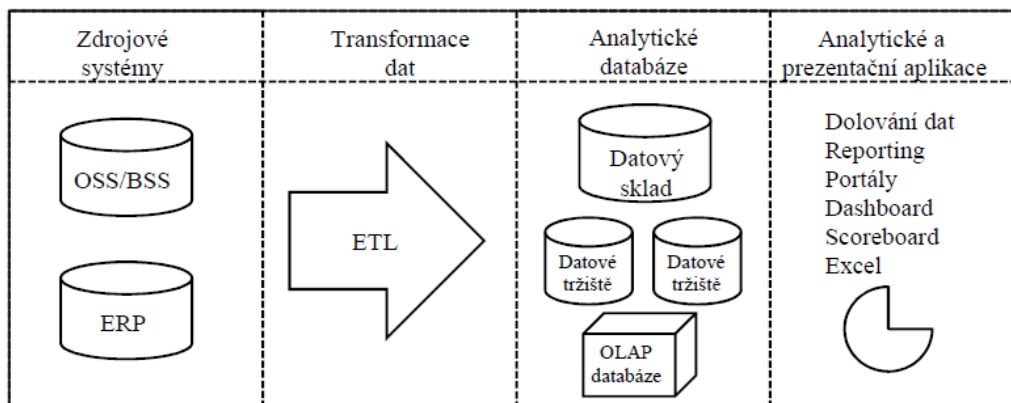
rá nástroje BI nevyužívá, jsou tak připraveny včas odhalit možná nebezpečí, a tím získat drahocenný čas na vypořádání se s těmito hrozbami odpovídajícím způsobem.

- Zlepšení vztahů se zákazníky – BI nástroje nabízejí možnost zjistit nedostatky v oblastech, které mají vliv na spokojenost zákazníků. Společnosti využívající BI řešení jsou schopny mnohem lépe shromažďovat a analyzovat informace o svých zákaznících a hledat v nich možné souvislosti. To v konečném důsledku vede k lepším, rychlejším a přesnějším službám pro zákazníky.

### 3.3 Architektura BI

Architektura BI popisuje jednotlivé součásti, které je zapotřebí spojit a rovněž zabezpečit jejich řádné fungování v rámci celého řešení [7]. Vymezení a uspořádání komponent v rámci BI řešení se však odvíjí od specifických potřeb a požadavků jednotlivých zákazníků. Můžeme se tak setkat s implementacemi různé složitosti, od těch nejjednodušších až po sofistikovaná, náročná a mnohdy velmi komplexní řešení. Na výsledné složitosti daného řešení pak závisí alokace pracovních, technologických a zejména finančních prostředků. [1]

I přesto, že se architektury jednotlivých řešení často liší, je možné najít společné rysy, a tím pádem i obecnou koncepci. Ta stanovuje, co by architektura BI řešení měla obsahovat a z jakých komponent by se měla skládat. [5] Tuto situaci ilustruje obrázek č. 1.



**Obr. 1 Obecná architektura BI**  
Zdroj: zpracováno dle [5]

Obecně lze architekturu BI rozdělit do několika základních vrstev, o nichž budou dále pojednávat následující podkapitoly. Jedná se o následující vrstvy [8]:

- Vrstva zdrojových systémů
- Transformační vrstva
- Databázová vrstva
- Analytická vrstva
- Prezentační vrstva

Je na místě rovněž podotknout, jak důležitý je správný návrh architektury modelovaného BI řešení. V případě, že architektura není navržena správně, může docházet k nekonzistencím, které mohou narušit chod celého systému. [1] Příkladem problému, který může nastat, je neschopnost přenášet informace mezi komponentami, neschopnost plnit stanovené požadavky nebo slabý výkon systému jako celku. V nejhorším případě pak může dojít k situaci, kdy jsou data nesprávně prezentována v nesprávnou dobu či nesprávné osobě [8].

### **3.3.1 Vrstva zdrojových systémů**

Jako zdrojové systémy můžeme označit různorodou množinu informačních systémů, aplikací, databází a souborů, ze kterých BI aplikace získávají data, ale nejsou přímou součástí BI řešení. Pro samotnou funkčnost mají však velký význam.

Jejich hlavním rysem je podpora ukládání a modifikace dat v reálném čase. [1] Mezi zdrojové systémy se ve většině BI řešení primárně řadí informační systémy, jako jsou například ERP (Enterprise Resource Planning, plánování podnikových zdrojů), CRM (Customer Relationship Management, řízení vztahů se zákazníky), SCM (Supply Chain Management, řízení dodavatelského řetězce), nebo jiné interní systémy určené ke správě účetnictví či lidských zdrojů. Jedná se o systémy, které jsou postaveny nad relačními databázemi a většinou obsahují informace o jednotlivých transakcích. [5]

Dalšími zdroji dat mohou být i soubory různých formátů, ať už textové podoby, tak soubory z tabulkových kalkulačků [1]. V dnešní době není výjimečná ani

integrace dat z externího prostředí organizace například od partnerů, dodavatelů, státní správy či z různých internetových zdrojů a sociálních sítí. Před implementací BI řešení je však důležité všechny zdrojové systémy pečlivě zanalyzovat a k integraci vybrat pouze ty relevantní a ty, ze kterých chceme data ukládat do datového skladu. [7]

### **3.3.2 Transformační vrstva**

Transformační vrstva, z anglických slov zkráceně označovaná jako ETL (Extract, Transform, Load) či pomocí českého ekvivalentu „datová pumpa“, patří mezi nejdůležitější součásti BI řešení. Její úlohou je získat (Extract) data ze zdrojových systémů, upravit (Transform) tato data do požadované struktury a uspořádání a následně je nahrát (Load) do specifických datových struktur a schémat datového skladu či datového tržiště. [9] ETL procesy jsou spouštěny v dávkovém režimu, data jsou tedy přenášena v předdefinovaných časových intervalech.

Transformace dat spojené s ETL představují jednu z nejnáročnějších činností v rámci BI řešení, ať už z časového, pracovního, nebo i z finančního hlediska a obvykle spotřebují cca 60 % vynaložených pracovních kapacit. Tyto transformace však představují nezbytný předpoklad k úspěšné realizaci daného řešení. [7]

Součástí transformační vrstvy jsou i nástroje EAI (Enterprise Application Integration). Tyto nástroje se primárně používají k redukci vzájemných komunikačních kanálů mezi zdrojovými systémy, které jsou mezi sebou navzájem propojeny. V BI se však využívají z důvodu přenosu dat do datových úložišť, čímž vzniká možnost analýzy dat v reálném čase. [10]

Jak již bylo zmíněno výše, ETL vrstva je jedna z nejdůležitějších komponent BI řešení. Z toho důvodu jí bude v této práci věnována větší pozornost v kapitole 7.

### 3.3.3 Databázová vrstva

Databázová vrstva obsahuje databázové komponenty a zajišťuje tak procesy ukládání, aktualizace a správy transformovaných a očištěných dat. Základní a zároveň povinnou součástí této vrstvy je datový sklad. V rámci databázové vrstvy se však mohou vyskytovat i další BI komponenty, jako jsou datová tržiště, dočasné úložiště či operativní úložiště dat. [5] Více o datových skladech a jeho komponentách pojednává kapitola 6.

Forma dat v datové vrstvě se může lišit od klasických přístupů databázového modelování, a to z důvodu vícedimenzionálního přístupu pro analýzu. V transakčních systémech jsou data uložena ve třetí normální formě, která zabezpečuje zachování datové integrity a odstranění duplicit. V analytických systémech je však potřeba dívat se na data z více hledisek současně. Uživatel buď musí vidět jednotlivé vztahy mezi daty, nebo je nutná rychlá dostupnost agregovaných atributů, které se v normalizované databázi do datové struktury ukládat nesmí. [11]

### 3.3.4 Analytická vrstva

Jak již z názvu vrstvy vyplývá, účelem analytické vrstvy je zpřístupnit data pro potřeby dalších analýz, ať už ve formě standardního reportingu pomocí databázových dotazů, či na pokročilejší analytické úlohy typu dolování dat. Tato vrstva je využívána zejména řídicím managementem organizace. Mezi hlavní komponenty analytické vrstvy patří multidimenzionální databáze OLAP (Online Analytical Processing), která je tvořena spojením několika vícedimenzionálních tabulek. [2]

Pro efektivní vytváření přehledů a analýz nad daty je potřeba mít možnost se na tato data podívat z několika pohledů (dimenzí) [7]. Jen tak je možné najít v datech na první pohled neviditelné souvislosti. Právě multidimenzionální databáze jsou navrženy tak, aby bylo možné získat přehledy o datech v různých kombinacích dimenzí. To vyžaduje specifickou organizaci dat v dané databázi. K uchování potřebné struktury se využívají dva postupy. Prvním z nich je multidimenzionalita na úrovni relačních databází a druhým multidimenzionalita na úrovni OLAP. [2] Podrobnější popis OLAP nabízí kapitola 8.



### 3.3.5 Prezentační vrstva

Poslední vrstvou referenčního modelu BI je vrstva prezentační. Tato vrstva slouží k samotné prezentaci výsledných dat, a to tak, aby podpořila firemní procesy a rozhodování. Je tedy zodpovědná za sběr požadavků na analytické operace, přičemž představuje nástroje pro koncové uživatele BI aplikace. [7]

Výsledky analýz mohou být zpřístupněny pomocí různých rozhraní například v prostředí internetového prohlížeče či v jiné univerzální aplikaci, jako je MS Excel. Do této oblasti můžeme zařadit i analytické a manažerské aplikace EIS (Executive Information Systems), které v sobě integrují nejdůležitější zdroje dat systému významné pro řízení společnosti jako celku. [1] Bližší specifikace těchto typů aplikací jsou následující:

- Analytické aplikace – zajišťují především sběr požadavků na dotazování a následné zobrazení výstupů od uživatelů. Umožňují sledovat tzv. manažerské informace (firemní procesy, naplnění cílů podniku atp.) nebo online analýzu dat, včetně analýzy trendů. Důraz je rovněž kladen na jednoduché ovládání aplikace pomocí grafického uživatelského rozhraní. [8]
- Manažerské aplikace – podporují manažerské procesy jako podnikové analýzy, plánování a rozhodování. Využívané mohou být všemi vrstvami managementu společnosti. Na rozdíl od reportování si manažerské aplikace vytvářejí vlastní vícedimenzionální vrstvu, prostřednictvím které uživatelé přistupují k datům. Díky uživatelskému prostředí je odstíněna potřeba znalosti dotazovacího jazyka SQL (Structured Query Language). Nevýhodou tohoto přístupu je však úzká škála předdefinovaných dotazů. [7]

## 4 Aktuální trendy a analýza trhu BI

Business Intelligence představuje jednu z mnohých oblastí informačních technologií, která se neustále a zároveň velmi rychlým tempem rozvíjí. Data jsou všudypřítomná a jejich objem neustále roste. Firemní prostředí, konkurence i vnější vlivy se navíc vyvíjejí stále rychleji. Manažeři na všech úrovních jsou tedy nuceni dělat čím dál větší počet rozhodnutí a mají na ně stále méně času. To vyvolává poptávku po BI nástrojích, které dokážou taková rozhodování podpořit, snadno se ovládají, ale kde jednoduchost použití není na úkor robustnosti a schopnosti integrovat různé datové zdroje.

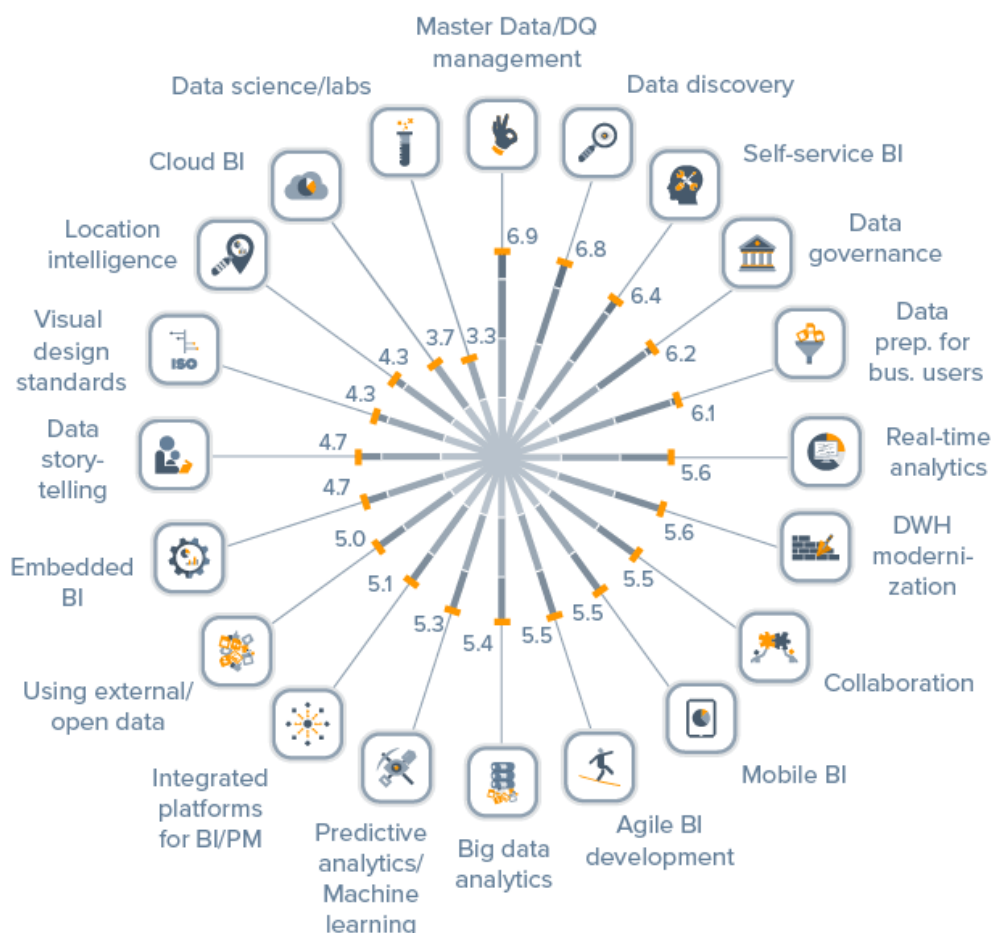
Není se tedy čemu divit, že je internet plný předpovědí a očekávaných trendů v rámci vývoje BI a jejího trhu. To ostatně dokazují rozborů předních analytických společností, jako jsou například BARC a Gartner, jejichž studie budou přibližně v následujících podkapitolách.

### 4.1 Aktuální trendy

Celý IT obor prochází dlouhodobě změnami a neustálým vývojem. V minulosti ho výrazně ovlivnilo rozšíření osobních počítačů, architektura client/server či rozvoj internetu. V dnešní době se hovoří zejména o oblastech jako digitalizace, propojení soukromého a firemního využití IT technologií, bezpečnost, cloud, mobilní zařízení či analýza informací. Všechny tyto trendy představují hlavní hybatele změn i při rozvoji a používání Business Intelligence. Z toho důvodu mnoho BI profesionálů každoročně přemýšlí o tom, jaké směry v jejich oboru převládnu v blízké budoucnosti. [12]

K identifikaci aktuálních trendů v oblasti BI můžeme využít studii „BI Trend Monitor 2018“ od společnosti BARC, což je jedna z předních evropských konzultáčních a analytických společností, která se zabývá podnikovými informačními systémy [13]. V rámci studie bylo téměř 2800 respondentů z řad dodavatelů BI, jejich uživatelů a odborných konzultantů dotázáno, které z nových trendů budou v roce 2018 z jejich pohledu ty nejdůležitější. Úkolem každého respondenta bylo stanovit důležitost daných trendů na stupnici od 0 do 10 (není vůbec důležitý –

velmi důležitý). [14] Výsledky zkoumaných trendů v rámci studie ilustruje následující obrázek číslo 2.



**Obr. 2 Výsledky studie BI Trend Monitor 2018**

Zdroj: převzato z [14]

Na prvních místech se umístily následující trendy: řízení kvality dat (Master Data, Data Quality management), vizualizace a interpretace dat (Data Discovery, Data Visualization) a BI ovládané uživatelem (Self-Service BI), též zvané jako samoobslužné BI. Porovnáme-li výsledky letošní studie s těmi loňskými, situace na prvních třech místech se změnila jen nepatrně, kdy se trend řízení kvality dat vyhoupl ze třetí pozice na pozici jedničky, čímž sesunul vizualizaci dat a samoobslužné BI vždy o jednu pozici níže. [14]

Nutno poznamenat, že tyto tři trendy se na prvních třech příčkách umístily i pro rok 2016. To jen potvrzuje skutečnost, že například samoobslužné BI již

dlouhou dobu patří mezi žádané trendy, a to zejména z důvodu navyšujícího se zájmu podniků o BI řešení, bez nutnosti zaměstnávat značný počet BI specialistů.

Respondenti zároveň vysokým hodnocením trendu řízení kvality dat rozpoznávají klíčovou důležitost pracovat se správnými a korektními daty v BI systémech, jelikož bez nich by tyto systémy byly de facto k ničemu. [15] Naopak některé z moderních přístupů, o kterých se uvažuje jako o budoucnosti Business Intelligence, skončily na zcela opačném konci preferencí BI specialistů. Zařadily se mezi ně datové laboratoře (Data Labs/Science) a překvapivě také cloudové BI (Cloud BI/Blaas).

Zdá se, že BI profesionálové dávají přednost praktické použitelnosti BI a kvalitě jeho výstupů před vizionářskými nápady a ideami. [12] Nicméně co se BI v cloudu týče, jeho nízká důležitost v očích respondentů může být způsobena především z bezpečnostních důvodů. Spousta společností se totiž obává, že jejich data budou v cloudu buď zneužita, nebo ztracena. Mají tedy větší pocit bezpečí, když jsou veškerá firemní data uložena na vlastních lokálních serverech (tzv. On Premise). [14]

#### **4.1.1 Řízení kvality dat**

Lidé mohou dělat správná rozhodnutí pouze na základě kvalitních a správných dat, a to je přesně to, co má řízení kvality dat za primární cíl. Úkolem je tedy sesbírat a vzájemně si vyměňovat kmenová data (Master Data) mezi několika zainteresovanými systémy. [14] Kmenová data jsou klíčovým zdrojem informací pro fungování každé společnosti. Jsou to data o objektech, například různé informace o dodavatelích, zákaznících a obchodních partnerech, o zdrojích (zejména o financích, majetku, lidských zdrojích), o produktech, projektech, materiálech či výrobních kapacitách. Na rozdíl od dat transakčních mají kmenová data dlouhodobý charakter a je u nich tedy důležitá kvalita. [16]

Kmenová data jsou sdílena různými obchodními procesy a mohou být uložena i ve více aplikacích, typicky v ERP, CRM, SCM a dalších systémech [14]. Je-

jich špatná kvalita je v praxi nejčastějším důvodem vzniku chyb, dohledávání a špatných rozhodnutí, tedy vzniku zbytečných procesních výdajů.

Kvalita kmenových dat, jejich ukládání a řízení jsou tak pro chod společnosti podstatné. Tato data by měla být ukládána a spravována tak, aby byla jediným zdrojem „pravdy“. Nezávisí na tom, zdali jsou ukládána distribuovaně, nebo centrálně. Důležité je, aby společnost měla procesy jejich změn a vytváření pod kontrolou a zabránila tak nežádoucí duplicitě, redundanci, či jiné možné nekvalitě. Řízení kvality dat však není jednorázová záležitost. Jedná se o nekončící cyklus, jelikož je potřeba neustále se o nová data starat a ta stará spravovat. [14, 16]

#### **4.1.2 Vizualizace a interpretace dat**

Vizualizace a interpretace dat patří v současnosti k nejdůležitějším trendům v oblasti BI, a to nejen podle průzkumu společnosti BARC. Vhodná a užitečná vizualizace dat však nepředstavuje pouze působivé barevné grafy a diagramy. [14] Každá společnost disponuje enormním množstvím dat, která přibývají každým dnem. Právě toto množství dat je potřeba pochopit a analyzovat tak, aby bylo možné objevit vzájemně skryté souvislosti mezi těmito daty. [20] Vizualizace a interpretace dat se tedy používá nejen pro účely reportingu, analýzy dat, ale funguje rovněž jako prostředek pro objevování nových informací v datech (Data Discovery). [14]

Novodobá vizualizace nahrazuje původní způsob interpretace informací v tabulkách, textech a statických grafech. V současnosti se tak stává nástrojem, ve kterém se používají barvy, tvary a různé grafické prvky, které mají za úkol urychlit a usnadnit poskytnutí informací, zpřehlednit je a zkrátit čas potřebný pro jejich porozumění. To v konečném důsledku vede k zefektivnění fungování podniků. [20]

Data v prostém formátu nic neříkající běžnému uživateli bez zvláštních technických znalostí se transformují do vizuální podoby, čímž umožňují těmto uživatelům daným datům a informacím v nich obsažených poměrně snadno a rychle porozumět. Běžní uživatelé z jednotlivých oddělení (obchod, marketing, finance)

tak mohou díky vizualizaci dat sami využívat analytické nástroje a objevovat tak nové, dosud skryté informace, predikovat trendy atd. [21]

Mezi pokročilé aspekty vizualizace a interpretace dat řadíme [21]:

- interaktivní grafické prostředí,
- integrované intuitivní a jednoduše dostupné analytické funkce,
- snadné vytváření reportů,
- snadnou distribuci reportů a analýz na webové portály a mobilní zařízení.

#### **4.1.3 Samoobslužné BI**

Nástroje typu samoobslužného BI (Self-Service BI), představují další z významných trendů v oblasti BI. Účelem je poskytnout nové technologie pomocí jednoduchého a uživatelsky intuitivního rozhraní, které umožní vykonávání pokročilých analýz a vizualizaci dat nejen odborným, ale i běžným uživatelům bez potřeby velkých technických znalostí a potřeby využívat složité BI systémy [18]. Netechničtí uživatelé tak mají více možností, větší flexibilitu a samostatnost. Mohou si prohlížet data, vytvářet vlastní reporty a nemusí tak čekat na pomoc odborných pracovníků, aby jim vygenerovali potřebné sady dat z různých zdrojů nebo samotné reporty dokonce vytvořili. [17]

Tento typ řešení představuje opačný přístup, než jakým je tradiční BI. Charakter BI aplikací se tak postupně přesouvá od aplikací, které byly závislé na IT oddělení a mění se na aplikace, které uživatel používá samostatně bez potřebné IT podpory. Tyto aplikace jsou intuitivnější, efektivnější a mnohem více agilní. Mohou být používány kýmkoliv, ať už pracovníkem finančního oddělení, či pracovníkem marketingu.

Z výše uvedeného dále vyplývá zrychlení a usnadnění celého procesu reportingu a analýzy dat, neboť uživatelé již nemusejí čekat na specialisty z IT oddělení. Tím se zrychluje a usnadňuje i rozhodování, které je klíčové v obchodním prostředí. Také se uvolní kapacity v IT oddělení, což znamená, že se BI specialisté budou moci zabývat opravdu komplexními a složitými požadavky či problémy týkajícími

se reportingu, analýzy dat, zabezpečení chodu BI řešení atd. [18] Snížením závislosti na IT oddělení se rovněž zvyšuje podniková flexibilita, což je ve světě podnikání velmi důležitá vlastnost. Dále je nutno poznamenat, že samoobslužné BI řešení má oproti tradičním BI řešením daleko nižší finanční, provozní i zdrojovou náročnost. [19]

Mezi určitá omezení či nevýhody samoobslužných BI řešení by se dalo zařadit, že komplexní a složité úlohy celopodnikového charakteru stále musí řešit odborníci z IT oddělení. Dále je nutné, aby i běžní uživatelé byli určitou formou proškoleni a ovládali tak alespoň základní principy analytických metod. [17]

## **4.2 Analýza trhu**

Jednotlivé segmenty trhu na poli informačních technologií se obvykle nachází v různých fázích vyspělosti, dodavatelé navíc volí různorodé konkurenční strategie, a tak není jednoduché udržet si dobrý přehled. Ne jinak je tomu i na přesyćeném a chaotickém trhu v případě Business Intelligence. [25]

K řešení této situace nám však poslouží každoroční studie „Magic Quadrant for Business Intelligence and Analytics Platforms“ od společnosti Gartner, což je přední světová analytická společnost zabývající se poradenstvím a výzkumem v oblasti IT. [22] Zmíněná studie, česky přezdívaná jako „magické kvadranty“, slouží k přehlednému zmapování situace na trhu BI. Nabízí tak dodavatelům, investořům či přímým uživatelům přehled o tom, jak si poskytovatelé BI platform vedou v souladu se stanovenou vizí a schopností ji realizovat, tedy dosáhnout kýžených výsledků. [24] V době psaní této práce je poslední dostupnou verzí studie pro rok 2017 (viz obrázek 3), nicméně se očekává, že v následujících týdnech či měsících bude k dispozici verze i pro rok 2018.

Metoda magických kvadrantů dělí poskytovatele, kteří spolu soupeří na rostoucích trzích v oblasti BI, do čtyř kvadrantů. Popis a úloha jednotlivých kvadrantů vypadají následovně [23]:

- Vyzyvatelé – se nacházejí v dobré pozici z hlediska realizace, ale nemusejí mít dostatečně silnou strategii, aby si dlouhodobě udržovali schopnost do-

dávat řešení s příslušnou přidanou hodnotou. Přestože mnohdy disponují dostatečnými finančními a lidskými zdroji, můžou postrádat vizi, schopnost inovace či celkové pochopení toho, kam trh v oblasti BI směřuje. V některých případech mohou vyzyvatelé nabízet služby či produkty, které dominují velké části trhu, jejíž význam se však časem snižuje. Mají rovněž příležitost přesunout se do kvadrantu lídrů, a to tím, že rozšíří svou vizi.

- Specializovaní hráči – si vedou dobře pouze ve specifické části trhu, jelikož mají omezené schopnosti inovace nebo podávání lepších výkonů nežli jiní dodavatelé. Důvodem může být to, že se zaměřují jen na určitý geografický region či specifickou funkcionalitu, případně jsou to „noví hráči“ na trhu. Klasifikace této skupiny poskytovatelů bývá náročnější než klasifikace poskytovatelů ve zbylých kvadrantech. I přesto, že někteří z nich mohou růst, jiní si nemusejí vést dobře a nemusejí být schopni držet krok s požadavky trhu.
- Lídři – poskytují pokročilé služby, které vyhovují požadavkům současného trhu BI. Tito poskytovatelé rovněž disponují potřebnou vizí k udržení svých vedoucích pozic v rámci vývoje trhu. Lídři se zaměřují zejména na aktuálně nabízené produkty a služby, zároveň však do nich investují, aby tak ovlivnili celkové směřování trhu.
- Vizionáři – mají na vývoj trhu stejný pohled jako společnost Gartner, ale není u nich jisté, zdali budou schopni poskytovat řešení v souladu se svou stanovenou vizí. Na nových trzích je toto chování normální, na těch vyspělejších však taková pozice může odrážet konkurenční strategii menšího dodavatele, který se snaží odlišit od zbytku trhu. Z pohledu zákazníků a dodavatelů patří vizionáři do skupiny, která podstoupením vyšších rizik může zajistit dosažení vyšších příjmů. Často totiž představují nové služby, obchodní modely či technologie, obvykle si však nejprve musejí vybudovat finanční sílu, podporu prodeje a distribuční kanály.





**Obr. 3 Analýza trhu 2017 dle Gartner**

Zdroj: převzato z [22]

Ze studie vyplývá, že trhu dnes dominují dodavatelé, kteří se zaměřují na jednoduchost používání a agilitu, zároveň však i na schopnost řešení uchopit pevnou ruku nad procesy a dohlížet tak na distribuci, tvorbu a využití analytického obsahu v rámci daného řešení. [25] Lídři na trhu BI musejí podle studie Gartner zaujmout jednoduchým, srozumitelným a cenově atraktivním modelem. V případě moderního BI a trhu s analytickými technologiemi jsou totiž nákupní rozhodnutí ovlivňována zejména byznys uživateli a ne IT odděleními. Tito uživatelé primárně vyžadují jednoduše zvládnutelné nástroje, které přinášejí jasnou obchodní hodnotu a podporují silnou analytiku ve společnosti. To vše bez nějaké technické expertízy a výraznější potřeby angažovat IT oddělení. [26]

Pouze tři dodavatelé vyhověli výše uvedeným požadavkům a zařadili se tak do kvadrantu lídrů. Konkrétně se jedná o společnosti Qlik, Tableau a Microsoft. Významní hráči z hlediska tržního podílu jako IBM, SAS, MicroStrategy, SAP, Micro-

soft a Oracle si sice díky svým BI platformám vytvořily v minulosti významnou klientskou základnu, ale až na Microsoft již nejsou uvedeni mezi lídry trhu. Pro umístění v této skupině je dle Gartner studie nutné mít schopnost přilákat novou generaci klientů s požadavky, které se zásadně liší od těch v minulosti [26]. Výsledkem je, že se mnoho z těchto organizací přemístilo do kvadrantu vizionářů. Díky změně hodnotících kritérií se navíc žádný dodavatel BI nedostal do kvadrantu vyzyvatelů, kam organizace Gartner zařazuje dodavatele s velkou schopností realizace, ale s nedostatečně silnou strategií. Do kvadrantu specializovaných hráčů bylo zařazeno několik známých společností jako například Oracle, Birst či open source Pentaho. [22]

## 5 Dimenzionální modelování

Dimenzionální modelování je dle [29] speciální způsob návrhu databázového modelu. Na rozdíl od klasického návrhu databázového modelu používaného v běžných databázových systémech (normalizovaných alespoň ve třetí normální formě), je dimenzionální model mnohem intuitivnější, srozumitelnější a také přehlednější pro koncové uživatele.

Normalizovaný databázový model je navržen tak, aby eliminoval duplicitní a redundantní data pomocí normalizačních pravidel. Kvůli tomu rozkládá data do většího počtu samostatných entit (databázových tabulek). [9] Snížení redundance dat má pak za následek snížení celkové velikosti normalizovaného systému [33]. Databáze vytvořená tímto způsobem je tak vhodná pro zdrojové systémy, zejména díky rychlosti a optimalizaci vykonávání operací insert, update a delete. Avšak z důvodu velkého množství udržovaných spojení mezi jednotlivými entitami a s tím související pomalou odezvou při rozsáhlých dotazech je tento model nepříliš vhodný pro analytické zpracování dat. [29]

Naopak dimenzionální model databáze je pro analytické dotazy velmi vhodný a tedy velice populární pro provádění složitých analýz za účelem podpory rozhodování v rámci společnosti [29]. Takto vytvořené databáze tak slouží jako podklad pro získávání agregovaných a sumarizovaných údajů, tedy vlastně informací [4]. Tento model se rovněž stal základem pro zpracování analytických dat prostřednictvím tzv. datových kostek – technologie OLAP (více viz kapitola 8), která umožňuje efektivně zpracovávat velké objemy dat. Dimenzionální modelování navíc umožňuje ukládat i redundantní data. Nesplňuje tak běžně používané databázové normalizace, nicméně razantně snižuje potřebný počet databázových spojení, což v konečném důsledku vede ke zvýšení výkonu při následném dotazování. [9]

Dimenzionální modely databází se skládají ze dvou typů struktur, a to z centrálních tabulek faktů a k nim přidružených tabulek dimenzí [33]. Každá tabulka faktů by přitom měla být v normalizovaném stavu (zpravidla ve třetí normální formě), zatímco tabulky dimenzí ve stavu denormalizovaném [31]. Denormalizo-

vaná databáze pak představuje databázi s určitým objemem redundantních dat, která prozatím neprošla procesem normalizace [3].

## **5.1 Fakta**

Tabulka faktů je jádrem dimenzionálního modelu. Jsou v ní uložena analyzovaná a sumarizovaná data společnosti s cílem dosáhnout lepšího pochopení vlastního podnikání. [34] Jeden záznam v tabulce faktů tak představuje určitou míru či hodnotu, která by v ideálním případě měla být vyjádřitelná v číselné podobě [2]. To z důvodu, aby bylo možné kvantifikovat rozsah analyzované události jako například množství prodaného zboží, dobu hovoru či počet objednávek.

Co se počtu sloupců měrných ukazatelů týče, jsou tabulky faktů velmi malé, zahrnují spíše cizí klíče připojených tabulek dimenzí. Obsahují však zpravidla velké množství záznamů a díky tomu zabírají až 90 % z celkové kapacity dimenzionální databáze. [31]

Významnou roli hraje v dimenzionálním modelu aditivita a granularita faktických záznamů. Aditivita je vlastnost, která určuje, zda lze fakta agregovat napříč dimenzemi. [31] Podle toho rozlišujeme fakta aditivní (metriky, které lze agregovat přes všechny dimenze), semiaditivní (metriky, které lze agregovat pouze přes některé dimenze) a neaditivní (metriky, které nelze agregovat podle žádné dimenze) [4].

Granularita pak představuje podrobnost záznamů uložených v tabulce faktů. Určení granularity dat je při tvorbě datového skladu jednou z nejdůležitějších úloh. Zvolíme-li příliš nízkou úroveň granularity, nebude možné pracovat s detailními daty. V případě zvolení vysoké granularity naopak dochází k explozi dat, manipulace s daty se stává časově náročnější a nároky na diskový prostor jsou mnohem vyšší. [31] Existují však obecná doporučení, která uvádějí, že pokud to technická kapacita umožňuje, měla by být data uložena v co nejnižší možné úrovni detailu, tzn. s použitím atomických dat. Další takové doporučení uvádí, že by se všechny tabulky faktů měly nacházet na stejné úrovni granularity, jinak by mohly být velmi nepřehledné. [34]

## **5.2 Dimenze**

Tabulky dimenzí obsahují logicky či organizačně uspořádané údaje, které kvalifikují metriky v tabulce faktů. Jsou to vlastně popisy obchodních procesů společnosti. [2]

Tabulky dimenzí jsou zpravidla menší než tabulky faktů a data v nich obsažená se nemění tak často. Dimenze tedy obecně obsahují relativně stabilní data [2]. Atributy dimenzí slouží jako primární zdroj různých dotazů a reportů, a tím pádem zastávají v systému datového skladu nepostradatelnou úlohu. Jsou navíc klíčovým prvkem pro vybudování srozumitelného datového skladu. [31]

Ve správně navrženém dimenzionálním modelu obsahují jednotlivé tabulky dimenzí značný počet atributů. Z toho důvodu nejsou výjimkou ani tabulky skládající se například z 50 či 100 sloupců. [33] I přesto jsou však poměrně málo objemné a nezabírají tak více než 10 % z celkové kapacity úložiště datového skladu [31].

Síla datového skladu je pak přímo úměrná hloubce a zejména kvalitě atributů jednotlivých dimenzí. Tabulky dimenzí totiž vysvětlují všechna co, kdo, kdy, kde, proč a jak v rámci obchodních procesů a transakcí společnosti. [34]

## **5.3 Schéma propojení tabulek faktů a dimenzí**

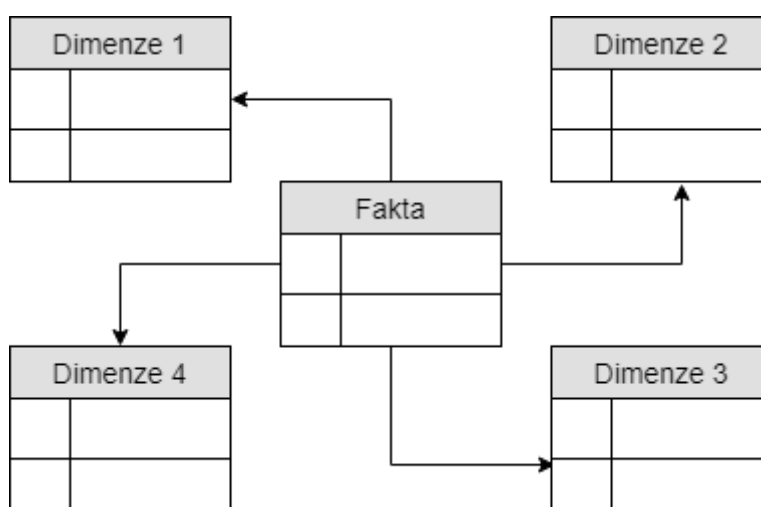
Spojení tabulek faktů a dimenzí tvoří v dimenzionálním modelu specifické topologické uspořádání, které nazýváme schéma [2]. Mezi nejběžněji používaná schémata řadíme schéma hvězdy a schéma sněhové vločky, která se liší uspořádáním tabulek dimenzí.

V případě schématu hvězdy jsou v tabulkách dimenzí zahrnuty i všechny další podřízené úrovně výsledné hierarchie dimenzí, zatímco v případě schématu sněhové vločky je pro každou úroveň hierarchického uspořádání vytvořena samostatná tabulka, která je určitým způsobem napojena na další tabulky v hierarchii. Mezi dvěma takto uspořádanými sousedícími tabulkami je pak vždy použita multiplicita 1:N. [31]

Je na místě zmínit, že se v rámci datového skladu mohou vyskytovat některé z tabulek dimenzí ve schématu hvězdy, zatímco jiné ve schématu sněhové vločky. Vždy záleží na výsledné efektivitě řešení jednotlivých dimenzí [2].

### 5.3.1 Schéma hvězdy

Schéma hvězdy většinou obsahuje jednu rozsáhlou centrální tabulku faktů propojenou s přidruženými tabulkami dimenzí. Pro každou z dimenzí je vytvořena právě jedna tabulka s danými atributy. Velký důraz je přitom kladen na to, aby měl každý z cizích klíčů v tabulce faktů svůj primární klíč v odpovídající dimenzi. [2]

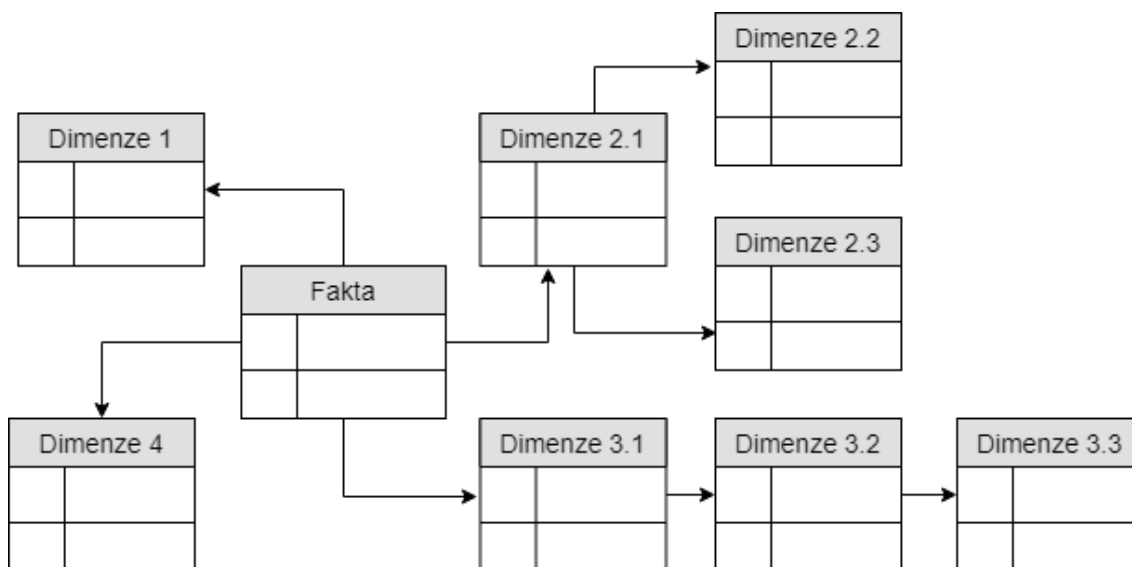


**Obr. 4 Schéma hvězdy**  
Zdroj: zpracováno dle [2]

Schéma hvězdy nemá normalizované tabulky dimenzí a ani neobsahuje jejich relační propojení. To znamená, že odpadá problém se spojováním tabulek v rámci dotazování (nutné je pouze spojení s tabulkou faktů). Schéma hvězdy je tak velice srozumitelné a umožňuje rychlejší získávání výsledků při dotazování. Mezi nevýhody však můžeme zařadit relativně pomalé vytvoření modelu, nemožnost agregace dat díky denormalizaci tabulek či nevhodnost použití při častých změnách prvků v hierarchiích dimenzí. [2] [31]

### 5.3.2 Schéma sněhové vločky

Schéma sněhové vločky obsahuje centrální tabulku faktů a skupinu tabulek dimenzí, které jsou normalizované (tzn. rozdělené do několika samostatných tabulek). Tímto rozpadem vzniká hierarchická struktura tabulek dimenzí (viz obrázek 5). [2]



Obr. 5 Schéma sněhové vločky

Zdroj: zpracováno dle [2]

Schéma sněhové vločky je díky normalizaci tabulek dimenzí velmi vhodné pro časté změny v jejich hierarchickém uspořádání prvků. Výhodou je i možnost vytvoření agregačních tabulek a rovněž menší nároky na potřebnou paměť než v případě předešlého schématu. Naopak mezi nevýhody můžeme zařadit nižší efektivitu analýz, nepřehlednost a větší časovou náročnost zapříčiněnou potřebou spojovat tabulky dimenzí v rámci dotazování. [2, 31]

## 6 Datový sklad

Téma Business Intelligence bylo zařazeno před problematiku datových skladů z důvodu, že se většina systémů datových skladů vytváří proto, aby přinesla určitou firemní hodnotu. Tuto hodnotu zpravidla představuje právě Business Intelligence. Podívejme se nyní na detailnější popis datového skladu.

Datový sklad (Data Warehouse, DWH) je systém, který umožňuje shromažďovat, uchovávat, organizovat a sdílet historická data. Obsahuje „použitá“ data pocházející ze zdrojových provozních systémů společnosti, které data zaznamenávají a používají v kontextu své funkce. Zdrojových systémů využívá datový sklad obvykle více. [4] Datový sklad se běžně koncipuje pro celou společnost, ale v mnoha případech může být zaměřen jen na určitý obor činnosti, jako je obchod, marketing nebo výroba [27].

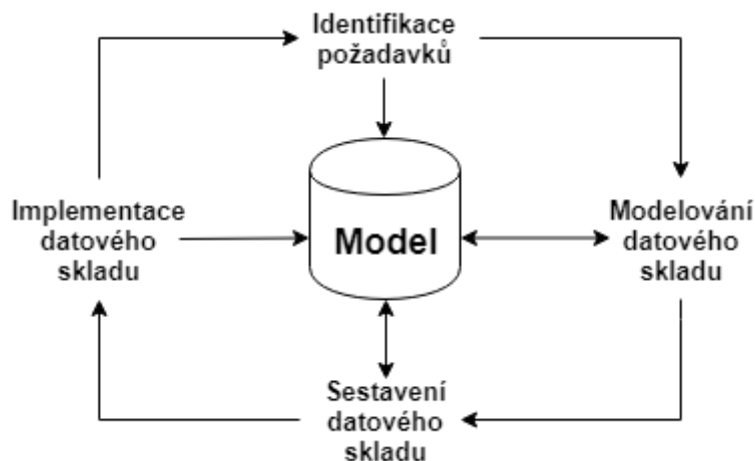
Koncoví uživatelé využívají datový sklad pro podporu rozhodování. Ta je v této souvislosti synonymem Business Intelligence technologie, která je založena na použití dat a způsobu jejich uchování, shromažďování a prezentace v datovém skladu. [4]

Koncoví uživatelé provádějí dotazy na data uložená v datovém skladu. Odpovědi na tyto dotazy jim následně pomáhají provádět potřebná firemní rozhodnutí. Dotazy přitom mohou mít různou podobu od elementárních dotazů, analýz trendů dat za specifické období, komparativní analýzy, prediktivní analýzu budoucího vývoje, dolování dat pro asociativní analýzu až po různé kombinace těchto a mnoha dalších postupů v závislosti na potřebách koncových uživatelů. [29]

### 6.1 Životní cyklus datového skladu

Podle [4] mnoho lidí nerozlišuje mezi projektem datového skladu a systémem datového skladu. Projekt, bez ohledu na svůj účel nebo obor činnosti, má určité datum zahájení a ukončení. Na druhou stranu datový sklad je systém, který se jako každý jiný systém vyznačuje životními cykly, jež znázorňuje obrázek 6.





**Obr. 6 Životní cyklus datového skladu**

Zdroj: zpracováno dle [4]

Životní cyklus datového skladu začíná identifikací nejen firemních potřeb či požadavků, ale také identifikací technických požadavků. Datový sklad je totiž založen na symbiotickém vztahu IT a podnikových potřeb. [33]

Jakmile společnost zjistí, z jakých důvodů usiluje o datový sklad a dospěje ke schválení rozpočtu na jeho budování, začíná fáze návrhu neboli modelování datového skladu. [4] Tato fáze zahrnuje například datovou a technickou architekturu, datové modelování a modelování procesů [33].

Po fázích návrhu, plánování a modelování přichází na řadu vlastní výstavba datového skladu. V rámci této etapy probíhají dílčí úlohy fyzické výstavby prostředí, např. vytváření databáze, indexace, programování ETL procesů, tvorba sestav a podobně. [4]

Na závěr dochází k samotné implementaci datového skladu. Současně se nastavuje produkční stav veškerých komponent a systém se zpřístupňuje koncovým uživatelům, aby mohl aktivně podpořit rozšíření firemních znalostí [27].

## 6.2 Systém datového skladu

V systému datového skladu jsou stejně jako u jiných systémů základními komponentami vstup, zpracování, výstup a také zpětná vazba (viz obrázek 7).



Obr. 7 Systém datového skladu

Zdroj: zpracováno dle [4]

Vstup je v systému datového skladu spojen s identifikací, záznamem a shromažďováním dat, přičemž zde hraje klíčovou roli jejich kvalita. Nesprávná data totiž vedou k nepřesným výstupům, jelikož se přenášejí do dalších procesů, subsystémů, prováděných analýz, a tím pádem i do firemních rozhodnutí [2].

Fáze zpracování má za úkol převod, transformaci, uspořádání, analýzu a ukládání dat ve smysluplné či užitečné podobě. Transformace a nahrávání dat do centrálního prostředí může zahrnovat jak fázi vstupu, tak i fázi zpracování. [4] Jádro celého prostředí tvoří většinou jedna velká databáze, ale stejně tak se může jednat i o sadu databází umístěných na různých serverech [27]. Ty by však bez ohledu na konfiguraci měly být řízeny centrálně, jelikož centrální oblast zastává výkonnou roli při uchovávání a transformaci dat uspořádaným a strukturovaným způsobem. [33]

Výstupní část systému poskytuje přenos dat uživatelům, kteří je potřebují k prezentaci informací ukrývajících se právě v těchto datech. Může zahrnovat materializované dotazy, virtuální pohledy, specializované přístupové cesty apod. [2] Oblast uživatelské prezentace tak představuje část datového skladu, se kterou koncoví uživatelé skutečně pracují.

Zpětná vazba je založena na vstupech a dosažených výstupech systému. Výsledky agregace či odvozování dat mohou být totiž užitečné později. Výstupní data tedy fungují jako cyklický vstup celého systému (jak je vyznačeno na obrázku 5).

Nově odvozené metriky tak pocházejí ze samotného datového skladu a mohou být použity jako podklad pro budoucí dotazy. [4]

### **6.3 Architektura datového skladu**

Pojem architektura může představovat různé perspektivy, vrstvy, komponenty a jejich vzájemné vztahy. Jednoduše řečeno, architektura se váže na způsob, jakým je organizováno a plánováno řešení systému. [4] Klade zároveň důraz na návrh komponent a jejich sdružení na základě převládajícího prostředí a předem specifikovaného použití [33].

Architektura datového skladu tak závisí na mnoha faktorech, jako je strategická vize, uvedení na trh, povědomí o datové architektuře, rozsah dat ve společnosti a regulační zásady. [29] Závisí také na praktických otázkách: Jedná se o vybudování prostředí pro vykazování nebo o iniciativu správy dat jako aktiv společnosti? Je správa dat zajištěna lokálně v odděleních, nebo data spravuje nějaká centrální autorita? Je vhodné data distribuovat nebo centralizovat? [4]

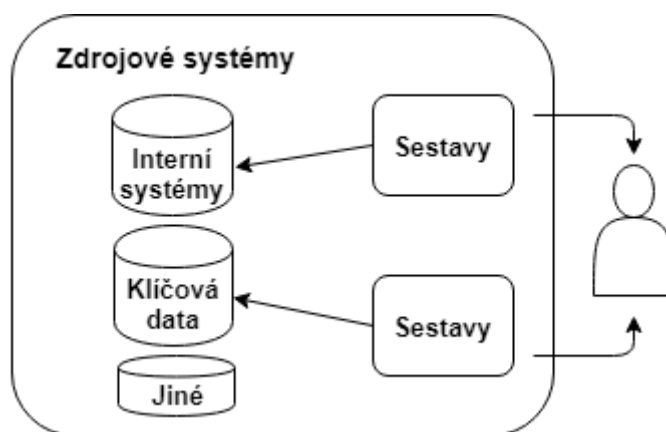
V každém případě je podstatné stanovit strategickou vizi datového skladu v rámci společnosti a rozhodnout o postupu vývoje, jenž musí být ve shodě se směřováním společnosti. Další krok spočívá v určení současného a budoucího použití datového skladu z pohledu správy dat a ve vývoji dané architektury založené na předpokládaném vývoji dat od výchozí implementace k plnému nasazení ve společnosti [2]. Na základě těchto konceptů je možné uskutečnit základy datové architektury, použití, správy a dat jako aktiv společnosti. Následně je možné vytvořit plán činností implementace datového skladu a Business Intelligence řešení. [27]

#### **6.3.1 Architektura dle počtu vrstev**

Systém datového skladu lze vybudovat z několika propojených systémů, které se jednotně označují jako vrstvy (Layer). Díky izolaci těchto vrstev a objasnění jejich vztahů se pak můžeme místo globálního pohledu zaměřit na jednotlivé systémy. [33] Nejběžnější vrstvy datového skladu budou popsány v následujících podkapitolách.

### 6.3.1.1 Architektura s jednou vrstvou

Architektura s jednou vrstvou (viz obrázek 8) získala svůj název z toho důvodu, že nijak nerozlišuje mezi systémem datového skladu a zdrojovým systémem. Architektura tohoto typu se však v oboru datových skladů vyskytuje již jen vzácně, jelikož značně narušuje funkčnost provozního systému. [4]



**Obr. 8 Architektura s jednou vrstvou**

Zdroj: zpracováno dle [4]

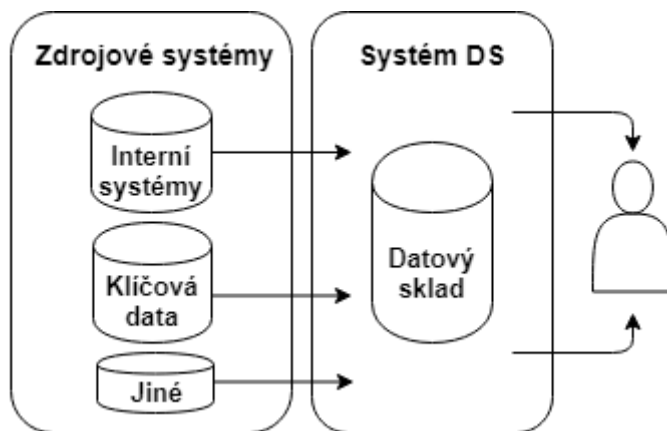
U každé rutiny vykazování nebo analytického dotazu dochází k přímému přístupu ke zdrojovému provoznímu systému. To představuje podstatné navýšení zátěže procesoru a operační paměti serveru a rovněž vstupně-výstupního podsystému, což následně ovlivňuje podnikové operace. V krajním případě může dojít i k narušení klíčových podnikových funkcí. [2]

Jednovrstvá architektura navíc nesplňuje požadavky kladené na architekturu datového skladu, jelikož nerozlišuje mezi transakčními a analytickými oblastmi. Proto se tento přístup z velké části přestal používat již koncem 80. let 20. století. Na druhou stranu pro tento typ architektury platí, že nedochází k duplikaci žádných dat do jiných systémů. Kvůli tomu se nevyskytuje žádná redundance dat a odpadá i režie procesů ETL. [4]

### 6.3.1.2 Architektura se dvěma vrstvami

Architektura datového skladu se dvěma vrstvami zahrnuje systém datového skladu a zdrojové systémy, což představuje architekturu s centrálním úložištěm dat, jak je

znázorněno na obrázku 9. Podstatné je z provozní perspektivy rozlišení mezi zdrojovými systémy a systémem datového skladu. [4]



**Obr. 9 Architektura se dvěma vrstvami**

Zdroj: zpracováno dle [4]

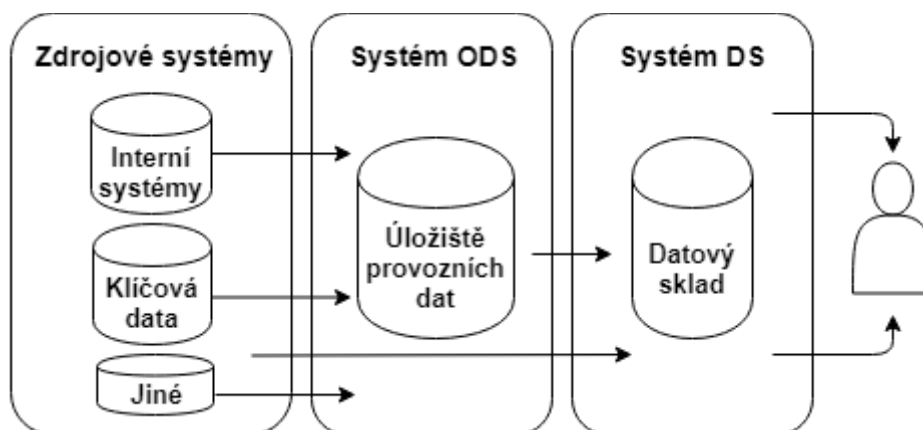
Díky oddělení různých částí systémů je možné se zaměřit na systém datového skladu nezávisle na zdrojových systémech. Na tomto principu je možné dále oddělit základní dílčí komponenty, aby bylo možné identifikovat tok dat a jejich původ. [1]

Rozložení architektury se dvěma vrstvami na jednotlivé komponenty umožňuje jednoduchý přístup k libovolné vrstvě. Současně je možné porozumět funkci a struktuře propojení těchto komponent. Díky architektuře se dvěma vrstvami je tak možné vybudovat nezávislý datový sklad, který lze následně navrhnout s ohledem na maximálně optimalizované použití zaměřující se na jednotlivé vrstvy datového skladu. [4]

Perspektiva nezávislých systémů dovoluje, aby zdrojové či transakční systémy fungovaly nejen bez vzájemných závislostí, ale také bez závislosti na systému datového skladu. Izolace systémů rovněž umožňuje integraci dalších systémů do datového skladu prostřednictvím nových doplňků, avšak při dodržení definované podnikové terminologie [27].

### 6.3.1.3 Architektura se třemi vrstvami

Architektura se třemi vrstvami, jež je znázorněna na obrázku 10, rozšiřuje předešlou dvouvrstvou architekturu o další systém, kterým obvykle bývá úložiště provozních dat (ODS - Operational Data Store). [4] Úložiště provozních dat je zásobováno daty ze zdrojových systémů a zpravidla poskytuje „aktuální“ obraz zákaznických informací. Této vrstvy se rovněž týká správa a kvalita dat, což umožňuje snazší integraci do systému datového skladu. [2]



Obr. 10 Architektura se třemi vrstvami

Zdroj: zpracováno dle [4]

Základní myšlenka spočívá v tom, že ODS představuje samostatný systém čerpající data ze zdrojových systémů, která následně poskytuje systému datového skladu [27]. Z toho vyplývá pojmenování architektury se třemi vrstvami. Systém datového skladu obvykle získává data i z dalších zdrojových systémů, nicméně ODS funguje jako primární zdroj většiny dat. ODS lze rovněž použít i nezávisle na systému datového skladu, ale zpravidla se implementuje společně s datovým skladem dle definic a struktur datového skladu. [31]

Architektura se třemi vrstvami vypadá na první pohled jako optimální kombinace rychlé reakční doby jednovrstvé architektury a izolace datového skladu u architektury dvouvrstvé [4]. Realita je však poněkud složitější, jelikož je u architektury tohoto typu nutné zohlednit sladění datového skladu s ODS prostřednictvím struktur pro správu dat a výkonu. Řešení je tak mnohem pracnější, časově náročnější a i finančně nákladnější. [31]

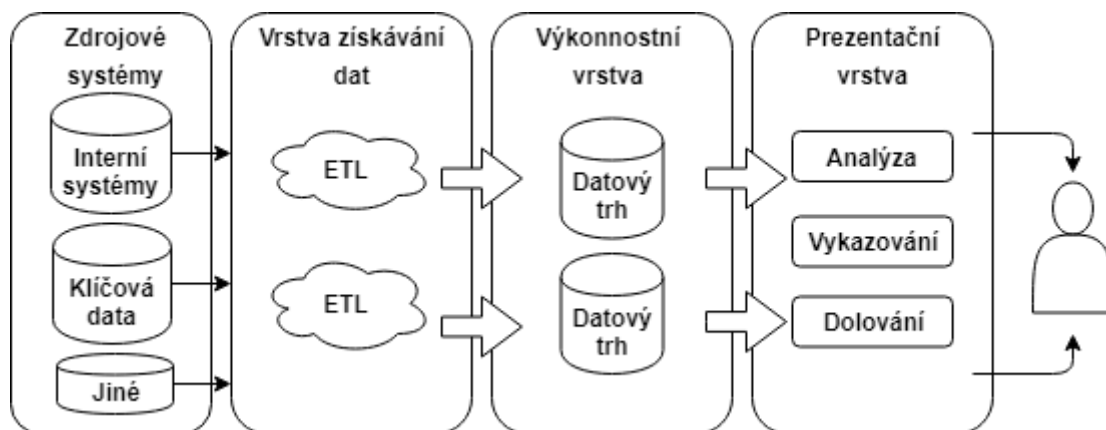
### 6.3.2 Architektura dle komponent

Nejobtížnější aspekt systému datového skladu nespočívá v jeho technických a technologických obtížích, ale ve volbě nejlepší metody jednak pro vybudování architektury datového skladu vhodné pro kulturu a strukturu společnosti a jednak pro zodpovězení organizačních otázek a otázek podnikové politiky, které budou vznikat během její implementace. [2]

Architektury datového skladu můžeme rozdělit na čtyři základní přístupy v rozsahu od oddělení až po sdílení dat na úrovni celé společnosti. Tyto přístupy budou blíže specifikovány v následujících podkapitolách.

#### 6.3.2.1 Architektura samostatných datových trhů

Samostatný datový trh nebo lépe řečeno nezávislé datové trhy (datových trhů totiž může existovat více, jen nejsou nijak propojeny) představují nejprostší formu datového skladu. [4] Datový trh je menší verze centrálního prostředí, která se řídí fyzickými nebo logickými centrálními datovými modely, ale je optimalizována s ohledem na určité koncové použití [27]. Na obrázku níže je vyobrazen typický tok dat architektury, který směřuje od zdrojů dat k samostatným datovým trhům.



Obr. 11 Architektura samostatných datových trhů

Zdroj: zpracováno dle [4]

Vybudování této architektury zpravidla začíná podnětem na úrovni oddělení podniku, jehož cílem je centralizovat data do jedné databáze. Po vytvoření datových trhů se následně nahrají data ze zdrojových systémů do centrální oblasti, přičemž

správa dat probíhá způsobem vyhovujícím danému oddělení. Často jsou však data načítána přímo do datových trhů a spravována až poté. [4]

V případě, že společnost začne generovat stále více datových trhů, práce s identifikací, správou a načítáním dat pak může představovat příliš velkou zátěž. Bude tedy nutné početné datové trhy optimalizovat (změnit jejich návrh) a na minimum omezit načítací rutiny, aby se tak opět optimalizovalo fungování společnosti. [2]

### 6.3.2.2 Architektura sběrnice

Architekturu sběrnice, kterou znázorňuje obrázek 12, vyvinul Ralph Kimball, jedna z nejdůležitějších osob pro oblast datových skladů a Business Intelligence, jelikož jako první charakterizoval využití dimenzionálního modelování a konceptualizoval datové trhy (včetně schémat hvězdy a sněhové vločky). [30]



Obr. 12 Architektura sběrnice

Zdroj: zpracováno dle [4]

Architektura sběrnice se orientuje na podnikovou analytiku a zaměřuje se na finální výsledky. Jedná se o modifikaci předešlé architektury samostatných datových trhů doplněnou o aspekt správy, který zaručí zachování struktur a definic tabulek v celé společnosti. [29] Tím se dosahuje potvrzené terminologie, názvů a struktur, čímž také dochází k opakovatelné použitelnosti veškerých dat v různých odděleních společnosti. Díky tomu je architektura sběrnice též označována jako „potvrzené dimenze“. [4]



Potvrzené dimenze jsou základní koncepční dimenze, které představují informační pilíře společnosti. Tyto dimenze je možné v rámci datových trhů využívat opakovaně, návrh architektury je tedy založen na synchronizaci dimenzí v celé organizaci. To přináší zjednodušení celého řešení, jelikož všichni při svých analýzách využívají stejné dimenze. Potvrzené dimenze se navíc popisují pomocí podnikových termínů, tudíž jsou relativně srozumitelné a snadno použitelné. Mezi typické potvrzené dimenze patří Zákazník, Obchod, Produkt, Zaměstnanec a další. [4]

Na rozdíl od architektury samostatných datových trhů vyžaduje tato architektura centrální správu dat, která zajistí, že budou dimenze skutečně potvrzené. Zpravidla je potřebná realizační vrstva, a to výhradně pro přípravu dat k jejich distribuci, neslouží tedy jako úložiště (viz obrázek 12). Datové trhy se navrhují zejména ve hvězdicových schématech, je však možné je normalizovat do návrhů ve formě sněhové vločky.

Architektura sběrnice se vyskytuje na výkonnostní vrstvě a zahrnuje potvrzené dimenze jako celopodnikový sdílený zdroj dat, díky kterému mezi sebou jednotlivé datové trhy komunikují a data sdílejí. Vlastní dimenze uchovávají veškerá data po delší dobu, zatímco faktické tabulky seskupují dimenze v daném analytickém kontextu. [29]

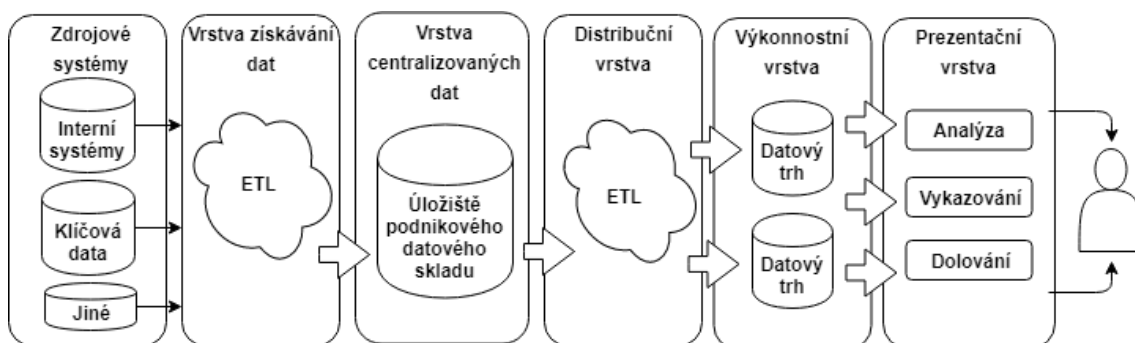
Bez sdílení dat a přizpůsobení se faktům a dimenzím se datový trh stává samostatnou aplikací. Izolované datové trhy, které nelze vzájemně propojit, však představují zhoubu pro další rozvoj celého datového skladu. Proto je integrace datových trhů pomocí sběrnice nezbytná.

Definováním standardního rozhraní sběrnice v prostředí datových skladů mohou být izolované datové trhy propojeny do různých seskupení v různém čase a koexistovat tak spolu prospěšným způsobem. [31]

### **6.3.2.3 Architektura centrálního úložiště**

Další velice významnou osobou v oblasti datových skladů je William H. Inmon. V roce 1991 jako první definoval pojem datový sklad. Z toho důvodu bývá považován za „otce datových skladů“ a bezesporu se jedná o významného průkopníka

a odborníka v této oblasti. [30] To potvrzuje i architektura centrálního úložiště znázorněná na obrázku 13, kterou Inmon vyvinul a která bývá hojně využívána.



**Obr. 13 Architektura centrálního úložiště**  
Zdroj: zpracováno dle [4]

Princip architektury centrálního úložiště spočívá ve vytvoření centrální oblasti, ve které je možné ukládat data v čase společně s jejich změnami, a to na úrovni největších podrobností [32]. Architektura tak zdůrazňuje aspekt úložiště, které ukládá atomární data společnosti normalizovaným způsobem, ale při zachování plné časové historie.

Z výše uvedeného plyne, že se návrh zaměřuje především na správu podnikových dat. Je však nezbytné vytvořit takový formát struktury dat, který zaručí její pružnost a také umožní růst v případě, že se do datového skladu budou přidávat nové hodnoty i jiné projekty. [4]

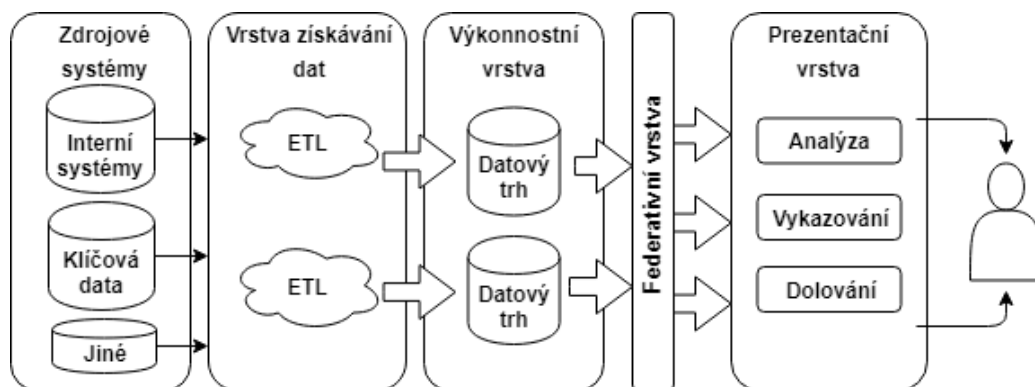
Celkový princip řešení je relativně prostý. Jádrem architektury je úložiště, které posiluje škálovatelnost a pružnost návrhu. Data jsou čerpána ze zdrojových systémů či z úložiště provozních dat, převedena do vyžadované podoby pomocí procesů ETL v realizačním prostředí a následně sloučena s úložištěm datového skladu, které je navrženo normalizovaným způsobem. Úložiště tak ze své podstaty představuje centrální zdroj potvrzených dat. Po nahrání dat do úložiště k nim poté může přistupovat jakýkoliv koncový uživatel pomocí datových trhů, které jsou určeny pro specifické použití. Slouží však zejména pro podporu rozhodovacích procesů jednotlivých oddělení společnosti. [32]

Datové trhy v tomto případě tvoří podmnožinu celého úložiště, ale nemusí být nezbytně navrženy stejným způsobem. Mohou být navrženy jak normalizovaným, tak i denormalizovaným způsobem. Záleží na tom, který ze způsobů je vhodnější pro podnikové použití. [27]

Vzhledem k tomu, že jsou datové trhy navrženy pro firemní použití, bývají optimalizovány pro konkrétní aplikaci. Mohou být ve hvězdicovém schématu, schématu sněhové vločky nebo ve třetí normální formě. [4]

### 6.3.2.4 Federativní architektura

Federativní architekturu znázorňuje obrázek 14. Hlavní důraz je kladen na přístup k datům (reporty, dolování dat, analýzy atd.), jejichž zdrojem jsou různé již existující datové trhy či přímo datové sklady. Tato architektura tedy povstává z bezprostředního použití samostatných systémů, zpravidla v případech, kdy je nutné integrovat dvě oddělení nebo pokud dochází ke slučování společností. [4] Provozování několika nezávislých datových skladů může být finančně náročné, proto se často tento návrh slučuje do architektury s jedním centrálním úložištěm.



**Obr. 14 Federativní architektura**  
Zdroj: zpracováno dle [4]

Federativní datový sklad čerpá data z existujících datových trhů nebo datových skladů pomocí procesů ETL a nahrává je do centrálního dimenzionálního úložiště. Z důvodu rozdílných časů nahrávání dat do zdrojových datových skladů je však důležité správné nastavení periody spouštění ETL procesů. Současně je potřeba zjistit, zdali data načítaná ze zdrojových datových skladů nejsou vzájemně duplicitní a pokud ano, tak je sloučit. [33]

Sjednocení několika datových trhů je obvykle založeno na federativních pohledech, tudíž je možné datové trhy využívat při minimálních nárocích na změny současných systémů. Udržování federativní architektury však může komplikovat správu dat a rovněž zvyšovat náklady. [4]

## **6.4 Metody budování datového skladu**

Jedním z nejpodstatnějších kroků při budování datového skladu je volba nejvhodnější metody. Musíme brát v potaz nejen organizační strukturu, ale předvídat i možné potíže, které se během implementace datového skladu mohou objevit. [4] Nejznámější a nejčastěji používané metody budou přiblíženy v následujících podkapitolách.

### **6.4.1 Metoda velkého třesku**

Mnoho společností, vývojářů a nejspíše i konzultantů se podle [2] domnívá, že je možné implementovat datový sklad prostřednictvím jediného projektu. Vývoj datového skladu je však náročná záležitost, proto se ji patrně nepodaří vyřešit naráz a rozhodně ne v rozumném čase. To je největší slabina tohoto přístupu, protože i kdybychom nakonec projekt datového skladu zrealizovali metodou velkého třesku, mohou se mezitím změnit nejen požadavky koncových uživatelů, ale i technologie používané v rámci vývoje [27]. Metoda velkého třesku se skládá ze tří po sobě jdoucích kroků, které jsou následující [2]:

- analýza požadavků společnosti,
- vybudování datového skladu,
- vytvoření přímého přístupu nebo přístupu pomocí datových trhů.

Jediným přínosem metody velkého třesku je to, že celý plán projektu můžeme kompletně vytvořit ještě před zahájením jeho realizace [27]. Převládají tedy spíše nevýhody tohoto přístupu, které jsou však velmi závažné. Je tu nejen velké riziko změny požadavků, ale hlavně trvá dlouhou dobu, než se prokážou první výsledky nemalých investic do datového skladu a dostaví se kýžený obchodní zisk. [2]

## 6.4.2 Přírůstková metoda

Přírůstková metoda předpokládá realizaci datového skladu po jednotlivých fázích. Namísto vytvoření celého datového skladu najednou tak postupně přibývají přírůstková řešení, která zapadají do výsledné architektury datového skladu. Fáze přírůstkové metody jsou [27]:

- strategie společnosti,
- definice rozsahu přírůstkového vývoje,
- analýza uživatelských potřeb a požadavků,
- návrh řešení,
- implementace a otestování navržené struktury,
- produkce a údržba datového skladu.

Začneme realizací několika málo oblastí (jednou či dvěma), toto částečné řešení implementujeme (např. jako rozšiřitelný datový trh) a zpřístupníme ho koncovým uživatelům. První subsystémy celého řešení tak začnou fungovat a přinášet výhody již po krátké době od spuštění projektu. [27]

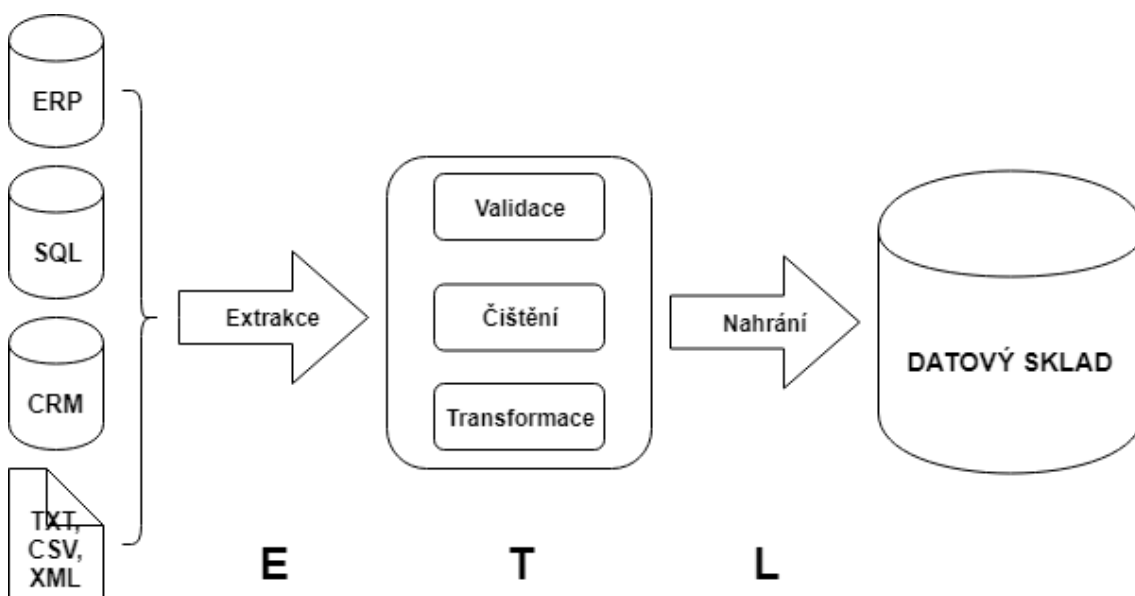
V případě, že se částečné řešení důkladně otestované koncovými uživateli osvědčí, můžeme do systému zakomponovat novou předmětnou oblast. Tím získáme další funkcionalitu. [2] Stejným způsobem se pokračuje až do kompletního vyřešení datového skladu.

Budování datového skladu pomocí přírůstkové metody tedy představuje iterativní proces udržující stálou vazbu mezi budovaným datovým skladem a požadavky koncových uživatelů [27]. Na rozdíl od metody velkého třesku zde převažují výhody nad nevýhodami. Mezi hlavní výhody přírůstkové metody můžeme zařadit [2]:

- Iterativní budování datového skladu zachovává spojitost projektu s potřebami a požadavky uživatelů.
- Umožňuje implementovat škálovatelnou architekturu.
- Zabezpečí rychlejší zisk, a tím pádem i rychlejší návratnost investice.

## 7 ETL

Předtím, než je možné využívat datový sklad, je potřeba jej naplnit daty, přesněji řečeno vytvořit, nastavit a zajistit proces jeho plnění. A přesně to je úlohou vrstvy ETL v rámci projektu Business Intelligence. [2] ETL představuje posloupnost procesů, jejichž úkolem je extrakce dat ze zdrojových systémů, jejich transformace a očištění a následně ve vhodné formě nahrání do cílového úložiště (zejména datového skladu). [28] Na vrstvu ETL tak můžeme pohlížet jako na článek mezi zdrojovými systémy a datovým skladem, který tyto dva segmenty propojuje (viz obr. 15).



Obr. 15 Architektura ETL procesu  
Zdroj: vlastní zpracování

Obecně lze říci, že data, která chceme, aby vstupovala do procesu BI, mnohdy pocházejí z nehomogenních a různorodých zdrojů. Mohou to být například data ze souborových databází (MS Access), data z databází spravovaných databázovým serverem (MS SQL Server, Oracle, Informix), či data vyexportovaná nějakým informačním systémem (ERP, CRM, SCM) do tzv. flat file, dokumentu XML apod. [27]

Data se ve fázi ETL nejen nahrávají, ale různým způsobem i zpracovávají, například sumarizují, indexují, zjišťují se potenciální změny struktury zdrojových dat potřebných pro datový sklad, udržují se metadata (předpisy a definice pro přenos a správu dat) a podle potřeby se mění rovněž struktury klíčů. [27]

Prvotním naplněním datového skladu daty ze zdrojových systémů úloha ETL pochopitelně nekončí, datový sklad se totiž v pravidelných časových intervalech (např. den, týden, měsíc) naplňuje novými či aktualizovanými daty. Proces je tedy komplexní a v drtivé většině případů časově velmi náročný. U implementací některých řešení si může vyžádat i více než polovinu z celkového času, úsilí a tedy i značnou část nákladů potřebných na vytvoření datového skladu.

Je rovněž velice důležité si uvědomit, že ve fázi ETL se pracuje s daty, ze kterých se později stanou informace. Z toho důvodu by tato data měla být velmi přesná, kvalitní, předmětná, aktuální, a tedy užitečná pro koncového uživatele. Kromě kvality dat je také důležitá i jejich dostupnost. To proto, aby mohli jednotliví uživatelé datového skladu používat datový sklad efektivně a účinně. [2]

Hlavní cíle a požadavky kladené na fázi ETL při vývoji BI řešení jsou [28]:

- centralizace dat z mnoha různorodých a nehomogenních zdrojových systémů,
- transformace dat do vhodného formátu,
- očištění zdrojových dat od chybných a duplicitních záznamů,
- dostatečný výkon pro práci s velkým objemem dat,
- nezávislost na použité technologické platformě,
- naplnění datového skladu přesnými a kvalitními daty,
- schopnost plnění datového skladu novými či aktualizovanými daty.

Jednotlivými fázemi procesu ETL se budeme podrobněji zabývat níže.

## **7.1 Extrakce dat**

První fází celého procesu ETL je extrakce dat. Jak již bylo řečeno, data, která chceme nahrát do datového skladu, jsou umístěna v různorodých nehomogenních prostředích, hardwarových platformách, operačních systémech, archivních systémech, podnikových systémech, databázových systémech a dalších různých dokumentech [27]. Tímto výčtem však výsledný počet kombinací zdrojů pro proces integrace zdaleka nekončí, ten je totiž dále navyšován díky rozličným formátům dat, kte-

ré se ve výše zmíněných zdrojích používají k ukládání například kalendářních dat nebo čísel s plovoucí řádovou čárkou. Úlohou extrakce dat je v rámci ETL procesu získat data právě z takových zdrojů. [2]

Dalším faktorem, který zvyšuje počet kombinací, je vzájemné propojení podnikových procesů. Kromě interních dat z podnikového prostředí je totiž občas potřeba pracovat i s daty externími. [27] Může se jednat o data od obchodních partnerů, analýzu konkurenčního prostředí či zakoupená data o zákaznících. Nezanedbatelným zdrojem informací jsou v dnešní době i data přístupná z prostředí internetu jako například historie měnových kurzů, cen akcií či data ze sociálních sítí. U externích dat však nemůžeme odebírat vzorky periodicky, jako je tomu v případě interních dat. Externí data tedy vyžadují neustálé monitorování za účelem stanovení doby dostupnosti. [2]

K provedení extrakce dat můžeme použít různé postupy, nástroje a technologie. Například můžeme vytvořit vlastní aplikaci v nějakém vyšším programovacím jazyce (C#) nebo v procedurální nadstavbě dotazovacího jazyka SQL (PL/SQL, T-SQL). Někdy můžeme využít i výstupy z interních podnikových systémů, které poskytují konverzi a očištění dat. [9] U správně navrženého procesu ETL máme rovněž k dispozici metadata pro všechny fáze tohoto procesu, která obsahují informace o místě, typu, struktuře a přístupových právech datového zdroje.

Z výše uvedeného vyplývá, že ve fázi extrakce dat musí být ETL schopno získat všechna data nezávisle na platformě, v níž jsou zdrojové systémy implementovány. Extrakce dat by rovněž neměla ovlivňovat chod zdrojových systémů, ze kterých jsou data čerpána. Z toho důvodu se pro provedení extrakce dat obvykle volí čas od večerních do brzkých ranních hodin, tedy doba, kdy nejsou zdrojové systémy využívány pro jejich primární účely. [2]



## 7.2 Transformace a čištění dat

Zřejmě nejvýznamnější a mnohdy implementačně nejnáročnější fází procesu ETL je fáze týkající se transformace a čištění dat. Během této fáze se data převádí z podoby, v níž byla uložena ve zdrojových systémech, do podoby vhodné pro jejich uložení v datovém skladu [28]. Celý proces probíhá na základě předem specifikovaných transformačních pravidel. [27]

Data ze zdrojových systémů mají určitou kvalitu, která často bývá nejen proměnlivá, ale dokonce nedostačující pro zavedení dat do datového skladu. Špatná kvalita dat může být způsobena zejména nepozorností při manuálním zadávání údajů (například nevyplnění všech polí formuláře, gramatické chyby, překlepy, nedodržení formátu dat). Z toho důvodu je potřeba data nejprve očistit. Předmětem čištění mohou být de facto všechny atributy datových objektů jako identifikační údaje, adresy či kontakty. Čištění dat však může být někdy velmi náročné, a tedy i nákladné. [2]

Transformace dat pak představuje sadu úkonů a úloh, které vedou ke zvýšení kvality dat získaných ze zdrojových systémů [2]. Rovněž mají za úkol odstranit anomálie, které ve zdrojových systémech vznikají například přechodem na jiný operační systém nebo změnou kódové stránky, měny, formátu času apod. [27]

Pod transformací dat si můžeme představit celou škálu operací od různých konverzí dat, filtrování dat, matematických operací, až po sofistikované způsoby vytváření multidimenzionálních struktur. Ty nejpoužívanější z nich shrnuje následující výčet [9]:

- filtrace – výběr určité množiny řádků podle filtrační podmínky,
- třídění – vzestupné nebo sestupné setřídění dat dle zvoleného sloupce,
- selekce – výběr pouze potřebných sloupců,
- konverze – transformace hodnoty sloupce na požadovaný formát například pomocí matematické, textové nebo časové funkce,

- agregace – zpracování hodnot do skupin použitím agregační funkce, např. AVG, MIN či MAX (z mnoha hodnot je tak vytvořena pouze jediná hodnota za celou skupinu),
- sjednocení – sjednocení řádků pocházejících z různých datových zdrojů,
- spojení – spojení sloupců dvou a více záznamů do jednoho celistvého záznamu pomocí klíčového atributu, který spojení umožní,
- pivotace – záměna sloupců za řádky a naopak.

Je rovněž důležité upozornit na skutečnost, že se v rámci fáze transformace dat můžeme potýkat s různými problémy, se kterými je třeba se vypořádat. Mezi ty nejčastěji se vyskytující řadíme [2]:

- Nejednoznačnost dat – Nejednoznačnost dat se projevuje zejména při uložení hodnot stejného typu, ale v různé podobě. Například údaj o pohlaví zákazníka může být uložen různým způsobem (muž, Muž, M). Tyto odlišnosti je tedy potřeba odstranit transformací na jednotný tvar.
- Chybějící hodnoty a duplicitní záznamy – Snížení kvality dat mohou způsobovat i chybějící hodnoty (sloupce obsahující hodnotu NULL), případně skrytá nebo otevřená duplicita záznamů. Duplicita údajů je menší problém, jelikož pokud je něco navíc, vždy se to dá odstranit. V některých případech však odhalení a odstranění duplicit může být časově náročné. Daleko větší problém představují chybějící data. U malého objemu chybějících dat je však můžeme ignorovat nebo doplnit manuálně z jiných zdrojů.
- Konvence názvů pojmů a objektů – Slučujeme-li data z různých zdrojů, které popisují stejný jev, ale mají jednotlivé objekty vedeny pod odlišnými názvy, musíme sloučit terminologii a vytvořit jednotnou konvenci názvů.
- Referenční integrita – Kromě hodnot jsou v datech ukryty i různé vztahy, například hierarchická struktura zaměstnanců, organizační struktura podniku a podobně. Data jsou však dynamická, organizační struktura se v čase mění, a to často bez dokumentace změn ve zdrojových systémech.

Pokud se tedy zruší jedno z oddělení a zůstanou po něm nějaké záznamy, mohou tyto záznamy zkreslit data a tím nepříznivě ovlivnit jejich kvalitu.

- Chybějící datum – Čas plní v datovém skladu velmi důležitou úlohu, jelikož se od něj vše odvíjí a téměř každá analytická databáze obsahuje časovou dimenzi. Časový údaj tak musí buď být přítomný v datech před jejich nahráním do datového skladu, nebo se musí zjistit a přidat při nahrávání dat.

### **7.3 Nahrání dat**

Završením celého procesu ETL je fáze nahrání dat, ve které se transformovaná a očištěná data ukládají do cílové databáze nebo datového skladu [27]. Nahrávání dat spočívá v přesunu dat a jejich fyzickém uložení do výsledných databázových tabulek. Mělo by být plánované, automatizované a co nejvíce optimalizované. [2]

Při počátečním naplnění datového skladu se může jednat o obrovská množství dat. Poté se však nová data nahrávají v pravidelných časových intervalech (například každý měsíc). Jejich objem přitom závisí na tom, jaké množství dat za dané období ve zdrojových systémech vznikne. [28]

Pro naplnění datového skladu můžeme použít hned několik metod. Volbou metody plnění ovlivňujeme nejen dostupnost, ale i dobu plnění datového skladu. Rovněž je vhodné při výběru dané metody vzít v úvahu, na jaké platformě je datový sklad postaven. Pro plnění datového skladu existují následující metody [9]:

- Metoda přímého nahrávání dat – po dobu zavádění dat musí být datový sklad odstaven, jelikož je implementovaný pouze na jednom počítači. V závislosti na technologické platformě, v jaké je datový sklad vybudován, je možno využívat jejích vlastností. Některé z platforem umožňují data do datového skladu nahrávat přímo za běhu. Podmínkou však je, že jsou po dobu plnění data v datovém skladu dostupná pouze ke čtení. Výhodou tohoto řešení je jeho nízká cena. Naopak nevýhodou lze spatřovat

ve výskytu technologických odstávek. Při zavádění dat do větších datových skladů rovněž v tom, že čas odstávky nemusí být dostačující.

- Metoda postupného nahrávání dat – změny prováděné v podnikových informačních systémech jsou zpracovávány v menších objemech dat. ETL proces je tak spouštěn v kratších časových intervalech, přičemž provedené změny jsou označeny příznakem neviditelnosti. Důvodem používání tohoto příznaku je konzistence obchodních záznamů. Po zpracování většího množství dat je příznak neviditelnosti změněn.
- Metoda flip-flop – tato metoda vyžaduje použití dvou shodných serverů. Jeden z nich je provozní, zatímco druhý slouží k nahrávání dat. Provozní server je tak dostupný i po dobu nahrávání dat, jelikož ta se nahrávají na druhý server. Po ukončení nahrávání dat se role obou serverů prohodí. Tím se snižuje nedostupnost datového skladu pouze na dobu, kdy dochází ke změně rolí serverů.
- Metoda odděleného ETL – fáze extrakce a transformace dat jsou vykonávány na zařízení, které je odděleno od zařízení, na němž je implementován datový sklad. Po dokončení obou fází jsou data nejprve ověřena, a až poté přesunuta do datového skladu. To umožňuje zkrátit potřebnou dobu odstávky datového skladu.

Po nahrání dat dále probíhá jejich indexace. To z důvodu, aby byl přístup k těmto datům optimalizovaný pro následné dotazování. [27] Pro jednoznačnou identifikaci dat se rovněž využívají i uměle generované klíče. Pomocí nich jsme schopni zajistit jednoznačnost každého řádku v databázové tabulce. Data v datovém skladu jsou totiž mnohdy kombinací několika transformovaných záznamů, které nejsou opatřeny žádným přirozeným klíčem, který by se dal využít k jednoznačné identifikaci daného záznamu. [2]

## 8 OLAP

Zkratka OLAP (Online Analytical Processing) představuje databázovou technologii, která umožňuje efektivně zpracovávat velké objemy dat pro účely tvorby dotazů, sestav a analýz podnikových dat uložených v datovém skladu [36]. Jinými slovy slouží k taktickému a strategickému rozhodování společnosti, jelikož umožňuje vykonávat komplexní výpočty pro potřeby různých analýz v dimenzionálních systémech, které by nebylo snadné napodobit pomocí čistě relačního řešení [33]. Pro analýzu dat se tak na rozdíl od relačních databází, ve kterých jsou data uložena v jednotlivých řádcích a sloupcích, používají analytické databáze s multidimenzionálními strukturami (viz kapitola 5).

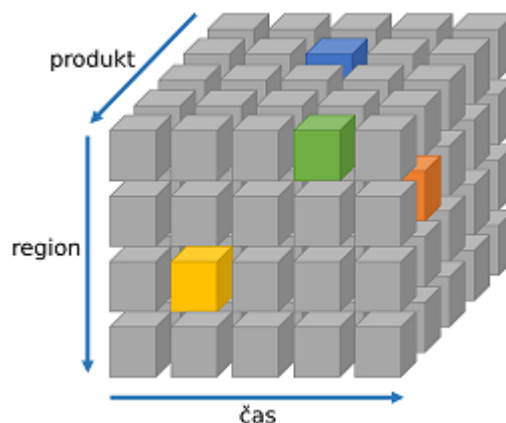
Do OLAP vstupují data, která jsou odvozována z historizovaných dat v datovém skladu a slučována do datových struktur umožňujících jejich důkladnou analýzu. [35] Data jsou rovněž organizována hierarchicky a namísto tabulek ukládána do tzv. OLAP kostek. Ty dovolují pružně a hlavně rychle měnit jednotlivé dimenze, čímž umožňují měnit úhel pohledu uživatelů na sledovanou modelovanou realitu.

Jedná se tedy o důmyslnou technologii využívající vícedimenzionálních struktur, která při provádění analýz poskytuje rychlý přístup k datům. [2]

### 8.1 OLAP kostka

Jak již bylo zmíněno výše, v souvislosti s technologií OLAP často hovoříme o OLAP kostkách. Ty se skládají z dat, které pocházejí z jedné či více tabulek faktů a rovněž z údajů, které jsou uloženy v tabulkách představujících dimenze. Pojem kostka přitom vychází z logiky, že datová struktura může obsahovat více dimenzí, než pouhé dvě jako je tomu v případě tabulek relačních databází. [27]

V reálných multidimenzionálních databázích je počet použitých dimenzí zpravidla vyšší (i desítky dimenzí). OLAP kostka tak dovoluje, aby data byla modelována a prohlížena z několika úhlů pohledu (dimenzí), přičemž hledané informace se pak nacházejí na průsečících použitých dimenzí. [2] Pohled na OLAP kostku nabízí obrázek 16.



**Obr. 16 OLAP kostka**  
Zdroj: upraveno dle [2]

OLAP kostky běžně obsahují značný počet předzpracovaných agregací, aby se minimalizovala doba potřebná k následným výpočtům v rámci dotazování. OLAP databáze tak díky předkalkulovaným agregacím dosahují vynikajícího výpočetního výkonu. [37]

Je však důležité zmínit, že není ve všech případech nutné, aby byla OLAP kostka součástí datového skladu. Její zařazení závisí na účelu a zvolené architektuře daného řešení. Existují tedy datové sklady, které OLAP databáze neobsahují. Vyznačují se pak tím, že jsou levnější a méně náročné nejen co se realizace, ale i údržby týče. Není totiž nutné vytvářet vlastní OLAP kostku, ETL rutiny plnění apod. [37]

Mezi velké nevýhody OLAP kostek můžeme zařadit větší nároky na paměťový prostor, jelikož OLAP kostka obvykle obsahuje maximální počet předzpracovaných agregací dat. Tyto výpočty rovněž zatěžují i samotný server datového skladu. [2]

V uložení velkého množství různých kalkulací však tkví i hlavní výhoda, jelikož díky tomu OLAP kostka umožňuje dosahovat velmi vysokého dotazovacího výkonu, a to i při velkých objemech dat. [2]

## Operace s OLAP kostkou

Důvodem pro vytvoření OLAP kostky je snadná možnost realizovat přijatelným způsobem analýzy nad shromážděnými daty v multidimenzionálním modelu. Mezi základní operace spojené s OLAP kostkou řadíme následující [35, 38]:

- Drill-down – operace, která umožňuje zaměřit se na detailnější hodnoty, tzn. sestoupit na nižší agregační úroveň (například v produktové dimenzi z kategorie na jednotlivé podkategorie). Po provedení operace Drill-down je do OLAP kostky přidána jedna či více dimenzí.
- Roll-up – operace Roll-up je přesným opakem operace Drill-down. Ve zvolených dimenzích tedy nastavuje vyšší úroveň agregace, tzn. zobecňuje data (z předchozího příkladu „sbalí“ podkategorie zpět do kategorií). Po provedení této operace je tak z OLAP kostky jedna či více dimenzí odstraněno.
- Pivoting – Pivoting je operace, která umožňuje s OLAP kostkou rotovat a měnit tak úhel pohledu na agregovaná data dle dimenzí.
- Slicing – Slice představuje operaci výběru jedné určité dimenze z celé OLAP kostky, z níž je vytvořena kostka nová (sub-kostka). Tato operace současně snižuje dimenzionalitu kostek.
- Dicing – Operace Dice je velmi podobná operaci Slice, s rozdílem, že je v tomto případě vybráno dvě a více dimenzí. Po provedení operace zde rovněž vzniká nová kostka a dochází ke snížení počtu hodnot v dimenzích.

## 8.2 Varianty OLAP

Se vzrůstajícím množstvím rozměrů v multidimenzionálních databázích zároveň velmi rychle dochází i k nárůstu požadavků na velikost úložného prostoru. Z tohoto důvodu se v praxi používají různé technologie pro ukládání, kompresi a zpracování dat potřebných k provedení analýz. [37] Mezi nejčastěji používané technologie řadíme MOLAP (Multidimensional OLAP), ROLAP (Relational OLAP), HOLAP (Hybrid OLAP) a DOLAP (Desktop OLAP).

### **8.2.1 MOLAP**

Technologie MOLAP získává data z datového skladu či zdrojových systémů, která následně ukládá do multidimenzionálních datových struktur, tedy OLAP kostek. Během procesu ukládání dat je předzpracováno tolik výpočtů, kolik jich je z časového a technického hlediska možné. Databáze je tedy organizována pro účel rychlého získávání příslušných hodnot z více dimenzí. Hlavní výhodou této technologie spočívá ve vysokém dotazovacím výkonu. Nevýhodou je pak redundance dat z důvodu jejich uložení v multidimenzionální i relační databázi či omezené množství dat, které je možné tímto způsobem zpracovat. Díky tomu se MOLAP nehodí pro práci s většími objemy dat. [2, 38]

### **8.2.2 ROLAP**

Relační OLAP čerpá data a metadata přímo z relačních databází datového skladu, datových trhů či z relačních databází zdrojových systémů. Tato data jsou po zpracování předložena uživateli v podobě multidimenzionálního pohledu na příslušná data. Metadata jsou v rámci ROLAP ukládána jako záznamy v relačních databázích, přičemž jsou dynamicky používána pro generování SQL příkazů, které jsou potřebné pro získávání dat požadovaných uživatelem. Oproti způsobu ukládání dat v případě MOLAP, zůstávají data v ROLAP pouze v relačních databázích, čímž je odstraněn problém s redundancí dat. Zpracování dotazů je však obvykle pomalejší, než při použití technologie MOLAP. Na druhou stranu ROLAP dovoluje uživatelům přímo přistupovat k aktuálním a historickým datům ze zdrojové databáze. Navíc zde neexistují žádné limity týkající se velikosti použité databáze. [2, 38]

### **8.2.3 HOLAP**

Hybridní OLAP je de facto kombinací obou předešlých přístupů, přičemž je zde snaha o maximální využití výhod a potlačení nevýhod každého z nich. Data pro podrobné analýzy (pracující s velkými objemy dat) zůstávají uložena v relačních databázích, avšak agregovaná data, jejichž získání je časově mnohem náročnější, jsou ukládána v multidimenzionální databázi. Díky tomu HOLAP poskytuje propojení mezi značnými objemy dat v relačních tabulkách a současně nabízí výhodu vyššího dotazovacího výkonu n-dimenzionálně uložených agregací. [27, 38]



#### **8.2.4 DOLAP**

Desktopová OLAP technologie umožňuje koncovým uživatelům připojit se k centrálnímu úložišti OLAP dat a stáhnout si OLAP kostku či její potřebnou podmnožinu na lokální disk. Veškeré analytické operace v rámci prováděných analýz jsou pak vykonávány nad touto lokální kostkou, a to bez nutnosti připojení k OLAP serveru, tedy offline. DOLAP tak nachází uplatnění zejména u mobilních aplikací jako podpora mobilních uživatelů. [10, 38]

## 9 Praktická část

Praktická část práce se zabývá návrhem a pilotní implementací BI/DWH řešení pro oblast školství v Hradci Králové jako obce s rozšířenou působností. Celý projekt byl přitom realizován pod záštitou a odborným dohledem společnosti GIST s.r.o.

Společnost GIST s.r.o. je česká softwarová a poradenská společnost se sídlem v Hradci Králové, která se orientuje nejen na vývoj softwaru, ale rovněž na Business Intelligence a Consulting. V synergii zmíněných oblastí přitom spatřuje svou konkurenční výhodu, kterou ocenilo již více, než 400 zákazníků ze všech segmentů obchodu, průmyslu, služeb a veřejné správy.

### 9.1 Představení oblasti školství

Ministerstvo školství, mládeže a tělovýchovy (MŠMT) zajišťuje prostřednictvím Odboru školské statistiky, analýz a informační strategie sběry a zpracování statistických údajů, jejichž dílčí součástí tvoří data z oblasti regionálního školství.

Krajské a obecní úřady mají tedy k dispozici řadu výkonových, ekonomických a personálních údajů a na ně navazující normativní způsob rozdělování prostředků. Statutární město Hradec Králové pak jako zřizovatel zodpovídá společně se školami a školskými zařízeními za zpracování těchto statistických údajů.

Ministerstvo poskytuje na vyžádání také specifické údaje (tzv. přehledky), z nichž lze sestavit informace například o spádovosti žáků. To znamená vytvořit informace o tom, ze kterých okolních obcí žáci danou školu a dokonce ročník školy navštěvují. Takový výstup pak lze využít v rámci zkvalitňování dopravní obslužnosti v širším regionu.

Je třeba zmínit stávající systém normativního přidělování prostředků, které byly dosud poskytovány prostřednictvím výkonů (počtu dětí, žáků apod.). Změna ve financování by měla začít od září 2018, kdy budou školy prostředky dostávat podle počtu odučených hodin.

Od reformy financování se očekává předvídatelné stabilní financování. Skončí tak udržování žáků na školách za každou cenu, přestože nedosahují dostatečných výsledků na to, aby se ve škole udrželi.

## **9.2 Současný stav**

V současnosti je z oblasti technologií Business Intelligence a datových skladů na Královéhradeckém kraji využíván softwarový systém GIST Controlling, na kterém je postaveno stávající BI řešení vytvořené na základě požadavků zaměstnanců odboru školství.

GIST Controlling je informační systém, který je úspěšně provozován u mnoha desítek zákazníků v České a Slovenské republice, a to téměř ve všech podnikatelských segmentech i veřejné správě. Tento systém plní úlohu široce pojatého nadstavbového manažerského a controllingového systému, který zahrnuje podporu plánování, reportingový systém a manažerský portál. V rámci své činnosti GIST Controlling přebírá základní hodnotová data z jiných systémů a dále je zpracovává do tvaru souhrnných přehledů a výkazů včetně grafického znázornění. Umožňuje sledování vývojových trendů jednotlivých položek a poskytuje možnosti porovnání plánových hodnot s hodnotami skutečnosti. Umožňuje i vkládat data ručně přímo do systému, což je zejména využíváno při sestavování plánu a při modelování navrhovaného opatření. Systém je postaven na využití technologií společnosti Microsoft, které společnost GIST vlastním vývojem doplňuje o zatím jinak nedostupné, ale potřebné funkcionality.

V současnosti tedy pracovníci odboru školství používají systém kategorie Business Intelligence, který je určen k integraci dat z různorodých datových zdrojů z oblasti školství. Jedná se zejména o integraci dat ze školských výkazů, údajů o kapacitách školských zařízení, o kapacitách jednotlivých studijních oborů a údajů z ekonomických výkazů. Nad uloženými daty z analytické a relační vrstvy je přitom zpracováno cca 200 výstupů využívající software Microsoft Excel v podobě tabulek, kontingenčních tabulek a grafů.

Kromě výše zmíněného systému však Královéhradecký kraj v současnosti nedisponuje jinými nástroji Business Intelligence ani technologiemi datových skladů. Není tak schopen efektivním způsobem získávat data, a to jak z interních systémů, tak z externích subjektů, jako jsou například Český statistický úřad, Ministerstvo Vnitra atd. V řešení, které nevyužívá datový sklad, se ukrývá řada slabých míst, mezi která můžeme zařadit následující:

- Nedostatečná systémová výtěžnost obrovského rozsahu informací soustředěných v resortních výkazech škol a zařízení, statistickém výkazu, případně v účetních výkazech.
- Nestabilita používaných výkazů a nutnost ručního dopracování dat.
- Ruční úpravy datových zdrojů nesoucí vysoké riziko chybovosti.
- Nepřehlednost výstupů při zaměření se na velký detail.
- Absence systému rychlého dotazu a rychlé odpovědi.
- Omezená možnost další práce s předávanými podklady (např. tištěné výstupy).
- Problematická vypovídací schopnost rozsáhlého reportu pro manažerské řízení a rozhodování (absence dynamického reportingu).

Pro odstranění výše zmíněných slabín a problémů je potřeba zrealizovat projekt zahrnující transformaci stávajícího BI řešení na nově vzniklé BI/DWH řešení, které bude obsahovat centralizovaný datový sklad a rozšíření o nové funkcionality.

### **9.3 Požadavky na nové řešení**

Nedílnou součástí každého BI/DWH projektu je identifikace a analýza požadavků zákazníka, podle kterých je nutné se po celý čas návrhu řešení i jeho následné implementaci řídit. Pravděpodobně nejběžnějším způsobem získání požadavků je interview se zadavatelem. V tomto případě se však jedná o projekt realizovaný pouze pro účely diplomové práce. Z toho důvodu byl použit soupis stanovených požadavků zadavatele na stávající řešení rozšířený o požadavky vedoucích k odstranění výše zmíněných slabín. Pro doplnění těchto požadavků byly uskutečněny

konzultace se znalcem problematiky školství ze společnosti GIST, jelikož byl u zrodu současného řešení.

Z poskytnutého soupisu a uskutečněných konzultací byly identifikovány následující požadavky:

- Obecné požadavky na implementaci:
  - Plná lokalizace v českém jazyce.
  - Implementace nesmí nijak omezit a především ohrozit provoz zdrojových systémů a databází.
  - Po čas implementace bude vykonáváno funkční testování jednotlivých částí řešení.
  - Součástí implementace budou veškeré práce nezbytné pro řádné a úplné zprovoznění systému včetně implementačních popisů obsahujících jednotlivé kroky implementace všech částí.
- Požadavky na technický popis řešení:
  - Detailní seznam zdrojových dat s uvedeným popisem jejich zdroje, struktury a metodik pro aktualizaci.
  - Popis realizace datových pump s popisem jednotlivých kroků. U datové pumpy musí být základní popis funkce, kam a co importuje.
  - Popis realizace datové kostky s uvedením zdroje dat, tabulky a výstupů, které jsou nad kostkou vytvořeny.
  - Popis možností reportingu a dalších možných výstupů z datového skladu.
- Požadavky na nové funkcionality:
  - Doplnění nahrávaných školních výkazů počínaje rokem 2012 do datového skladu.
  - Vytvoření základního zdroje informací pro rozhodování o optimalizaci sítě školních zařízení jako součást datového skladu.
  - Doplnění specifických výstupů pro gestora oblasti školství a školský výbor.

## **9.4 Návrh řešení**

Krajské a obecní úřady mají k dispozici řadu výkonových, ekonomických a personálních údajů a na ně navazující normativní způsob rozdělování prostředků. Prostřednictvím implementace nástrojů datových skladů a nad nimi vystavených analytických a prezentačních nástrojů pak mohou budovat svůj informační potenciál zpřístupňující zainteresovaným subjektům relevantní informace o školství v daném území.

Cílem této podkapitoly je tedy navrhnout BI/DWH řešení jednotného prostředí spojujícího analýzu dat takovým způsobem, aby manažeři veřejné správy a další cílové skupiny uživatelů mohli využívat k rozhodování aktuální a ověřené informace zpracované transparentními postupy, které umožňují hodnocení efektivity úrovně řízení a výkonu veřejné správy.

### **9.4.1 Analýza datových zdrojů**

Datovými zdroji jsou zpravidla statistika školských výkonů, výkaz o pracovnících a mzdách (PAM), školské rejstříky a účetní výkazy. Datový sklad tak může nabízet v analytické a prezentační vrstvě nejen statický pohled na uspořádaná data, ale poskytovat také průřezové informace propojující výkonnost, mzdy a ekonomiku. Tyto informace mohou být také využívány pro benchmarking srovnatelných skupin škol a školských zařízení.

Pro účely diplomové práce však postačí sledovat pouze některé ze všech možných datových zdrojů, aby implementované řešení nebylo až příliš robustní. Jako vhodná volba se zdá být zúžení zkoumané oblasti z krajské úrovně na úroveň obce s rozšířenou působností, tedy města Hradce Králové. Toto zúžení ovlivní počet zkoumaných výkazů a s nimi spojených číselníků, jelikož na obecní úrovni je možno zkoumat především mateřské, základní a základní umělecké školy. Rovněž není možné sledovat ekonomické a účetní výkazy, neboť k tomu nebyl poskytnut souhlas krajského úřadu.

#### 9.4.1.1 Číselníky ve školství

Základním nástrojem standardizace školství je používání číselníků a klasifikací. Jedná se o datový zdroj (formátu XML), který bude sloužit k naplnění dimenzí datového skladu. Sledované číselníky shrnuje následující tabulka 1.

Tabulka 1 Číselníky ve školství

ID	Název číselníku
AKDT	Druhy a typy škol a školských zařízení
AKSO	Obory vzdělávání regionálního školství
BAZS	Zřizovatel školy
NAJS	Vyučovací jazyk školy
NBUO	Umělecké obory na základní umělecké škole
RADS	Délka vzdělávacího programu
RAFS	Forma vzdělávání
RAOR	Kód okresu trvalého pobytu žáka, případně kód státu u žáků s trvalým pobytem mimo ČR
RARO	Číselník ročníků
RASD	Způsob plnění povinné školní docházky na základní škole
RASS	Stupeň základní školy
RAST	Číselník států

Zdroj: vlastní zpracování

#### 9.4.1.2 Výkazy škol a školských zařízení

Statutární město Hradec Králové jako obec s rozšířenou působností zodpovídá společně se školami a školskými zařízeními za zpracování školských výkazů (ukládáných ve formátu XML). Protože data jsou sbírána každoročně, lze tyto informace vyhodnocovat v průběhu času a modelovat trend očekávaného vývoje. Současně lze informace smysluplně matematicky propojovat, například vypočítávat počet žáků na učitele, vývoj průměrného počtu žáků ve třídě, v oboru apod. Ukázkou první strany školského výkazu o základní škole nabízí příloha 1.

V následující tabulce jsou uvedeny školské výkazy, ze kterých se budou načítat všechny údaje do datového skladu. Jejich identifikace v datovém skladu bude z důvodu přehledného setřídění označena písmenem V a příslušnými arabskými číslicemi. Datový sklad bude strukturován dle Výkazu / oddílu / řádko-sloupce.

**Tabulka 2 Výkazy škol a školských zařízení**

<b>Zkratka</b>	<b>ID v DS</b>	<b>Název výkazu</b>
<b>S 53-01</b>	V53	Výkaz o zahájení povinné školní docházky v základní škole
<b>S 1-01</b>	V01	Výkaz o mateřské škole
<b>Z 2-01</b>	V02	Výkaz o školní družině – školním klubu
<b>M 3</b>	V03	Výkaz o základní škole
<b>S 4c-01</b>	V04c	Výkaz o přípravné třídě základní školy a přípravném stupni základní školy speciální
<b>R 13-01</b>	V13	Výkaz o ředitelství škol
<b>Z 24-01</b>	V24	Výkaz o základní umělecké škole

Zdroj: vlastní zpracování

## 9.4.2 Dimenzionální model

Jak již bylo řečeno v kapitole 5, dimenzionální modely databází se skládají ze dvou typů struktur, a to z centrálních tabulek faktů a k nim přidružených tabulek dimenzí. Cílem této podkapitoly je tedy určit jaká fakta a dimenze budou v datovém skladu ukládána.

### 9.4.2.1 Dimenze

Na základě provedené analýzy datových zdrojů, v tomto případě sledovaných číselníků, byly určeny jednotlivé tabulky dimenzí. Jejich výčet i s potřebnými atributy je uveden v příloze 3.

### 9.4.2.2 Fakta

Sledované výkazy či jejich kombinace představují jednotlivé tabulky faktů. Datový sklad tak bude obsahovat jedinečné údaje o jednotlivém faktu charakterizovaném v daném výkazu, oddílu (tabulce) číslem řádku a číslem sloupce. Z tohoto důvodu je potřeba z pohledu zachování věcné kontinuity dat v čase řešit případy, kdy tvůr-



ci na MŠMT mění v průběhu času text v řádku či sloupci u daného faktu (ukazatele).

V datovém skladu dále mohou být vypočítávány i takové údaje, které nejsou ve výkazech uváděny, například vedle počtu dívek také počty chlapců, rozdílové hodnoty mezi celkovým údajem a dílčími údaji apod.

Všechny tabulky faktů, které i s popisem atributů nabízí příloha 4, mají obdobnou strukturu. Obsahují uměle generované klíče (SK, Surrogate Key) odkazující do přidružených dimenzí a atribut Hodnota představující v kombinaci s daným ukazatelem (atribut Ukazatel SK) sledovaný faktický údaj.

#### **9.4.2.3 Schéma**

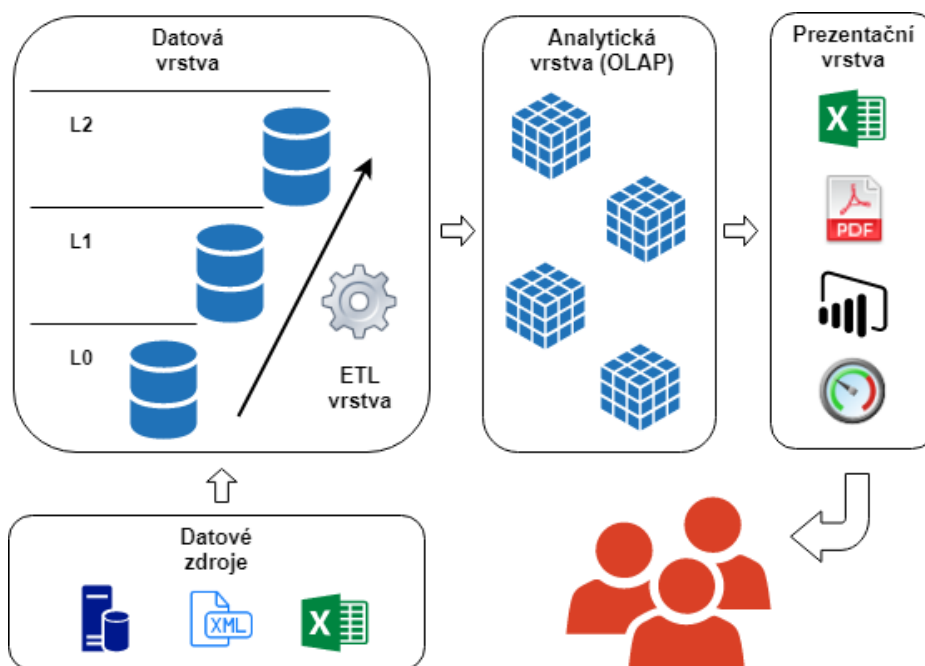
Schéma dimenzionálního modelu představuje způsob propojení tabulek faktů a k nim přidružených tabulek dimenzí. V tomto případě se jedná o schéma hvězdy se sdílenými dimenzemi, které znázorňuje příloha 5. Pro přehlednost schématu jsou však zahrnuty pouze tři z celkových sedmi sledovaných tabulek faktů.

#### **9.4.3 Architektura**

Architektura navrhovaného BI/DWH řešení se skládá z několika vrstev, jak je znázorněno na obrázku 17, přičemž bližší specifikace jednotlivých vrstev navrhovaného řešení budou uvedeny v následujících podkapitolách.

Stručně řečeno, řešení obsahuje dostatečně dimenzovaný datový sklad, který umožňuje časté dotazy nad velkými objemy dat. Pro rozsáhlé agregační výpočty je použita robustní agregační vrstva, která je schopna zpracovávat požadavky v reálném čase. Co se prezentační vrstvy týče, je navržena takovým způsobem, aby data mohla být zobrazována nejen přehledně, ale rovněž bez velkých časových odezev.

Je rovněž dobré zmínit, že architektura datového skladu, analytických a reportovacích systémů obsahuje postupy a zkušenosti, které se v dnešní době při budování obdobných systémů používají a jsou pro sledovanou oblast a pro navrhované řešení nejlepší praxí.



**Obr. 17 Architektura navrženého BI/DWH řešení**  
Zdroj: vlastní zpracování

#### 9.4.3.1 Datová vrstva (Datový sklad)

Pro vybudování datového skladu byl zvolen návrh s jednotným byznys modelem, který je určen k přípravě metadat. Jednotný byznys model umožní administrátorům intuitivní správu atributů, tabulek dimenzí, tabulek faktů, hierarchií, sledovaných metrik a rovněž jednotné nastavení přístupových práv koncových uživatelů.

Navržená architektura datového skladu přitom vychází z principů třívrstvé architektury (viz kapitola 6.3.1.3) a architektury sběrnice neboli architektury spojených datových trhů (viz kapitola 6.3.2.2), kterou vyvinul Ralph Kimball. Architektura jednotné datové základny (datového skladu) se tak skládá z následujících relačních vrstev:

- Relační vrstva L0 – Vrstva L0 slouží pro přesun a dočasné uložení vstupních dat z primárních datových zdrojů. Tato část datového skladu rovněž umožňuje oddělit proces extrakce a přenosu dat od procesu jejich zpracování a následného uložení v dalších vrstvách datového skladu. Jednorázově bude také využita pro prvotní načtení historických

dat (tzv. iniciální load). Pravidelná denní aktualizace pak bude obsahovat pouze ta data, u kterých lze předpokládat, že došlo k jejich aktualizaci na zdroji. Vrstva L0 je vždy 100% obrazem datového zdroje a nedochází v ní k žádnému čištění dat (například z důvodu datových typů).

- Relaçní vrstva L1 – Vrstva L1 je základní relační vrstva, která slouží pro účely uložení historických dat na co nejnižší možné úrovni granularity (detailní data) a plní tak především archivační funkci (odpovídá za dlouhodobou správu dat). Jedná se o kumulaci dat načítaných z vrstvy L0, tedy konsolidaci iniciálního načtení dat a denních nápočtů. Data jsou přitom uložena ve třetí normální formě.
- Relaçní vrstva L2 – Jedná se o vrstvu, která slouží jako podklad pro analytickou úroveň datového skladu. Poskytuje data aplikacím a uživatelům v podobě optimalizované pro jejich potřeby. V principu vrstva L2 obsahuje tabulky faktů a číselníky pro dimenze, jedná se tedy o uložení dat do jednotlivých datových trhů. Vrstva L2 může být kdykoliv naplněna daty z vrstvy L1, přičemž mezi vrstvou L1 a L2 dochází ke značné míře transformací dat.

#### 9.4.3.2 ETL vrstva

ETL vrstva představuje velmi důležitou komponentu celé architektury, jelikož zahrnuje transformační a čistící mechanismy dat. Jejím hlavním úkolem je tedy zajistit extrakci dat ze zdrojových systémů různého charakteru (relaçní databáze, souborové systémy, XML data apod.), jejich následnou transformaci a uložení do datových struktur datového skladu.

Vrstva zároveň zajišťuje očištění a validaci dat na základě uživatelem specifikovaných pravidel. Chybná data, která neprojdou povinnými datovými kontrolami, je přitom možné opravit a opakovaně načíst.

Ve vrstvě ETL jsou také obsaženy procedury na zpracování dat i samotné plánování kdy a za jakých podmínek se potřebné transformace provedou. Jedná se o předem naplánované workflow opakující se zpravidla jednou denně. Distribuce dat (i těch detailních) je poskytována i do okolních systémů.

Shrnutí vlastností vrstvy ETL:

- Deklarativní návrh a vývoj modulů ETL pro zajištění jednoduché správy a údržby oproti ručně psaným kódům ETL procesů.
- Vysoký výkon a propustnost ETL procesů využitím konceptu ETL – nejprve proběhne extrakce a nahrání dat do cílového úložiště a fáze transformace dat proběhne až v cílovém úložišti s využitím výkonu databázového serveru datového skladu.
- Možnost extrakce dat z různorodých datových zdrojů (relační databáze, CSV, XML, TXT).
- Možnost definice pravidel datové kvality nebo korekčních mechanismů pro datovou kvalitu.
- Podpora sjednocování jednotlivých ETL procesů do jednoho a definování závislostí jednotlivých procesů (workflow).
- Možnost automatizace spouštění a monitorování ETL procesů.
- Podpora logování průběhu a vzniklých chyb v rámci ETL procesů.
- Možnost restartovatelnosti ETL procesů.
- Podpora vysledování toku dat ze zdroje dat na cíl v rámci ETL procesů.
- Podpora analýzy a hromadného zpracování (indexace) nestrukturovaných dat (PDF dokumenty, MS Office dokumenty).

#### **9.4.3.3 Analytická vrstva**

Pro prezentaci dat koncovým uživatelům slouží data uložená v analytické vrstvě. Základem této vrstvy je analytický server, který představuje centrální bod k jednotnému přístupu k datům. Jinými slovy, analytický server umožňuje snazší přístup k datům a jejich prezentaci skrze jednotné uživatelské rozhraní, v jazyce a pojmech, kterým koncoví uživatelé rozumí.

Analytický server pracuje s již konsolidovanými a transformovanými daty z relační vrstvy L2 a poskytuje možnost nad těmito daty vytvářet různé reporty a analýzy dat. Díky metadatům uložených v datovém skladu analytický server umožňuje koncovým uživatelům provádět i ad-hoc dotazování.

Analytický server také umožňuje bezpečný přístup k datům. To znamená, že autentizace a autorizace uživatelů (přístup a oprávnění) jsou definovány na úrovni analytického serveru a ne na úrovni databáze.

Shrnutí vlastností analytické vrstvy:

- Snadná publikace a sdružování finálních reportů do podoby analytických panelů (tzv. dashboardů).
- Tvorba prezentací, dashboardů, sestav a ad-hoc analýz.
- Možnost použití HOLAP, který slučuje prvky MOLAP a ROLAP. Ponechává původní data v relačních tabulkách, zatímco agregace ukládá v multidimenzionální podobě. Více viz kapitola 8.2.
- Vysoká interaktivita sestav – možnost drillování, slicing a dicing z agregovaných informací, filtrování a parametrizace výstupů.
- Možnost nastavení různých typů hierarchie – například hierarchie s možností přeskokovat chybějící záznamy, hierarchie s navazujícími záznamy, hierarchie s šablonou organizační struktury či hierarchie „Parent-Child“.
- Možnost grafické prezentace dat – prezentace informací přes jednotné uživatelské rozhraní ve formě statických textů, tabulek, grafů, křížových tabulek, KPI ukazatelů apod.
- Možnost exportu výstupů do různých formátů (PDF, CSV, MS Excel).
- Možnost proaktivního monitorování a upozorňování na nestandardní hodnoty sledovaných metrik.
- Podpora periodické distribuce výstupů vybraným skupinám koncových uživatelů s podporou různých typů zařízení (mobilní telefon, e-mail).
- Podpora různorodých datových zdrojů (multidimenzionální databáze, relační databáze, soubory XML, TXT, CSV).
- Odstínění konzumentů dat od sémantiky a topologie datových zdrojů v podobě jednotné vrstvy metadat.

#### 9.4.3.4 Prezentáční vrstva

Prezentáční vrstva slouží k zobrazování dat z analytické vrstvy koncovému uživateli, a to jak pracovníkům, tak i široké veřejnosti a partnerům (intranet, internet). Tato vrstva kooperuje s analytickým serverem, přičemž se stará o správné zobrazování dat. Předává požadavky koncových uživatelů analytickému serveru a následně vrací výsledky ve vizualizované podobě.

Prezentáční vrstva je přitom rozdělena do několika přístupových bodů (reporty, webové rozhraní, mobilní přístup), přičemž poskytuje jednotný pohled na vizualizovaná data. Všechna data je navíc možno prohlížet, aniž by koncový uživatel musel mít nainstalovanou aplikaci.

Shrnutí vlastností prezentáční vrstvy:

- Přístup ke všem funkcionalitám BI (analýzy, reporty, sestavy apod.).
- Mobilní verze pro přístup z mobilního telefonu a tabletu.
- Export reportů do různých standardních formátů.
- Podpora automatické distribuce vytvořených reportů.
- Možnost ad-hoc analýzy dat a dotazování.
- Podpora administrace jednotlivých reportů.
- Jednotný konzistentní pohled na uložená data.
- Možnost samoobslužného přístupu k informacím.
- Možnost distribuce vytvořených sestav a reportů e-mailem.

## **9.5 Pilotní implementace**

Vybudování datového skladu je pracná a časově velmi náročná záležitost, a to i v celém týmu odborníků. Z toho důvodu bude tato část praktické práce věnována implementaci pouze pilotního projektu celého řešení. V následujících podkapitolách tak bude představeno vytvoření databází, tabulky faktů, datové pumpy a OLAP kostky a uvedeno použité hardwarové a softwarové vybavení.

### **9.5.1 Použitý HW a SW**

Pilotní implementace navrženého BI/DWH řešení je realizována v testovacím prostředí TC KHK (Technologické centrum Královéhradeckého kraje) a využívá jeho současné hardwarové a softwarové infrastruktury.

#### **Popis dostupného vybavení a zdrojů**

- Datové úložiště – Dva diskové systémy s kapacitou řádově desítek TB zapojené v režimu s vysokou dostupností (tzn. synchronní replikace mezi datovými centry).
- Databázový server – Server provozovaný v režimu s vysokou dostupností dle použité virtualizace VMware vSphere 5 Enterprise. Jedná se o databázový server MS SQL 2008R2, který je provozován v Microsoft 2008R2 clusteru licencovaný na CPU.
- Aplikační server – Server provozován v rámci serverové virtualizace v režimu s vysokou dostupností. Hardwarové omezení dle použité virtualizace VMware vSphere 5 Enterprise. Operačním systémem je Windows 2008R2 server edice DC.

## **MS SQL Server 2008 R2**

MS SQL Server 2008 je technologie, která byla využita k realizaci pilotní implementace navrženého BI/DWH řešení. Jedná se o produkt vyvinutý společností Microsoft pro datovou platformu, který díky dalším nástrojům, jež integruje, umožňuje tvorbu a následnou správu datových skladů. Mezi tyto nástroje řadíme následující:

- Databázový stroj – Poskytuje komplexní realizaci systému řízení báze dat (ukládání, zpracování a zabezpečení dat).
- Integrované služby (SQL Server Integration Services, SSIS) – Prostřednictvím SSIS balíčků je umožněna realizace ETL procesů (datových pump) pro extrakci, transformaci a nahrávání dat do datového skladu.
- Analytické služby (SQL Server Analysis Services, SSAS) – Tento nástroj umožňuje a usnadňuje vytváření analytických databází a OLAP kostek.
- Reportovací služby (SQL Server Reporting Services, SSRS) – Jedná se o serverovou platformu poskytující funkcionality potřebné pro prezentaci dat z databázového stroje či analytických databází.
- SQL Agent – SQL Agent je služba pro plánování úloh, která se nejčastěji využívá pro automatizaci spouštění datových pump (SSIS balíčků).

Implementace datových struktur byla realizována pomocí nástroje MS SQL Server Management Studio 2008 a ETL procesy potřebné k naplnění těchto struktur pomocí nástroje MS Visual Studio 2008.

### **9.5.2 Vytvoření struktur datového skladu**

První fází při vývoji navrženého řešení je vytvoření struktur datového skladu. Strukturami datového skladu se myslí všechny komponenty nezbytné pro jeho provoz. Jedná se tedy o databáze představující jednotlivé vrstvy třívrstvé architektury, dále tabulky obsažené v těchto databázích, procedury, triggeru apod.

K vytvoření relačních a analytických databází a jejich tabulek lze přitom využít grafické rozhraní Management Studia nebo SQL skripty, které se o vytvoření potřebných struktur datového skladu postarají.



Pro vytvoření všech výše zmíněných struktur je však nejprve potřeba se připojit k databázovému serveru. K tomu stačí vyplnit název požadovaného serveru a zvolit typ autentizace (viz obrázek 18). V tomto případě se jedná o Windows autentizaci, a tudíž není potřeba zadávat heslo.



**Obr. 18 Připojení k databázovému serveru KHK**  
Zdroj: vlastní zpracování

Dalším krokem je vytvoření tří databází představujících vrstvy L0, L1 a L2 systému datového skladu. K tomu lze využít SQL skript *CREATE DATABASE nazev\_vrstvy*, kde se za *nazev\_vrstvy* postupně dosadí *DS\_KHK\_L0*, *DS\_KHK\_L1* a *DS\_KHK\_L2*.

Po vytvoření databází je možno přistoupit k vytvoření tabulek, které tyto databáze budou obsahovat. Pro lepší přehlednost a případnou orientaci v rámci databází bude dodržena konvence názvů jednotlivých tabulek. Každá tabulka se tedy bude skládat z jednopísmenného prefixu určujícího typ tabulky (tabulka faktů – F, tabulka dimenze – D), následovat bude název tabulky oddělený z obou stran podtržítkem a na konci bude uveden sufix představující vrstvu datového skladu (L0, L1 či L2). Jako příklad můžeme uvést název pro tabulku faktů týkající se výkazu o ředitelství škol, jejíž názvy v jednotlivých vrstvách budou *F\_VykazV13\_L0*, *F\_VykazV13\_L1* a *F\_VykazV13\_L2*. Na této tabulce faktů budou zároveň předvedeny SQL skripty pro vytvoření tabulky v každé z vrstev datového skladu.

Jak již bylo zmíněno v návrhu a popisu vrstev datového skladu, vrstva L0 představuje 100% obraz datového zdroje a nedochází v ní k žádnému čištění dat.

Cílem je tedy bez velkých obtíží do této vrstvy nahrát data ze zdrojových souborů představujících vyplněné údaje daných výkazů a číselníků. Z toho důvodu budou veškeré sloupce všech tabulek této vrstvy bez ohledu na možný datový typ vytvořeny jako textový datový typ *nvarchar(255)* či *nvarchar(max)*, aby se předešlo problémům s konverzí a možným přetečením datových typů. Konverze na správný datový typ bude probíhat až při přesunu dat z vrstvy L1 do vrstvy L2 společně s dalšími transformacemi, jako je například očištění dat (viz podkapitola 9.5.3). Příklad SQL skriptu pro vytvoření tabulky *F\_VykazV13\_L0* je následující:

```
CREATE TABLE [dbo].[F_VykazV13_L0](
  [Misto] [nvarchar](255) NULL,
  [Redlzo] [nvarchar](255) NULL,
  ...
  [Hodnota] [nvarchar](255) NULL,
)
```

Stejně jako je vrstva L0 obrazem datového zdroje, tak vrstva L1 je obrazem vrstvy L0. Nedochozí zde tedy ke změně struktury tabulek či datových typů. Na rozdíl od vrstvy L0, která slouží k načtení pouhého přírůstku dat, však vrstva L1 uchovává data za celou historii. SQL skript pro vytvoření tabulky *F\_VykazV13\_L1* je následující:

```
CREATE TABLE [dbo].[F_VykazV13_L1](
  [Misto] [nvarchar](255) NULL,
  [Redlzo] [nvarchar](255) NULL,
  ...
  [Hodnota] [nvarchar](255) NULL,
)
```

Jak již bylo řečeno, ve vrstvě L2 jsou sloupce vytvořeny ve správném datovém typu, tudíž například pro celočíselné hodnoty použijeme datový typ *int*, pro časové a kalendářní údaje datový typ *datetime* apod. V rámci přesunu dat z vrstvy L1 do vrstvy L2 dochází tedy minimálně ke konverzi dat, ale zpravidla ještě k několika dalším transformacím. U tabulek faktů je tou nejdůležitější úlohou doplnění SK klíčů za hodnoty z jednotlivých číselníků (dimenzí). Všechny tabulky je rovněž ve vrstvě L2 nutné opatřit primárním klíčem. Každý faktický údaj je navíc napojen na časovou dimenzi.

SQL skript pro vytvoření tabulky *F\_VykazV13\_L2* vypadá následovně:

```
CREATE TABLE [dbo].[F_VykazV13_L2](
  [Cas_SK] [int] NOT NULL,
  [Redlzo_SK] [int] NOT NULL,
  [Ukazatel_SK] [int] NOT NULL,
  ...
  [Hodnota] [numeric](18, 2) NULL,
  CONSTRAINT [PK_F_VykazV13_L2] PRIMARY KEY CLUSTERED
  ([Cas_SK] ASC, [Redlzo_SK] ASC, [Ukazatel_SK] ASC)
)
```

V rámci vytváření tabulek pomocí výše uvedených skriptů dochází rovněž k nastavování tzv. rozšířených vlastností tabulky. Jedná se o vlastnosti, které budou využity k parametrizaci a dynamickému mapování v rámci ETL procesů (viz například *Cilova\_tabulka* a *Zdrojova\_tabulka* z obrázku 19), což v konečném důsledku umožní vytvořit šablonu SSIS balíčku pro naplnění tabulek faktů a dimenzí (viz následující podkapitola). Příklad SQL skriptu jak takovou vlastnost vytvořit, je následující:

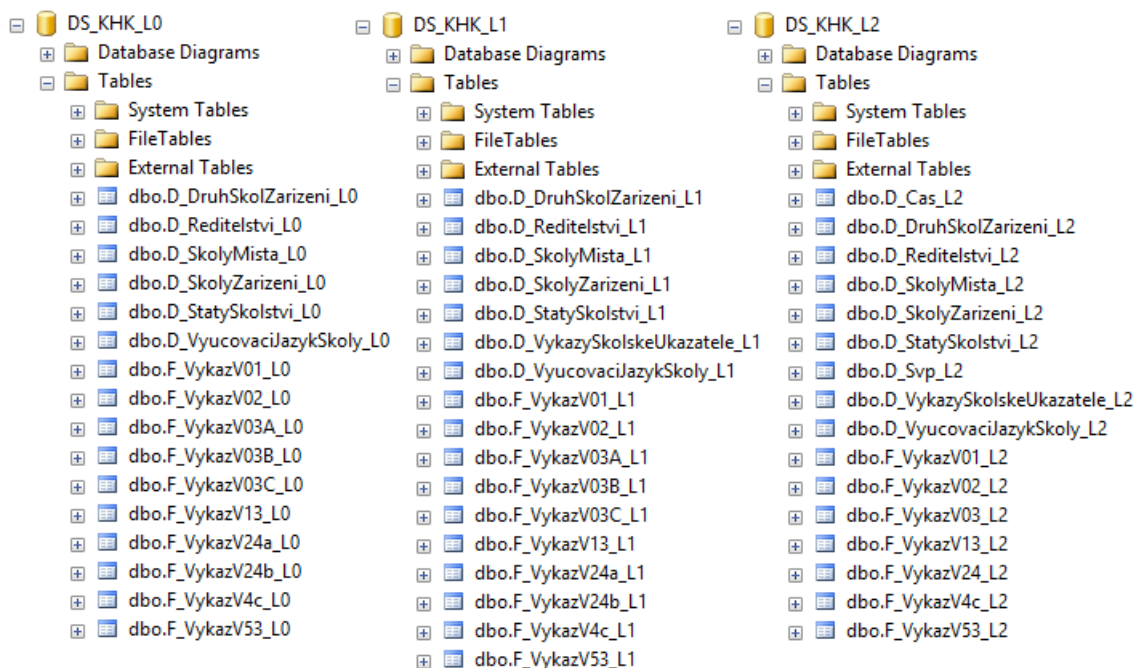
```
EXEC sys.sp_addextendedproperty @name=N'Zdrojova_tabulka',
@value=N'F_VykazV13_L1', @level0type=N'SCHEMA',@level0name=N'dbo',
@level1type=N'TABLE',@level1name=N'F_VykazV13_L2'
```

Database:	DS_KHK_L1
Collation:	Czech_CI_AS
Properties:	
Name	Value
Cilova_tabulka	F_VykazV01_L2
Externi_datovy_zdroj	
MS_Description	Výkaz V01 o mateřské škole
Zdrojova_tabulka	F_VykazV01_L0

Obr. 19 Ukázka rozšířených vlastností tabulky *F\_VykazV01\_L1*

Zdroj: vlastní zpracování

Veškeré představené postupy pro vytvoření tabulek i s jejich případnými rozšířenými vlastnostmi je nutno provést i pro všechny zbylé tabulky faktů a dimenzí. Výsledkem je struktura datového skladu, kterou znázorňuje obrázek 20.



**Obr. 20 Výsledná struktura datového skladu**  
Zdroj: vlastní zpracování

Je rovněž důležité zmínit, že některé zpracovávané výkazy jsou rozprostřeny do více tabulek faktů ve vrstvě L0 a současně ve vrstvě L1. Příčinou je, že dané výkazy jsou načítány z více zdrojových souborů. To odpovídá již zmíněné skutečnosti, že vrstva L0 obsahuje obraz každého datového zdroje a vrstva L1 obraz každé tabulky z vrstvy L0. Ve vrstvě L2, což je podkladová vrstva pro prezentaci dat, se však uchovávají všechna data pohromadě, a proto je potřeba provést jejich sjednocení do jedné výsledné tabulky.

### 9.5.3 Realizace ETL procesu

ETL procesy pro extrakci dat ze zdrojových systémů a načtení do datového skladu byly realizovány prostřednictvím nástroje SQL Server Integration Services (SSIS) v prostředí Microsoft Visual Studio 2008. Tento typ projektu je uzpůsoben pro tvorbu tzv. balíčků, které umožňují načítat data z různých zdrojů, následně je transformovat pomocí velkého množství komponent a poté nahrát do databázového úložiště, případně vyexportovat například do CSV či TXT souboru.

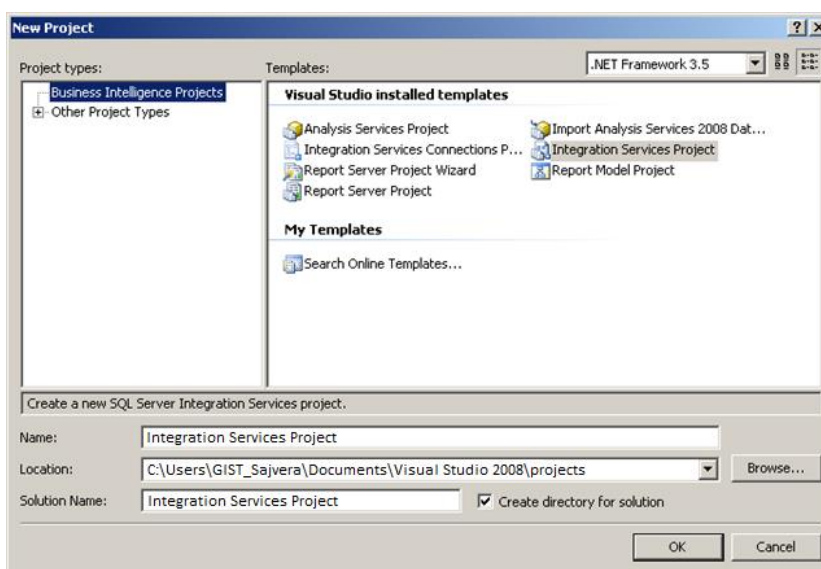
Obvykle se vytváří tolik SSIS balíčků, kolik je datových zdrojů, a tedy kolik je tabulek ve vrstvě L0, která představuje dočasné úložiště dat. Často se jedná o de-

sítky i více tabulek, respektive balíčků. Z toho důvodu je vhodné vytvořit šablonu SSIS balíčku pro každý typ datového zdroje, která vytváření těchto balíčků usnadní a která bude co nejvíce konfigurovatelná takovým způsobem, aby byla zajištěna možnost rozšířit ji o další parametry.

V následujícím textu bude představeno vytvoření šablony SSIS balíčku pro naplnění tabulek faktů, které představují školské výkazy a jsou plněny ze zdrojových souborů. Postup tvorby se skládá z několika následujících kroků.

### Vytvoření nového SSIS projektu

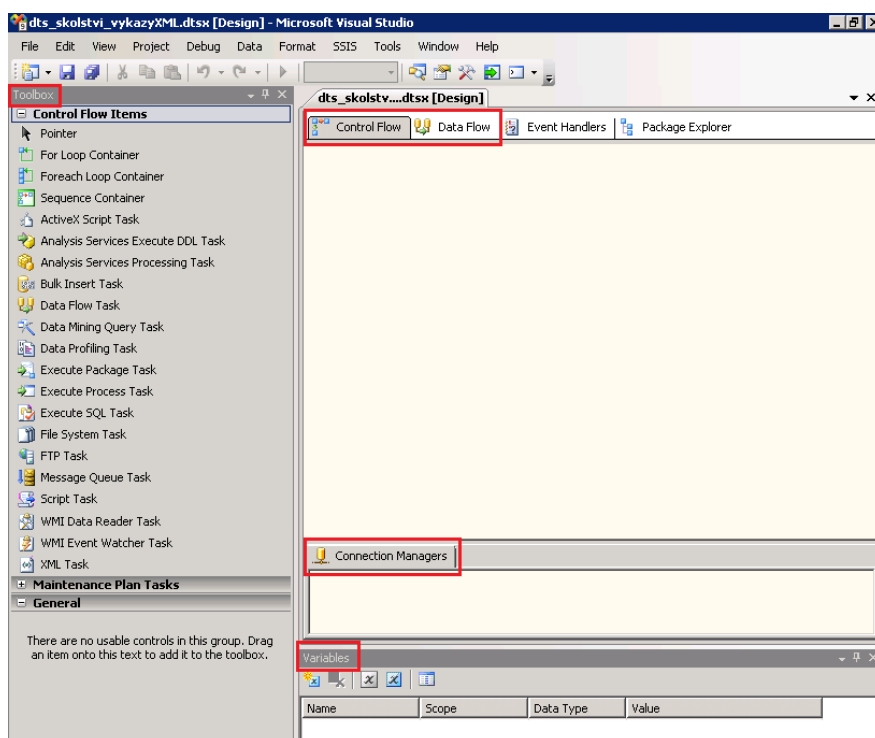
Nový projekt lze jednoduše vytvořit tak, že se z nabídky *File* zvolí možnost *New* a následně možnost *Project*. Pro rychlejší přístup je možno použít klávesovou zkratku *Ctrl+Shift+N*. V okně pro výběr typu nového projektu se z dostupných možností vybere možnost *Business Intelligence Projects*, v rámci které se konkrétně označí šablona *Integration Services Project* (viz obrázek 21). Pro lepší přehlednost je doporučeno projekt vhodně pojmenovat a následným stiskem tlačítka *OK* dokončit proces jeho vytvoření.



**Obr. 21 Vytvoření nového SSIS projektu**  
Zdroj: vlastní zpracování

Po vytvoření projektu lze spatřit několik panelů, ze kterých se skládá řešení celého projektu (viz obrázek 22). Jedná se o panel SSIS komponent (Toolbox), panel pro vytvoření připojení (Connection Managers), panel s proměnnými (Variables)

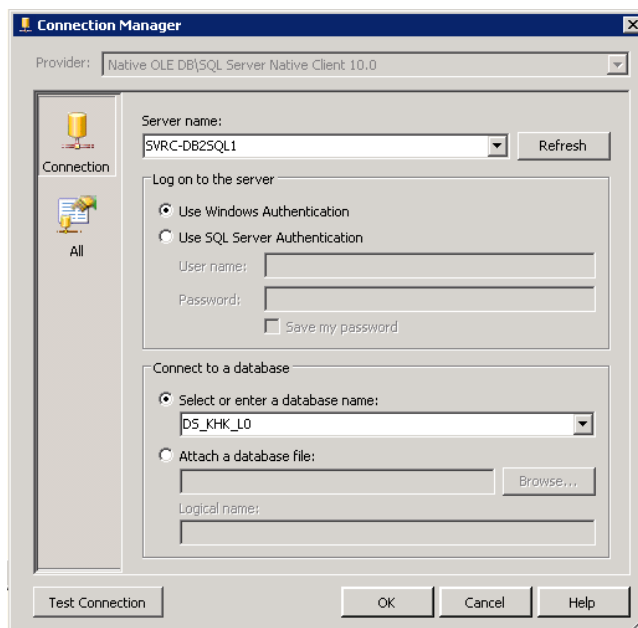
a hlavní panel pro vytváření obsahu balíčku, kde nejdůležitější roli zaujmají záložky kontrolního a datového toku dat (Control Flow, Data Flow).



**Obr. 22 Panely SSIS projektu**  
Zdroj: vlastní zpracování

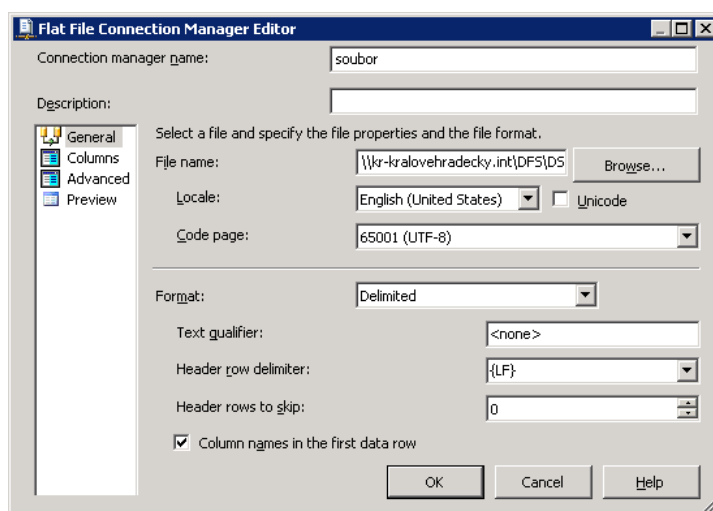
## Vytvoření potřebných připojení

V rámci vykonávání ETL procesu budou využity komponenty (například Execute SQL Task pro vykonání SQL skriptu), které ke své činnosti potřebují připojení do databáze. Takové připojení je možno vytvořit kliknutím pravého tlačítka v panelu *Connection Managers*, přičemž se zvolí možnost *New OLE DB connection*. Zobrazí se tabulka (viz obrázek 23), ve které je nejprve nutno vyplnit název databázového serveru, ke kterému se komponenty budou připojovat. Následně je potřeba vybrat typ autentizace, přičemž pro účely projektu se zvolí Windows autentizace. Nakonec se z rozbalovací nabídky vybere databáze, kde se nachází cílová tabulka. V této ukázce se jedná o databázi *DS\_KHK\_L0*, nicméně v rámci balíčku bude potřeba připojení i do databází *DS\_KHK\_L1* a *DS\_KHK\_L2*. Stejným způsobem se tedy vytvoří připojení i do těchto databází. Pro ověření spojení s daným serverem je možno využít tlačítko *Test Connection*, které poskytuje informaci, zda pokus o připojení proběhl úspěšně či nikoliv.



**Obr. 23 OLE DB připojení do databáze**  
Zdroj: vlastní zpracování

Rovněž je potřeba vytvořit připojení ke zdrojovému souboru, ze kterého budou načítána data do datového skladu. K tomu je nutné vytvořit *Flat File connection* (viz obrázek 24), ve kterém je nezbytné zdrojový soubor blíže specifikovat. Tím je myšleno určit kódovou stránku, oddělovač hodnot a také cestu, kde se soubor nachází. Cesta se zde uvádí z důvodu, aby bylo možné zjistit a případně upravit strukturu souboru (jednotlivé sloupce) na dalších záložkách *Columns* a *Advanced*. Tato cesta bude jinak načítána dynamicky z jedné z vytvořených proměnných.

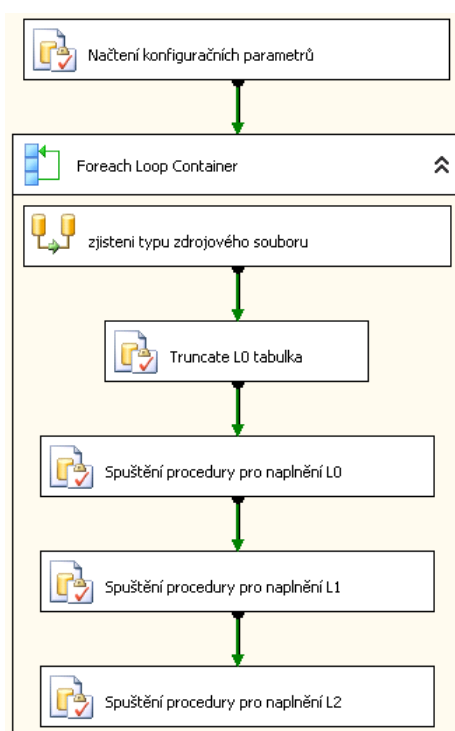


**Obr. 24 OLE DB připojení do databáze**  
Zdroj: vlastní zpracování

## Realizace Control Flow a Data Flow

Nejdůležitějším krokem je sestavení komponent kontrolního a datového toku, které lze jednoduše do těchto toků umístit přetažením pomocí myši z panelu komponent pro daný tok.

Kontrolní tok definuje průtok úloh, které mají být postupně vykonány. Může se například jednat o spuštění SQL skriptů, provedení cyklů, odeslání emailu či zavolání COM rozhraní. Také zde může být umístěn pouze odkaz do datového toku, ve kterém je možno pracovat přímo s načítanými daty. Kontrolní tok našeho SSIS balíčku ilustruje následující obrázek 25.



**Obr. 25 Control Flow**  
Zdroj: vlastní zpracování

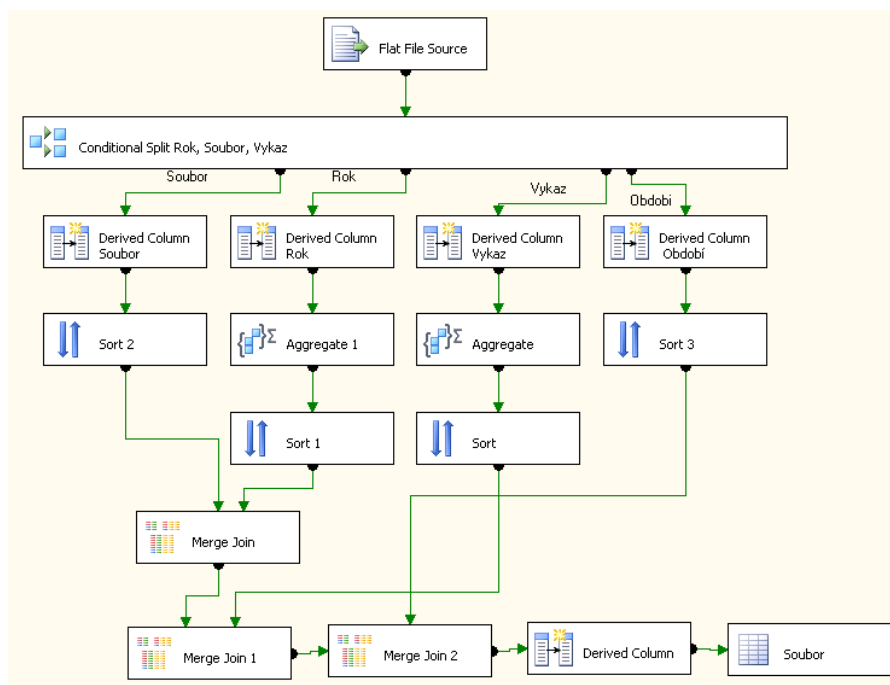
Nejprve je vykonán SQL skript pomocí komponenty *Execute SQL Task*, který do proměnných načte konfigurační parametry potřebné pro běh celého ETL procesu. Jedná se o parametry prostředí (názvy databází a tabulek) a parametry k určení struktury zdrojového souboru.

Další použitou komponentou je *ForeachLoop*, která postupně iteruje přes jednotlivé soubory obsažené ve zdrojové složce (cesta složky je získána z promě-



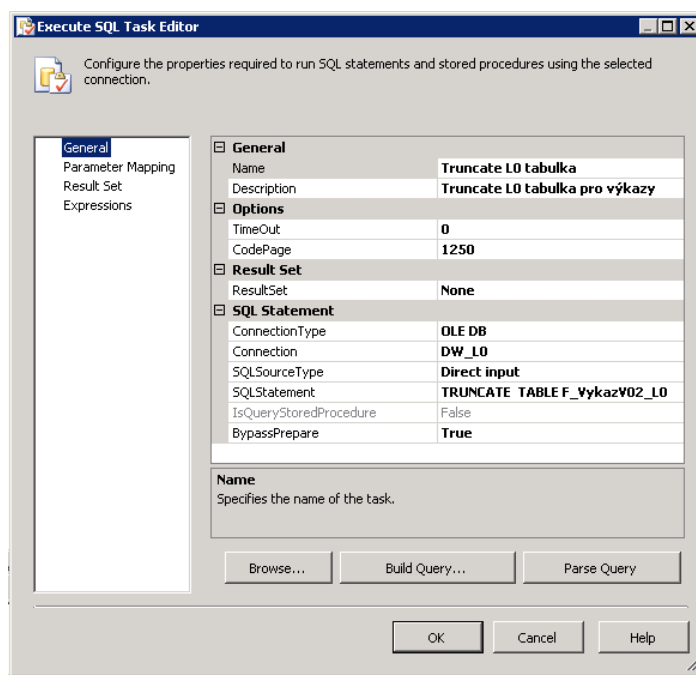
né) a podle masky vyhledává konkrétní zdrojové soubory, jelikož jak již bylo řečeno, data jedné tabulky faktů mohou být získávána z více zdrojových souborů. V případě nalezení požadovaného souboru je do proměnné uložena cesta k danému souboru, která je využita jako zdroj pro již vytvořené *Flat File connection* připojení. V rámci *ForEachLoop* je postupně vykonáno několik dalších úloh.

První obsaženou komponentou ve výše zmíněném cyklu je odkaz do datového toku. Ten určuje tok dat z právě zjištěného zdrojového souboru do cílové destinace, kterou je komponenta *Record Set*. V rámci tohoto toku je provedena celá řada transformací (viz obrázek 26). Souhrnně lze říci, že se jedná o transformace potřebné pro zjištění parametrů vstupního souboru. Ten je díky tomu rozebrán na dílčí atributy nezbytné pro načtení dat do datového skladu (atributy školní rok, jméno výkazu apod.), které jsou společně s faktickými údaji uloženy do cílové komponenty. Její obsah je namapován na proměnnou *Soubor* typu *Object*, která představuje kolekci načtených záznamů a figuruje jako vstup jedné z dalších komponent v rámci kontrolního toku.



**Obr. 26 Data Flow**  
Zdroj: vlastní zpracování

Výstupem datového toku je tedy kolekce záznamů, které je potřeba nahrát do datového skladu. K tomu je nutné vykonat čtyři po sobě jdoucí komponenty *Execute SQL Task*. První z nich je promazání tabulky ve vrstvě L0, aby bylo možno nahrát novou dávku dat. Nastavení této komponenty ilustruje následující obrázek 27.



**Obr. 27 Truncate tabulky ve vrstvě L0**  
Zdroj: vlastní zpracování

Jak je z obrázku patrné, komponentě bylo nastaveno připojení *DW\_L0*, tedy připojení do vrstvy L0 a má se provést příkaz *TRUNCATE TABLE*, přičemž název konkrétní tabulky je opět předáván pomocí proměnné.

Po promazání tabulky ve vrstvě L0 následuje její naplnění voláním odpovídající procedury v rámci dalšího *Execute SQL Tasku*, které je předána kolekce *Source* obsahující načítané záznamy. Ty jsou v poměru 1:1 nahrány do dané tabulky.

Předposlední úlohou kontrolního toku je volání procedury pro naplnění tabulky ve vrstvě L1. Tato vrstva ukládá data za celou historii. Je zde tedy velice důležité nastavení přepisu hodnot, ke kterému dochází při historizaci přírůstků dat v rámci vykonávání procedury. Jejím úkolem je tedy rozhodnout, zda mají být data dle nastavení přepisu na jednotlivých sloupcích před nahráním nejprve smazána či nikoliv.

Završením celého ETL procesu je volání procedury pro naplnění tabulky ve vrstvě L2. Tato vrstva je základem pro analytickou vrstvu a reporting. V rámci procedury tak dochází ke konverzi dat na správný datový typ, očištění dat (nahrazení NULL či nevyplněných hodnot za #), doplnění chybějících hodnot dimenzí z faktických záznamů, přidělení SK klíčů a následnému prostému překopírování dat z vrstvy L1. Jsou zde rovněž odstraněny pro OLAP nadbytečné sloupce. Platí tedy, že tabulky faktů obsahují pouze SK sloupce (odkazy do dimenzí) a hodnotové sloupce (metriky). Dimenze obsahují primární SK a ID sloupce, atributy dimenze a případně sekundární SK sloupce (odkazy na další dimenze v hierarchii).

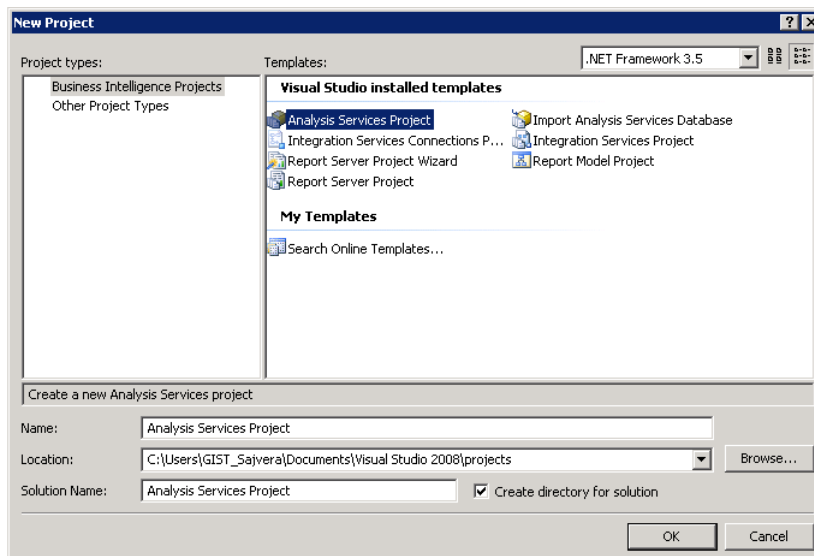
Dle výše uvedeného postupu byly vytvořeny balíčky pro naplnění všech tabulek faktů. Obdobným způsobem byla rovněž vytvořena šablona pro plnění tabulek dimenzí.

#### **9.5.4 Realizace OLAP kostky**

Data jsou koncovým uživatelům zpřístupněna prostřednictvím multidimenzionální OLAP kostky, která byla vytvořena pomocí nástroje SQL Server Analysis Services (SSAS) v prostředí Microsoft Visual Studio 2008. Postup její tvorby se skládá z několika kroků, které budou představeny níže. Pohled na atributy dimenzí a metriky, které kostka obsahuje, nabízí příloha 5.

#### **Vytvoření nového SSAS projektu**

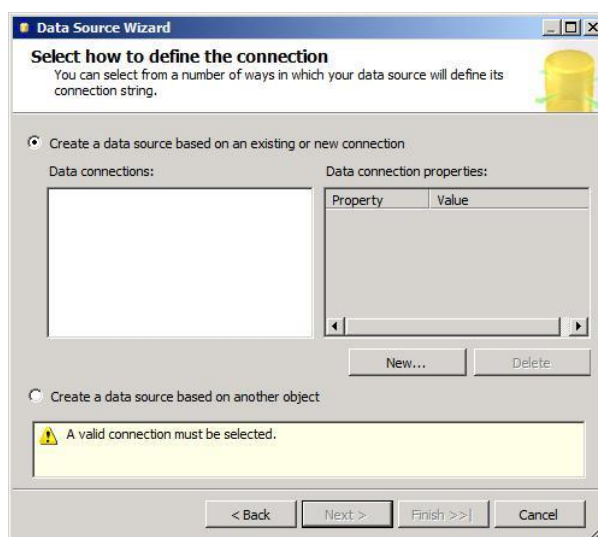
Nový projekt lze jednoduše vytvořit tak, že se z nabídky *File* zvolí možnost *New* a následně možnost *Project*. Pro rychlejší přístup je možno použít klávesovou zkratku *Ctrl+Shift+N*. V okně pro výběr typu nového projektu se z dostupných možností vybere možnost *Business Intelligence Projects*, v rámci které se konkrétně označí šablona *Analysis Services Project* (viz obrázek 28). Pro lepší přehlednost je doporučeno projekt vhodně pojmenovat a následným stiskem tlačítka *OK* dokončit proces jeho vytvoření.



**Obr. 28 Vytvoření nového SSAS projektu**  
Zdroj: vlastní zpracování

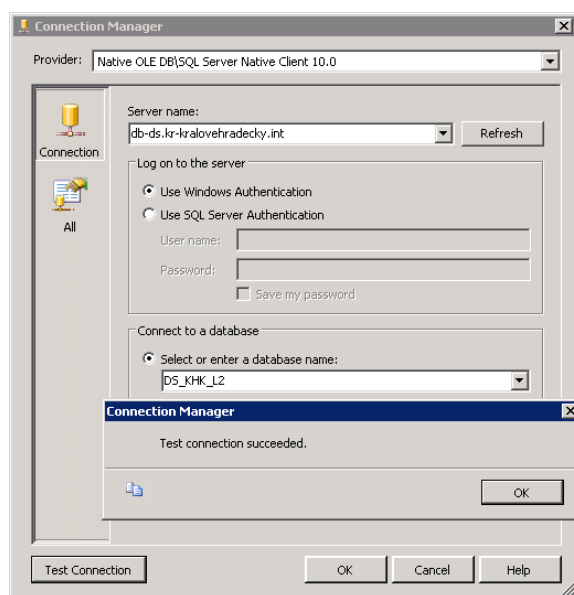
## Definice zdroje dat

Zdrojová data, která jsou základem pro vytvoření OLAP kostky, se definují v panelu průzkumníka řešení (Solution Explorer). V tomto případě se samozřejmě jedná o data uložená v datovém skladu, konkrétně ve vrstvě L2. Pomocí pravého tlačítka myši se klikne na *Data Sources* a následně se vybere možnost *New Data Source*, což způsobí, že se otevře průvodce datovými zdroji (viz obrázek 29).



**Obr. 29 Průvodce datových zdrojů**  
Zdroj: vlastní zpracování

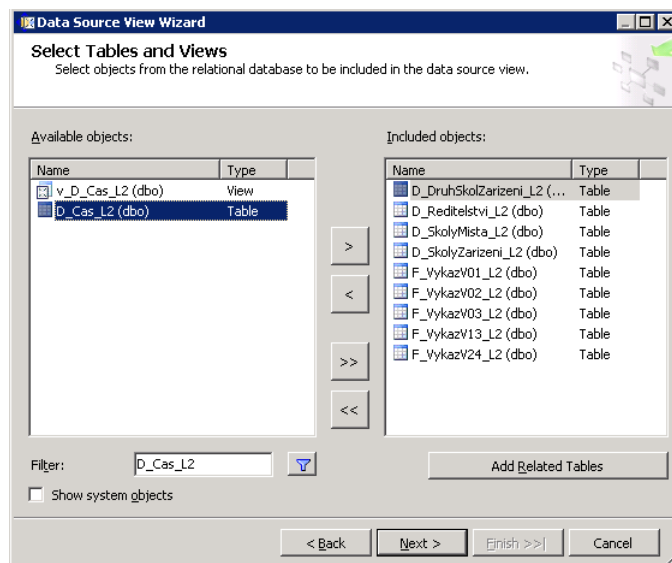
Kliknutím myši na *New* se definuje nové spojení pro zdrojová data. Z rozbalovací nabídky se vybere typ poskytovatele připojení a také se vyplní název serveru (viz obrázek 30). Následně je nutné vybrat způsob autentizace (v tomto případě Windows autentizace) a rovněž k jaké databázi je potřeba se na daném serveru připojit. Z rozbalovací nabídky se tedy zvolí databáze *DS\_KHK\_L2*, jelikož ta představuje podkladovou vrstvu pro OLAP a reporting. Pro jistotu je možno provést test spojení, a pokud vše proběhne v pořádku, potvrdí se definice zdroje nejprve stisknutím tlačítka *OK* a následně tlačítka *Finish*.



**Obr. 30 Průvodce datových zdrojů**  
Zdroj: vlastní zpracování

### Vytvoření pohledu na zdrojová data

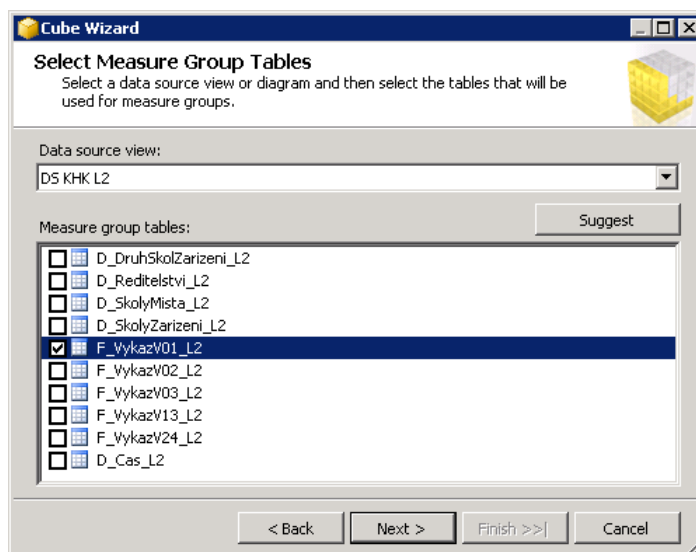
Nový pohled na zdrojová data lze vytvořit kliknutím pravého tlačítka na *Data Source View* v panelu průzkumníka řešení, přičemž se následně zvolí možnost *New Data Source View*. Otevře se průvodce vytvořením pohledu na zdrojová data. V sekci výběru tabulek a pohledů je možné určit požadované tabulky, které má výsledná OLAP kostka obsahovat (viz příloha 5, postup však shrnuje obrázek 31). Po provedení výběru tabulek lze vytvoření pohledu potvrdit nejprve kliknutím na tlačítko *Next* a následně na tlačítko *Finish*. Po vytvoření pohledu na zdrojová data se zobrazí schéma propojených tabulek, ve kterém je možné tabulky dále upravovat, například je přejmenovat, doplnit mezi nimi vazby nebo přidat vypočítávané položky na základě sloupců dané tabulky.



**Obr. 31 Vytvoření pohledu na zdrojová data**  
Zdroj: vlastní zpracování

### Vytvoření datové kostky

Posledním krokem je vytvoření samotné datové kostky. V průvodci řešení je tedy potřeba kliknout na *Cubes* a zvolit možnost *New Cube*. Otevře se průvodce pro vytvoření datové kostky, kde se nejprve vybere volba vytvořit kostku z existujících tabulek, tedy *Use existing tables*. Následně se v průvodci zaškrtně tabulka faktů (viz obrázek 32), tabulky dimenzí, zvolí se sledované metriky (tzv. Measures), kostka se pojmenuje a vytvoření se potvrdí tlačítkem *Finish*.



**Obr. 32 Výběr tabulek do OLAP kostky**  
Zdroj: vlastní zpracování

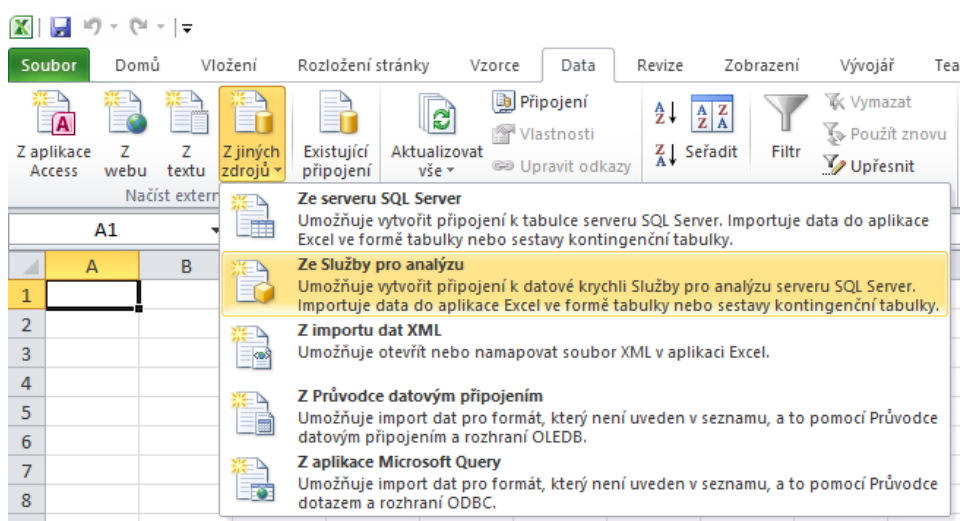
Tímto způsobem navržená datová kostka je pomocí nástroje SSAS fyzicky vytvořena spuštěním po sobě jdoucích procesů *Process* (test, zdali je kostka navržena správně) a *Deploy* (vytvoření a nahrání datové kostky na analytický server) v záložce *Build* v rámci průvodce řešení projektu.

## 9.6 Prezentace dat

Aplikace a nástroje pro vytváření analýz a reportů, se kterými v drtivé většině případů pracují běžní uživatelé bez větších technických znalostí, by měly mezi všemi nástroji BI patřit k těm méně složitým. Pro analýzu a prezentaci dat byl tedy zvolen nástroj MS Excel, ve kterém je vytváření reportů relativně snadnou záležitostí.

Prostřednictvím aplikace MS Excel budou koncoví uživatelé přistupovat k multidimenzionálním datům uložených v OLAP kostkách umístěných na analytickém serveru. Zejména pomocí kontingenčních tabulek si pak budou moci zobrazovat data a získávat z nich potřebné informace.

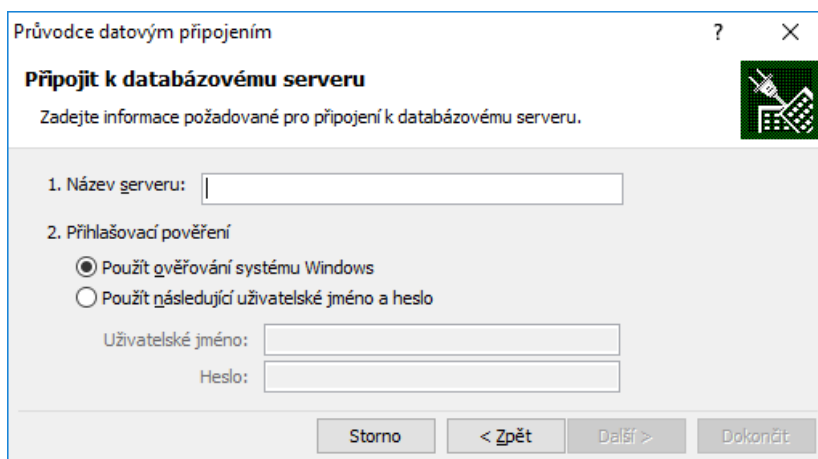
Jak znázorňuje obrázek 33, k OLAP kostce se dá připojit prostřednictvím aplikace MS Excel tak, že se na záložce *Data* zvolí v sekci externích zdrojů volba *Z jiných zdrojů* a následně se z rozbalovacího seznamu vybereme možnost *Ze služby pro analýzu*, která umožňuje definovat připojení k analytické databázi, kde je kostka uložena.



Obr. 33 Připojení k OLAP kostce v MS Excel

Zdroj: vlastní zpracování

Po vyplnění jména analytického serveru a způsobu autentizace v průvodci datovým připojením (viz obrázek 34) se vybere požadovaná datová kostka, přičemž je následně možné se na data multidimenzionálně dívat přímo v programu MS Excel a vytvářet nad nimi reporty.



**Obr. 34 Průvodce připojení k OLAP kostce v MS Excel**  
Zdroj: vlastní zpracování

Jak již bylo řečeno, data z OLAP kostky budou v prostředí MS Excel nejčastěji interpretována prostřednictvím kontingenční tabulky, která umožňuje zobrazovat vybraná fakta a dimenze v řádcích či sloupcích. Vždy záleží na požadovaném způsobu výsledné interpretace hodnot.

Příklady vytvořených kontingenčních tabulek jako možné prezentace dat nabízí příloha 6. Jedná se o tabulky znázorňující vybrané oddíly ze školského výkazu o základních školách, které obsahují sumární hodnoty za všechny základní školy. Pohled na prezentovaná data však lze jednoduše změnit pomocí různých filtrů. Z hodnot tabulek lze rovněž jednoduše vytvářet různé typy grafů.



## 9.7 Přínosy navrženého řešení

Přínosy navrženého BI/DWH řešení lze shrnout několika následujícími body:

- Propojením základních datových zdrojů lze komplexně hodnotit jednotlivé druhy škol a školských zařízení vzhledem k aspektům jejich výkonnosti a kapacitního vytížení.
- Poskytování relevantních informací pro rozhodování o optimalizaci sítě škol.
- Informační podpora při sestavování dlouhodobých záměrů vzdělávání a rozvoje výchovně vzdělávací soustavy v regionu.
- Možnost vybudování strukturovaného systému reportingu škol a školských zařízení v dlouhodobém horizontu spojený s predikcí dalšího vývoje.
  - Mateřské školy
  - Základní školy
  - Základní umělecké školy
  - Školní družiny
  - Školní jídelny
  - Střediska volného času
- Třídění informací dle typu zřizovatele (obecní, církevní, soukromý) s možností benchmarkingu.
- Vysoký potenciál možnosti využití dat za obecní školství v rámci ORP většími městy.
- Možnost podrobných analýz o výkonnosti škol a školských zařízení.
- Objektívni podklady pro prezentaci informací o školství pro interní i externí subjekty.

## 10 Závěry a doporučení

V souvislosti s neustále se měnící ekonomickou situací a způsobem, jakým se zvětšuje objem firemních dat, je mnohem více společností nuceno využívat stále větší počet informačních systémů pro podporu jejich rozhodování. Současně vzniká potřeba z podnikových dat získávat potenciálně využitelné informace pro rozvoj společnosti. Jednou z oblastí informačních systémů, které toto umožňují, je Business Intelligence a její důležitá součást, datový sklad. Významnost Business Intelligence tak v dnešní době stále roste, jelikož čím dál více společností spatřuje v těchto nástrojích konkurenční výhodu a možnost, jak lépe a efektivněji řídit podnik.

Cílem diplomové práce bylo zejména navrhnout BI/DWH řešení pro oblast školství v Hradci Králové a následně provést implementaci pilotního projektu tohoto řešení. Oblast školství je z pohledu Business Intelligence velmi zajímavé téma, jelikož poskytuje velké množství statistických údajů, které je možné různými způsoby analyzovat. Pro splnění stanoveného cíle však bylo nutné nejprve prozkoumat teoretické podklady týkající se Business Intelligence a datových skladů.

V rámci teoretické části práce tak byla představena problematika Business Intelligence ve spojení s datovým skladem, která zahrnovala popis architektury a hlavních komponent obou oblastí. Rovněž bylo popsáno dimenzionální modelování, tvorba ETL procesů a analýza OLAP, jakožto nezbytné součásti celého systému.

Praktická část práce se zabývala především návrhem a následně pilotní implementací již zmíněného BI/DWH řešení pro oblast školství v Hradci Králové. K tomu však bylo zapotřebí provést analýzu zkoumané oblasti a datových zdrojů a na základě zjištěných poznatků stanovit požadavky na nově vzniklé řešení. Teprve potom mohl být proveden samotný návrh řešení, v rámci kterého byl sestaven popis datových zdrojů, vytvořen dimenzionální model a určena architektura výsledného řešení. Ta byla popsána z funkčního pohledu všech vrstev, tedy datové, transformační, analytické i prezentační vrstvy. Po návrhu následovala implementace pilotního projektu, která zahrnovala realizaci datové vrstvy v prostředí

MS SQL Server Management Studio 2008 a realizaci datových pump a OLAP kostky v prostředí MS Visual Studio 2008. Na konci praktické části byl rovněž představen návrh uživatelských reportů společně s napojením na datovou kostku v prostředí všeobecně známé aplikace MS Excel.

Výsledné BI/DWH řešení umožňuje ukládat a zpracovávat statistické údaje ze školských výkazů, jejichž dílčí součástí tvoří data z oblasti regionálního školství. Jelikož jsou data sbírána každoročně, lze tyto informace vyhodnocovat v průběhu času a modelovat trend očekávaného vývoje. Současně lze informace smysluplně matematicky propojovat, například vypočítávat počet žáků na učitele, vývoj průměrného počtu žáků ve třídě či v oboru a podobně. V datovém skladu můžeme vypočítávat i takové údaje, které nejsou ve výkazech implicitně uváděny, například vedle počtu dívek také počty chlapců, vypočítávat rozdílové hodnoty mezi celkovým údajem a dílčími údaji atd. Všechny tyto informace pak lze vyhodnocovat jak za jednotlivé výkazy, tak i v celkových sumacích.

V rámci praktické aplikace je zcela jistě možné rozšíření zkoumané oblasti z obecní na krajskou úroveň, což by umožnilo sledovat výkonové výkazy i pro střední, vyšší odborné a vysoké školy. Vedle školských výkazů by město Hradec Králové rovněž mohlo do datového skladu ukládat finanční výkazy jednotlivých škol a školských organizací. Datový sklad by pak umožňoval vyhodnocovat informace z trojí perspektivy – výkonové, mzdové a finanční. Ve srovnatelných typech škol by se také dal realizovat benchmarking údajů jak z oblasti výkonů, tak z oblasti práce a mezd, a to jak v reálném čase, tak i v historii.

Závěrem lze říci, že oblasti Business Intelligence a datového skladu mají opravdu značný význam, a to nejen pro velké podniky, ale i pro ty středních a menších rozměrů, pro které je práce se znalostmi získaných z jejich dat rovněž důležitá. Nezáleží přitom, zda se jedná o podnikovou nebo veřejnou sféru. V dnešní době si už téměř každý začíná uvědomovat hodnotu informací, snaží se data určitým způsobem shromažďovat a využít jejich informační potenciál ve svůj prospěch ve formě znalostí vedoucích ke správným firemním rozhodnutím.

## 11 Seznam použité literatury

- [1] NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. *Business intelligence: jak využít bohatství ve vašich datech*. Praha: Grada, 2005. *Management v informační společnosti*. ISBN 80-247-1094-3.
- [2] LACKO, Luboslav. *Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby*. Brno: Computer Press, 2009. ISBN 978-80-251-2887-9.
- [3] RUD, Olivia Parr. *Business intelligence success factors: tools for aligning your business in the global economy*. Hoboken, N.J.: Wiley, c2009. *Wiley and SAS business series*. ISBN 978-0-470-39240-9.
- [4] LABERGE, Robert. *Datové sklady: agilní metody a business intelligence*. Brno: Computer Press, 2012. ISBN 978-80-251-3729-1.
- [5] GÁLA, Libor, Jan POUR a Zuzana ŠEDIVÁ. *Podniková informatika: počítačové aplikace v podnikové a mezipodnikové praxi. 3., aktualizované vydání*. Praha: Grada Publishing, 2015. *Management v informační společnosti*. ISBN 978-80-247-5457-4.
- [6] LOSHIN, David. *Business intelligence: the savvy manager's guide, getting onboard with emerging IT*. Boston: Morgan Kaufmann Publishers, c2003. ISBN 1-558-60916-4.
- [7] POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. *Business intelligence v podnikové praxi*. Praha: Professional Publishing, 2012. ISBN 978-80-7431-065-2.
- [8] ONG, In, Pei SIEW a Siew WONG. *A Five-Layered Business Intelligence Architecture*. *Communications of the IBIMA* [online]. 2011, , 1-11 [cit. 2018-02-25]. DOI: 10.5171/2011.695619. ISSN 19437765. Dostupné z: <http://www.ibimapublishing.com/journals/CIBIMA/2011/695619/695619.html>
- [9] KIMBALL, Ralph a Joe CASERTA. *The data warehouse ETL toolkit practical techniques for extracting, cleaning, conforming, and delivering data* [online]. Indianapolis, IN: Wiley, 2004 [cit. 2018-02-25]. ISBN 07-645-7923-1. Dostupné z: <http://users.itk.ppke.hu/~szoer/DW/Kimball%20&%20Caserta%20-The%20Data%20Warehouse%20ETL%20Toolkit%20%5BWiley%202004%5D.pdf>
- [10] GÁLA, Libor, Jan POUR a Prokop TOMAN. *Podniková informatika: počítačové aplikace v podnikové a mezipodnikové praxi, technologie informačních systémů, řízení a rozvoj podnikové informatiky*. Praha: Grada, 2006. *Management v informační společnosti*. ISBN 80-247-1278-4.
- [11] INMON, William H. *Building the data warehouse*. 4th ed. Indianapolis, Ind.: Wiley, c2005. ISBN 978-0-7645-9944-6.
- [12] *Trendy Business Intelligence* [online]. 2016 [cit. 2018-02-26]. Dostupné z: <https://www.noctuint.cz/blog/2016-12-15-trendy-bi>
- [13] *About BARC – Business Application Research Center* [online]. 2016 [cit. 2018-02-26]. Dostupné z: <http://barc-research.com/about-us/>
- [14] *BI Trend Monitor 2018: The World's Largest Survey on BI Trends* [online]. Würzburg: BARC, 2017 [cit. 2018-02-26]. Dostupné z: [http://barc-research.com/wp-content/uploads/2017/11/BARC-BI\\_Trend\\_Monitor\\_2018-Online.pdf](http://barc-research.com/wp-content/uploads/2017/11/BARC-BI_Trend_Monitor_2018-Online.pdf)
- [15] *10 Business Intelligence Trends for 2016* [online]. PCMag, 2015 [cit. 2018-02-26]. Dostupné z: <https://www.pcmag.com/article2/0,2817,2496370,00.asp>
- [16] *Kmenová data (Master Data)* [online]. *Management Mania*, 2016 [cit. 2018-02-26]. Dostupné z: <https://managementmania.com/cs/kmenova-data-master-data>
- [17] STOKLÁSKA, Ondřej. *Self Service BI: Když víte jaká data potřebujete, je to tak jednoduché* [online]. 2015 [cit. 2018-02-26]. Dostupné z: <https://www.linkedin.com/pulse/self-service-bi-kdy%C5%BE-v%C3%ADte-jak%C3%A1-data-pot%C5%99ebujete-je-tak-stoklaska>

- [18] POUR, Jan. *Self-service business intelligence. Systémová integrace* [Online]. 2014, 1-2. [cit. 2018-02-26]. ISSN 1804-2716. Dostupné z: [http://www.cssi.cz/cssi/system/files/all/SI\\_2014\\_1-2\\_8\\_Pour.pdf](http://www.cssi.cz/cssi/system/files/all/SI_2014_1-2_8_Pour.pdf)
- [19] BÉM, Martin. *Self-service BI: REPORT RYCHLE? UDĚLEJ SI HO SÁM!* [online]. 2016 [cit. 2018-02-26]. Dostupné z: <http://www.adastra.cz/ict-reseni/self-service-bi>
- [20] ČERNÝ, Michal. *Vizualizace dat: Jak odhalit utajené souvislosti* [online]. 2016 [cit. 2018-02-26]. Dostupné z: <http://vtm.e15.cz/vizualizace-dat-jak-odhalit-utajene-souvislosti>
- [21] KYJONKA, Vladimír. *Vizualizace v business intelligence. IT systems* [Online]. 2013, 6. [cit. 2018-02-26]. ISSN 1802-615X. Dostupné z: <https://www.systemonline.cz/clanky/vizualizace-v-business-intelligence.htm>
- [22] Gartner: *Magic Quadrant for Business Intelligence and Analytics Platforms* [online]. Gartner, 2017 [cit. 2018-02-28]. Dostupné z: <https://www.gartner.com/doc/reprints?id=1-3TYE0CD&ct=170221&st=sb>
- [23] ŽÁK, Čestmír a Pavel KREUZIGER. *Jak fungují magické kvadranty?. Inside* [online]. 2013, 17. 5. 2013, 14(1), 1 [cit. 2018-02-28]. Dostupné z: <https://kpc-group.cz/wp-content/uploads/2013/07/magicke-kvadranty.pdf>
- [24] *InterSystems je lídrem magického kvadrantu společnosti Gartner* [online]. iDNES.cz, 2015 [cit. 2018-02-28]. Dostupné z: [https://sdeleni.idnes.cz/intersystems-lidrem-magickeho-kvadrantu-spolecnosti-gartner-p68-/eko-sdeleni.aspx?c=A150203\\_114348\\_eko-sdeleni\\_ahr](https://sdeleni.idnes.cz/intersystems-lidrem-magickeho-kvadrantu-spolecnosti-gartner-p68-/eko-sdeleni.aspx?c=A150203_114348_eko-sdeleni_ahr)
- [25] *Gartner označil Qlik za jednoho z lídrů moderního BI už šestkrát po sobě* [online]. Emark, 2016 [cit. 2018-02-28]. Dostupné z: <http://www.emark.sk/cz/aktuality-1/emark-newsletter-brezen/gartner-oznacil-qlik-za-jednoho-z-lidru-moderniho-bi-uz-sestkrat-po-sobe>
- [26] *Gartner: How Markets and Vendors Are Evaluated in Gartner Magic Quadrants* [online]. Gartner, 2017 [cit. 2018-02-28]. Dostupné z: <https://www.gartner.com/doc/3188318?ref=SiteSearch&stkw=magic%20quadrant&fnl=search&srcId=1-3478922254>
- [27] LACKO, Luboslav. *Databáze: datové sklady, OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle*. Brno: Computer Press, 2003. ISBN 80-7226-969-0.
- [28] VAVRUŠKA, Jindřich. *ETL a kvalita dat. SystemOnLine* [online]. Adastra [cit. 2018-03-05]. Dostupné z: <https://www.systemonline.cz/clanky/etl-a-kvalita-dat.htm>
- [29] KIMBALL, Ralph a Margy ROSS. *The data warehouse toolkit: the definitive guide to dimensional modeling. Third edition*. Indianapolis, IN: John Wiley & Sons, 2013. ISBN 978-1-118-53080-1.
- [30] *Inmon vs. Kimball - An Analysis*. Nagesh [online]. Denver: Nagesh, 2005 [cit. 2018-03-20]. Dostupné z: <http://www.nagesh.com/publications/technology/173-inmon-vs-kimball-an-analysis.html>
- [31] KIMBALL, Ralph. a Margy ROSS. *The data warehouse toolkit: the complete guide to dimensional modeling. 2nd ed*. New York: Wiley, c2002. ISBN 04-712-0024-7.
- [32] INMON, William H. *Building the data warehouse. 4th ed*. Indianapolis, Ind.: Wiley, c2005. ISBN 07-645-9944-5.
- [33] RAINARDI, Vincent. *Building a data warehouse with examples in SQL Server*. Berkeley, CA: Apress ; Distributed to the book trade worldwide by Springer-Verlag New York, c2008. ISBN 978-1-59059-931-0.
- [34] MUNDY, Joy, Warren THORNTHWAITE a Ralph KIMBALL. *The Microsoft data warehouse toolkit: with SQL server 2008 R2 and the Microsoft Business intelligence toolset. 2nd ed*. Indianapolis, IN: Wiley Pub., c2011. ISBN 978-0-470-64038-8.

- [35] PANDRE, Andrei. *OLAP Cubes: Data Visualization*. OLAP[online]. Massachusetts: APandre, c2011 [cit. 2018-03-26]. Dostupné z: <https://apandre.wordpress.com/data/datacloud/datacube/>
- [36] *Overview of Online Analytical Processing (OLAP)*. OLAP [online]. Microsoft, c2018 [cit. 2018-03-26]. Dostupné z: <https://support.office.com/en-us/article/Overview-of-Online-Analytical-Processing-OLAP-15d2cdde-f70b-4277-b009-ed732b75fd6?ui=en-US&rs=en-US&ad=US>.
- [37] THOMSEN, Erik. *OLAP technology: building multidimensional system*. 2nd ed. New York: Wiley Computer Pub., c2002. ISBN 0471400300.
- [38] ČECH, Pavel a Vladimír BUREŠ. *Podniková informatika*. Hradec Králové: Gaudeamus, 2009. ISBN 978-80-7041-479-8.

## 12 Přílohy

- 1) Ukázka první strany výkazu o základní škole
- 2) Tabulky dimenzí
- 3) Tabulky faktů
- 4) Dimenzionální schéma se třemi tabulkami faktů
- 5) OLAP kostka Školství
- 6) Filtry a kontingenční tabulky pro výkaz o základních školách

## MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Povinnost předávat údaje stanoví § 28 odst. 5 zákona č. 561/2004 Sb., ve znění pozdějších předpisů.

Právnícká osoba předá individuální údaje elektronicky do 16. 10. 2017 na server podle pokynů MŠMT. Z předaných údajů se vytváří tento výkaz.

M 3

**VÝKAZ**  
**o základní škole**  
podle stavu k 30. 9. 2017

Škola \_\_\_\_\_

Obec \_\_\_\_\_ PSČ \_\_\_\_\_

Ulice \_\_\_\_\_ Čp. \_\_\_\_\_

Správní úřad \_\_\_\_\_

Resortní identifikátor právnické osoby (RED_IZO)		Resortní identifikátor školy (IZO)		
Území	Zřizovatel	Škola podle § 16 odst. 9	Organizace školy	Škola mimo provoz

## I. Počet žáků, kteří odešli ze základní školy

	Číslo řádku	Běžné třídy		Speciální třídy	
		Celkem	z toho dívky	Celkem	z toho dívky
a	b	2	3	4	5
Žáci, kteří ukončili povinnou školní docházku	0101				
z toho					
v 7. ročníku	0104				
v 8. ročníku	0105				
v 9. ročníku	0106				
v 10. ročníku	0107				
nezařazení do ročníku	0107a				
Žáci, kteří přešli do středních škol	0108				
z toho					
z 5. ročníku	0109				
ze 7. ročníku	0110				
Žáci 1. ročníku s dodatečným odkladem	0111				

## Dodatek za školní rok 2016/17

Součástí výkazu jsou **Vysvětlivky k vyplnění**, kde jsou podrobně vyloženy pojmy používané v tomto formuláři a popsáno, co je zahrnuto v řádcích a sloupcích jednotlivých oddílů.

Výklad pojmů „speciální třídy“ a „běžné třídy“ je uveden v části Společné poznámky, písm. b).

Ř. 0108 - 0110: Žáci, kteří pokračují v plnění povinné školní docházky ve víceletých gymnáziích a v 8letém vzdělávacím programu na konzervatoři.

Ř. 0111: Jen dodatečné odklady povinné školní docházky po zahájení minulého školního roku.

Ř. 0107a: Žáci, kteří se vzdělávali podle § 38, odst. 1, písm. a), u nichž nebylo možné zjistit uroveň dosavadního vzdělání, a nebyli zařazeni do ročníku.

## Školní rok 2017/18

## III. Třídy a žáci podle ročníků – běžné třídy

	Číslo řádku	Počet tříd	Počet žáků		z toho žáci se zdrav. postiž. <sup>1)</sup>		z celku žáci opakující ročník		z celku převedení do vyššího ročníku		z celku vzdělávající se podle § 50 odst. 3	
			celkem	dívky	celkem	dívky	celkem	dívky	celkem	dívky	celkem	dívky
a	b	2	3	4	5	6	7	8	9	10	10a	10b
Celkem	0301											
z toho 1. stupeň	0302											
v tom	1. ročník	0303							X	X		
	2. ročník	0304										
	3. ročník	0305										
	4. ročník	0306										
	5. ročník	0307										
	6. ročník	0308										
	z toho 10letý VP	0309	X									
	7. ročník	0310										
	8. ročník	0311										
	9. ročník	0312										
z toho 10letý VP	0313	X										
10. ročník	0314	X										

<sup>1)</sup> Uvedeni žáci se zdravotním znevýhodněním uvedeným v § 16 odst. 9 školského zákona, u nichž byly speciální vzdělávací potřeby zjištěny na základě speciálně pedagogického a psychologického vyšetření školským poradenským zařízením, vzdělávající se v běžných třídách (viz též odd. IX).

U 6. a 9. ročníku jsou v ř. 0309 a 0313 vyčleněni žáci, kteří se vzdělávají v 10letém vzdělávacím programu.

Sl. 9 - 10: Uvedeni pouze žáci mimořádně nadaní, kteří byli v rámci téhož vzdělávacího programu (na základě vykonané zkoušky) převedeni do vyššího ročníku bez absolvování ročníku předchozího.



## Dimenze Čas

Atribut	Popis	Řazení
<b>Den</b>	Den	A-Z
<b>Měsíc</b>	Měsíc	A-Z
<b>Kvartál</b>	Kvartál	A-Z
<b>Rok</b>	Rok	A-Z
<b>Školní rok</b>	Školní rok	A-Z

## Dimenze Ukazatel školských výkazů

Atribut	Popis	Řazení
<b>Kód</b>	Kód ukazatele školských výkazů (např. Rxxxx)	A-Z
<b>Název</b>	Název ukazatele školských výkazů	A-Z
<b>Výkaz</b>	Výkaz, ke kterému ukazatel přísluší	A-Z
<b>Oddíl</b>	Oddíl výkazu	A-Z
<b>Sloupec</b>	Název ukazatele ve sloupci	A-Z
<b>Řádek</b>	Název ukazatele v řádku	A-Z

## Dimenze Druh školního zařízení

Atribut	Popis	Řazení
<b>Kód</b>	Kód druhu školního zařízení	A-Z
<b>Název</b>	Název druhu školního zařízení	A-Z

## Dimenze Speciální vzdělávací potřeby

Atribut	Popis	Řazení
<b>Kód</b>	Kód speciálních vzdělávacích potřeb	A-Z
<b>Název</b>	Název speciálních vzdělávacích potřeb	A-Z

## Dimenze Státní občanství

Atribut	Popis	Řazení
<b>Kód</b>	Kód státního občanství	A-Z
<b>Název</b>	Název státního občanství	A-Z

## Dimenze Vyučovací jazyk

Atribut	Popis	Řazení
<b>Kód</b>	Kód vyučovacího jazyku	A-Z
<b>Název</b>	Název vyučovacího jazyku	A-Z

## Dimenze Ředitelství

Atribut	Popis	Řazení
<b>IZO</b>	IZO ředitelství	A-Z
<b>Název</b>	Název ředitelství	A-Z
<b>IČO</b>	IČO	A-Z
<b>Typ zřizovatele</b>	Kód typu zřizovatele	A-Z
<b>Ředitel</b>	Jméno ředitele	
<b>ORP</b>	ORP	A-Z
<b>Okres</b>	Okres	A-Z
<b>Ulice</b>	Ulice	A-Z
<b>ČP</b>	Číslo popisné	A-Z
<b>Místo</b>	Místo	A-Z
<b>Město</b>	Město	A-Z
<b>PSC</b>	PSC	A-Z
<b>Telefon</b>	Telefon	
<b>Správní úřad kód</b>	Kód správního úřadu	A-Z
<b>Správní úřad název</b>	Název správního úřadu	A-Z

## Dimenze Školy - Součásti

<b>Atribut</b>	<b>Popis</b>	<b>Řazení</b>
<b>IZO ředitelství</b>	IZO ředitelství	A-Z
<b>Název ředitelství</b>	Název ředitelství	A-Z
<b>IČO</b>	IČO	
<b>IZO</b>	IZO součásti	A-Z
<b>Název</b>	Název součásti	A-Z
<b>Typ zřizovatele</b>	Kód typu zřizovatele	A-Z
<b>ORP</b>	ORP	A-Z
<b>Okres</b>	Okres	A-Z
<b>Ulice</b>	Ulice	A-Z
<b>ČP</b>	Číslo popisné	A-Z
<b>Místo</b>	Místo	A-Z
<b>Město</b>	Město	A-Z
<b>PSČ</b>	PSČ	A-Z
<b>Telefon</b>	Telefon	A-Z
<b>Správní úřad kód</b>	Kód správního úřadu	A-Z
<b>Správní úřad název</b>	Název správního úřadu	A-Z
<b>Druh PAM</b>	Druh školního zařízení	A-Z
<b>Druh provozu školy</b>	Druh provozu školy	
<b>Organizace školy</b>	Organizace školy	A-Z
<b>Typ kód</b>	Kód typu školy, zařízení	
<b>Typ název</b>	Název typu školy, zařízení	
<b>Kapacita cílová</b>	Cílová kapacita	A-Z
<b>Kapacita aktuální</b>	Aktuální kapacita	A-Z

## Dimenze Školy - Místo

<b>Atribut</b>	<b>Popis</b>	<b>Řazení</b>
<b>IZO ředitelství</b>	IZO ředitelství	A-Z
<b>IZO součásti</b>	IZO součásti	
<b>IZO místa</b>	IZO místa	A-Z
<b>Název</b>	Název součásti	A-Z
<b>Typ zřizovatele</b>	Kód typu zřizovatele	A-Z
<b>ORP</b>	ORP	A-Z
<b>Okres</b>	Okres	A-Z
<b>Ulice</b>	Ulice	A-Z
<b>ČP</b>	Číslo popisné	A-Z
<b>Místo</b>	Místo	A-Z
<b>Město</b>	Město	A-Z
<b>PSČ</b>	PSČ	A-Z
<b>Typ kód</b>	Kód typu školy, zařízení	
<b>Typ název</b>	Název typu školy, zařízení	

**Fakta Výkaz V01 – Výkaz o mateřské škole**

<b>Atribut</b>	<b>Popis</b>
<b>Čas SK</b>	SK klíč do tabulky dimenze Čas
<b>IZO SK</b>	SK klíč do tabulky dimenze Školy - Součásti
<b>IZO místa SK</b>	SK klíč do tabulky dimenze Školy - Místo
<b>IZO ředitelství SK</b>	SK klíč do tabulky dimenze Ředitelství
<b>Stát SK</b>	SK klíč do tabulky dimenze Státní občanství
<b>SVP SK</b>	SK klíč do tabulky dimenze Speciální vzdělávací potřeby
<b>Ukazatel SK</b>	SK klíč do tabulky dimenze Ukazatel školských výkazů (ukazatel určující faktický údaj)
<b>Hodnota</b>	Sledovaný faktický údaj určený dle daného ukazatele

**Fakta Výkaz V02 – Výkaz o školní družině, školním klubu**

<b>Atribut</b>	<b>Popis</b>
<b>Čas SK</b>	SK klíč do tabulky dimenze Čas
<b>IZO SK</b>	SK klíč do tabulky dimenze Školy - Součásti
<b>IZO místa SK</b>	SK klíč do tabulky dimenze Školy - Místo
<b>IZO ředitelství SK</b>	SK klíč do tabulky dimenze Ředitelství
<b>Stát SK</b>	SK klíč do tabulky dimenze Státní občanství
<b>SVP SK</b>	SK klíč do tabulky dimenze Speciální vzdělávací potřeby
<b>Ukazatel SK</b>	SK klíč do tabulky dimenze Ukazatel školských výkazů (ukazatel určující faktický údaj)
<b>Hodnota</b>	Sledovaný faktický údaj určený dle daného ukazatele

**Fakta Výkaz V03 – Výkaz o základní škole**

<b>Atribut</b>	<b>Popis</b>
<b>Čas SK</b>	SK klíč do tabulky dimenze Čas
<b>IZO SK</b>	SK klíč do tabulky dimenze Školy - Součásti
<b>IZO místa SK</b>	SK klíč do tabulky dimenze Školy - Místo
<b>IZO ředitelství SK</b>	SK klíč do tabulky dimenze Ředitelství
<b>Stát SK</b>	SK klíč do tabulky dimenze Státní občanství
<b>SVP SK</b>	SK klíč do tabulky dimenze Speciální vzdělávací potřeby
<b>Ukazatel SK</b>	SK klíč do tabulky dimenze Ukazatel školských výkazů (ukazatel určující faktický údaj)
<b>Vyučovací jazyk SK</b>	SK klíč do tabulky dimenze Vyučovací jazyk
<b>Hodnota</b>	Sledovaný faktický údaj určený dle daného ukazatele

**Fakta Výkaz V04c – Výkaz o přípravné třídě ZŠ**

<b>Atribut</b>	<b>Popis</b>
<b>Čas SK</b>	SK klíč do tabulky dimenze Čas
<b>IZO SK</b>	SK klíč do tabulky dimenze Školy - Součásti
<b>IZO místa SK</b>	SK klíč do tabulky dimenze Školy - Místo
<b>IZO ředitelství SK</b>	SK klíč do tabulky dimenze Ředitelství
<b>Stát SK</b>	SK klíč do tabulky dimenze Státní občanství
<b>Ukazatel SK</b>	SK klíč do tabulky dimenze Ukazatel školských výkazů (ukazatel určující faktický údaj)
<b>Hodnota</b>	Sledovaný faktický údaj určený dle daného ukazatele

**Fakta Výkaz V13 – Výkaz o ředitelství škol**

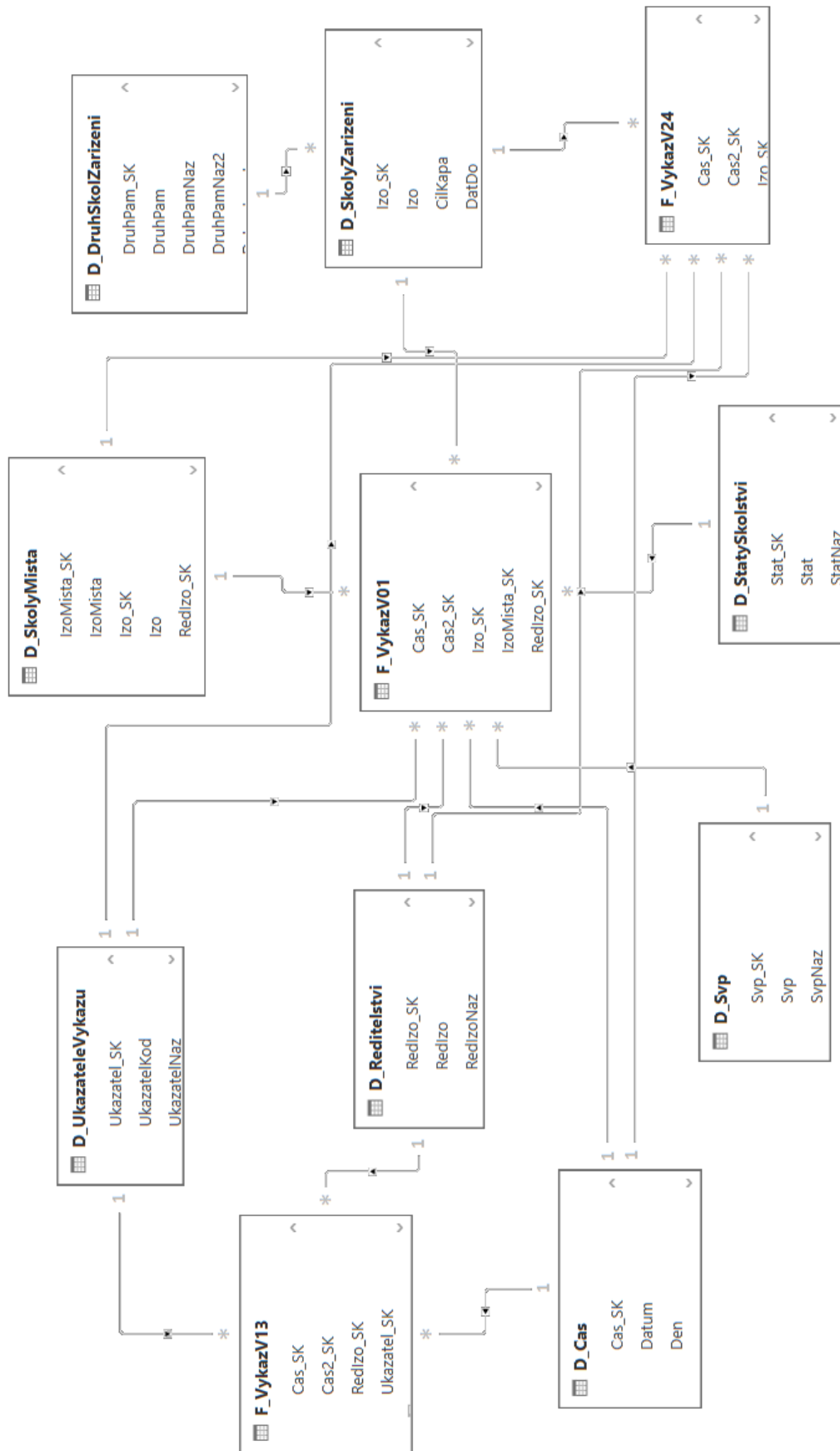
<b>Atribut</b>	<b>Popis</b>
<b>Čas SK</b>	SK klíč do tabulky dimenze Čas
<b>IZO ředitelství SK</b>	SK klíč do tabulky dimenze Ředitelství
<b>Ukazatel SK</b>	SK klíč do tabulky dimenze Ukazatel školských výkazů (ukazatel určující faktický údaj)
<b>Hodnota</b>	Sledovaný faktický údaj určený dle daného ukazatele

**Fakta Výkaz V24 – Výkaz o základní umělecké škole**

<b>Atribut</b>	<b>Popis</b>
<b>Čas SK</b>	SK klíč do tabulky dimenze Čas
<b>IZO SK</b>	SK klíč do tabulky dimenze Školy - Součásti
<b>IZO místa SK</b>	SK klíč do tabulky dimenze Školy - Místo
<b>IZO ředitelství SK</b>	SK klíč do tabulky dimenze Ředitelství
<b>Ukazatel SK</b>	SK klíč do tabulky dimenze Ukazatel školských výkazů (ukazatel určující faktický údaj)
<b>Hodnota</b>	Sledovaný faktický údaj určený dle daného ukazatele

**Fakta Výkaz V53 – Výkaz o zahájení povinné školní docházky v ZŠ**

<b>Atribut</b>	<b>Popis</b>
<b>Čas SK</b>	SK klíč do tabulky dimenze Čas
<b>IZO SK</b>	SK klíč do tabulky dimenze Školy - Součásti
<b>IZO místa SK</b>	SK klíč do tabulky dimenze Školy - Místo
<b>IZO ředitelství SK</b>	SK klíč do tabulky dimenze Ředitelství
<b>Ukazatel SK</b>	SK klíč do tabulky dimenze Ukazatel školských výkazů (ukazatel určující faktický údaj)
<b>Hodnota</b>	Sledovaný faktický údaj určený dle daného ukazatele





<b>Skupina metrik</b>	<b>Typ</b>	<b>Objekt v OLAP</b>	<b>Agregace</b>
Výkaz V01	Dimenze	Čas	-
Výkaz V01	Dimenze	Školy-Ředitelství	-
Výkaz V01	Dimenze	Školy-Součásti	-
Výkaz V01	Dimenze	Školy-Místo	-
Výkaz V01	Dimenze	Státní občanství	-
Výkaz V01	Dimenze	Ukazatel školských výkazů	-
Výkaz V01	Proměnná	Hodnota	ANO
Výkaz V02	Dimenze	Čas	-
Výkaz V02	Dimenze	Školy-Ředitelství	-
Výkaz V02	Dimenze	Školy-Součásti	-
Výkaz V02	Dimenze	Školy-Místo	-
Výkaz V02	Dimenze	Státní občanství	-
Výkaz V02	Dimenze	Ukazatel školských výkazů	-
Výkaz V02	Proměnná	Hodnota	ANO
Výkaz V03	Dimenze	Čas	-
Výkaz V03	Dimenze	Školy-Ředitelství	-
Výkaz V03	Dimenze	Školy-Součásti	-
Výkaz V03	Dimenze	Školy-Místo	-
Výkaz V03	Dimenze	Státní občanství	-
Výkaz V03	Dimenze	Vyučovací jazyk oboru	-
Výkaz V03	Dimenze	Ukazatel školských výkazů	-
Výkaz V03	Proměnná	Hodnota	ANO
Výkaz V13	Dimenze	Čas	-
Výkaz V13	Dimenze	Školy-Ředitelství	-
Výkaz V13	Dimenze	Ukazatel školských výkazů	-
Výkaz V13	Proměnná	Hodnota	ANO
Výkaz V24	Dimenze	Čas	-
Výkaz V24	Dimenze	Školy-Ředitelství	-
Výkaz V24	Dimenze	Školy-Součásti	-

<b>Výkaz V24</b>	Dimenze	Školy-Místo	-
<b>Výkaz V24</b>	Dimenze	Obor	-
<b>Výkaz V24</b>	Dimenze	Státní občanství	-
<b>Výkaz V24</b>	Dimenze	Ukazatel školských výkazů	-
<b>Výkaz V24</b>	Proměnná	Hodnota	ANO
<b>Výkaz V4c</b>	Dimenze	Čas	-
<b>Výkaz V4c</b>	Dimenze	Školy-Ředitelství	-
<b>Výkaz V4c</b>	Dimenze	Školy-Součásti	-
<b>Výkaz V4c</b>	Dimenze	Školy-Místo	-
<b>Výkaz V4c</b>	Dimenze	Ukazatel školských výkazů	-
<b>Výkaz V4c</b>	Proměnná	Hodnota	ANO
<b>Výkaz V53</b>	Dimenze	Čas	-
<b>Výkaz V53</b>	Dimenze	Školy-Ředitelství	-
<b>Výkaz V53</b>	Dimenze	Školy-Součásti	-
<b>Výkaz V53</b>	Dimenze	Školy-Místo	-
<b>Výkaz V53</b>	Dimenze	Ukazatel školských výkazů	-
<b>Výkaz V53</b>	Proměnná	Hodnota	ANO

**Rok**

2006 2007 2008 2009 2010 2011 2012 2013

**Ředitelství - název**

Základní škola a Mateřská škola, Hradec Králové - Kukleny, Pražská 198

Základní škola a Mateřská škola, Hradec Králové - Malšova Lhota, Lhotec

Základní škola a Mateřská škola, Hradec Králové - Svobodné Dvory, Spojc

Základní škola a Mateřská škola, Hradec Králové, Jiráskovo nám. 1166

Základní škola a Mateřská škola, Hradec Králové, Štefcova 1092

Základní škola a Mateřská škola, Hradec Králové, Úprkova 1

Základní škola a mateřská škola, Kratonohy, okres Hradec Králové, příspě

**Typ zřizovatele původní - název**

Církev Kraj Obec Privátní ...

### I. Počet žáků, kteří odešli v min. škol. roce

Řádek	Sloupec	
	Běžné třídy-celkem	Speciální třídy-celkem
Žáci, kteří ukončili školní docházku	1 093	2
v tom v 1. - 5. ročníku		
v tom v 6. ročníku		
v tom v 7. ročníku	6	
v tom v 8. ročníku	15	
v tom v 9. ročníku	1 072	2
v tom v 10. ročníku		
Žáci, kteří přešli do středních škol	182	
z toho z 5. ročníku	44	
z toho ze 7. ročníku	137	
Žáci 1.r.s dod. odkladem šk.doch.	4	

### III. Třídy a žáci podle ročníků - běžné třídy

Řádek	Sloupec						Naplněnost tříd
	Počet tříd	Počet žáků - celkem	z toho postižení žáci - celkem	z celku žáci opakující - celkem	z celku převedení do vyššího ročníku - celkem	z celku vzdělávající se podle § 50 odst. 3 - celkem	
Celkem	553	11 878	870	43	2		21,48
... z toho 1.stupeň	343	7 215	395	25	2		21,03
v tom 1. ročník	80	1 575	10	10			19,69
v tom 2. ročník	74	1 545	14	4			20,88
v tom 3. ročník	69	1 468	81	4	1		21,28
v tom 4. ročník	61	1 316	123	3	1		21,57
v tom 5. ročník	59	1 311	167	4			22,22
v tom 6. ročník	54	1 233	135	9			22,83
v tom 7. ročník	51	1 174	125	2			23,02
v tom 8. ročník	52	1 099	110	6			21,13
v tom 9. ročník	53	1 157	105	1			21,83

## V. Žáci učící se cizí jazyk

Řádek	Sloupec	Běžné třídy - z toho povinné předměty	Běžné třídy - navíc: nepovinné předměty	Speciální třídy - celkem	Speciální třídy - z toho povinné předměty	Speciální třídy - navíc: nepovinné předměty
Celkem	9 859	9 859	95	25	25	
... anglický jazyk	9 859	9 859	52	25	25	
... franc. jazyk	91					
... německý jazyk	1 726		27			
... ruský jazyk	658		16			
... španělský jazyk	28					
... italský jazyk						
... jiný evropský jazyk						
... jiný jazyk						

## XXIV. Věkové složení žáků

Řádek	Sloupec	běžné třídy - z celku nově přijatí do 1.r.celkem	spec. třídy - počet celkem	spec. třídy - z celku nově přijatí do 1.r.celkem
06 let a méně	9	9		
07 let	1 310	1 293	7	7
08 let	1 498	261	14	2
09 let	1 479	2	3	
10 let	1 341		3	
11 let	1 292		1	
12 let	1 225		2	
13 let	1 204		3	
14 let	1 082		1	
15 let	1 134		4	
16 let	292		2	
17 let	12			

Univerzita Hradec Králové  
Fakulta informatiky a managementu  
Akademický rok: 2017/2018

Studijní program: Aplikovaná informatika  
Forma: Prezenční  
Obor/komb.: Aplikovaná informatika (ai2-p)

**Podklad pro zadání DIPLOMOVÉ práce studenta**

PŘEDKLÁDÁ:	ADRESA	OSOBNÍ ČÍSLO
Bc. Sajvera Miroslav	Štefánikova 316/15, Hradec Králové - Moravské Předměstí	I1500696

**TÉMA ČESKY:**

Datový sklad a Business Intelligence pro použití ve veřejné správě

**TÉMA ANGLICKY:**

Data warehouse and Business Intelligence in public administration

**VEDOUcí PRÁCE:**

Ing. Zuzana Němcová, Ph.D. - KIT

**ZÁSADY PRO VYPRACOVÁNÍ:**

Cíl práce: Prozkoumat problematiku Business Intelligence se zaměřením na datové sklady, navrhnout BI/DWH řešení pro oblast školství v Hradci Králové a provést pilotní implementaci navrženého řešení.

Osnova:

1. Úvod
2. Business Intelligence
3. Aktuální trendy a analýza trhu BI
4. Dimenzionální modelování
5. Datový sklad
6. ETL
7. OLAP
8. Praktická část
9. Závěr

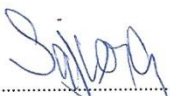
**SEZNAM DOPORUČENÉ LITERATURY:**

NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. Business intelligence: jak využít bohatství ve vašich datech. Praha: Grada, 2005. Management v informační společnosti. ISBN 80-247-1094-3.

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

LABERGE, Robert. Datové sklady: agilní metody a BI. Brno: Computer Press, 2012. ISBN 978-80-251-3729-1.


Podpis studenta:

  
.....

Datum:

15.4.2018

Podpis vedoucího práce:

  
.....

Datum:

15.4.2018