

Univerzita Hradec Králové
Fakulta informatiky a managementu
Katedra informatiky a kvantitativních metod

Využití technologie OLAP v prostředí maloobchodní sítě
Návrh řešení a implementace
Bakalářská práce

Autor: Jaromír Šplíchal
Studijní obor: Aplikovaná informatika

Vedoucí práce : doc. Mgr. Tomáš Kozel, Ph.D.

Prohlášení:

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a s použitím uvedené literatury.

V Hradci Králové dne 11.10.2021

Jaromír Šplíchal

Poděkování

Rád bych poděkoval vedoucímu práce doc. Mgr. Tomáši Kozlovi, Ph.D. za rady a připomínky při tvorbě bakalářské práce.

Anotace

Tato bakalářská práce se zabývá možnostmi využití technologie OLAP v prostředí maloobchodní sítě. V práci jsou popsány motivace i důvody pro využití moderních řešení *Business Intelligence*, obvyklá architektura těchto řešení včetně charakteristiky role a funkcí analytických nástrojů *OLAP* a *Reporting*. Práce obsahuje popis prostředí konkrétní maloobchodní sítě a návrh implementaci systému *Business Intelligence* v tomto prostředí včetně kroků, které je nutné před implementací provést. Jsou zmíněny i některé potenciální problémy implementace. V závěru je krátce popsáno řešení, které bylo uvedeno do produkčního provozu.

Klíčová slova: OLAP, Business Intelligence, Reporting

Annotation

This Bachelor Thesis deals with the possibilities of using OLAP technology in the retail network environment. The thesis describes the motivations and reasons for the use of modern Business Intelligence solutions, the usual architecture of these solutions, including the characteristics of the role and functions of analytical tools OLAP and Reporting. The thesis contains a description of the environment of a specific retail network and a proposal for the implementation of the Business Intelligence system in this environment, including the steps that must be performed before implementation. Some potential implementation issues are also mentioned. In the end, the solution, which was put into production operation, is briefly described.

Keywords: OLAP, Business Intelligence, Reporting

Obsah

1. Úvod.....	1
2. Od transakčních databází k analytickým.....	3
2.1 Role a možnosti produkčních systémů.....	3
2.2 Business intelligence jako řešení.....	5
2.3 Datový sklad.....	6
2.4 Informační systém BI s datovým skladem.....	8
2.5 Extrakce, transformace a přenos.....	9
2.6 Analytické nástroje.....	10
2.6.1 OLAP.....	11
2.6.2 Schémata dimenzí.....	14
2.6.3 Dělení OLAP podle uložení dat.....	15
2.6.4 Reporting.....	16
3. Popis prostředí pro plánovanou implementaci BI.....	17
3.1 Struktura obchodního řetězce APC.....	17
3.2 Produkční systém.....	17
3.3 Jednotné číselníky.....	18
3.4 Výstupy, reporty, analýzy.....	19
3.5 Uživatelé.....	20
3.6 Konektivita.....	20
3.7 Sehrávání dat, aktualizace číselníků.....	20
4. Příprava na implementaci BI v APC.....	21
4.1 Sjednocení číselníků.....	21
4.1.1 Zavedení JSO.....	21
4.1.2 Zavedení COP.....	24
4.2 Mapování požadavků.....	26
4.3 Souběh starého a nového řešení.....	27
5. Volba řešení BI.....	28
5.1 Volba řešení z pohledu časového a organizačního.....	28
5.2 Volba řešení z pohledu architektury.....	29
5.3 Procesy ETL.....	29
5.3.1 Provedení ETL.....	30
5.3.2 Rozsah dat - časově.....	30

5.3.3 Rozsah dat – obsahově.....	30
5.3.4 Ochrana osobních a citlivých dat.....	30
5.4 Služba CENTDATA.....	31
5.5 Datový sklad.....	31
5.6 OLAP databáze.....	32
5.7 Reporting.....	32
5.8 Aplikace DOHLED.....	33
6. Návrh dimenzí a tabulky faktů.....	34
6.1 Návrh dimenzí „Prodejních transakcí“.....	34
6.2 Návrh tabulky faktů „Prodejních transakcí“.....	37
6.2.1 Granularita tabulky faktů.....	38
6.2.2 Zdroje dat v produkčním systému.....	38
6.2.3 Struktura tabulky faktů.....	38
6.3 Další řešené oblasti.....	39
7. Rizika a problémy implementace.....	41
8. Implementace, volba platformy a dodavatele, finální řešení.....	44
8.1 Úvod.....	44
8.2 Popis a základní charakteristika vybraných platforem.....	45
8.2.1 SQL Server Business Intelligence.....	45
8.2.2 Oracle Business Intelligence.....	45
8.2.3 QlikView.....	46
8.2.4 Porovnání, výběr platformy.....	46
8.3 Jak to všechno dopadlo, finální řešení implementace BI.....	47
8.3.1 Výběr finálního řešení a jeho realizace.....	48
8.3.2 Aplikace REPORTING.CZ.....	50
9. Shrnutí výsledků.....	53
10. Závěr.....	54
11. Seznam použité literatury.....	55
12. Seznam obrázků.....	57
13. Seznam tabulek.....	58
14. Zadání práce (kopie).....	59

1. Úvod

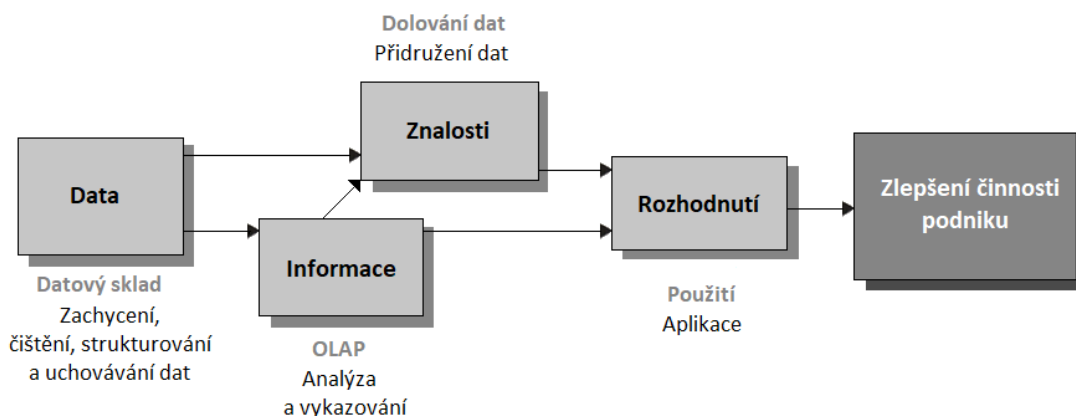
Nasazení informačních technologií v různých oblastech obchodní a výrobní činnosti v současných podmínkách vede zákonitě ke vzniku velkého množství dat a následné potřebě tato data shromažďovat a uchovávat. Jsou shromažďována různorodá data z různých částí a fází výrobních, obchodních a účetních procesů a agend. Procesy probíhající v moderním výrobním a obchodním prostředí mohou být realizovány nebo být podporovány na různých softwarových i hardwarových technologických platformách [4, s.14].

Současné technologie pro ukládání dat známé jako SQL servery umožňují proces ukládání dat spolehlivě zvládnout a to i s ohledem na různý způsob vzniku dat a i s ohledem na jejich velikost.

Pokud by práce se shromážděnými daty skončila jejich uložením a možností pouze si je kdykoli vyvolat a určitým způsobem je prezentovat (typicky ve formě nějakého přehledu nebo souhrnu), pak by zdánlivě byl naplněn účel tvorby úložiště dat, kdy data byla zaznamenána a lze je znovu více či méně efektivně nějakým způsobem přečíst a vyhodnotit. Takový způsob využití dat je vhodný spíše v situaci, kdy je třeba číst právě tato primární data, kdy může být účelem data prezentovat anebo rekonstruovat do původní podoby např. pro kontrolní činnost, obnovení dat ze záloh nebo přípravě určitého speciálního reportu nebo výstupu.

Současná úroveň vědění rozlišuje mezi pojmem údaj, informace a poznatek a formuluje procesy označované pojmem *Business Intelligence*, které transformují obvykle velké objemy údajů na informace a převádějí informace na poznatky a znalosti [3, s.19]. Tyto poznatky koncový uživatel s rozhodovací pravomocí využívá v procesech spojených s řízením a optimalizací své podnikatelské aktivity.

Obrázek 1 znázorňuje diagram toku, kde data se nacházejí na jedné straně a výsledný cíl ve formě zlepšení fungování podniku nebo organizace realizovaný správnými rozhodnutími na základě z dat získaných informací a znalostí na druhé straně.



Obr. 1 Diagram toku [2, s.272]

Důležitým aspektem v procesu využití dat je také to, aby si vlastník dat uvědomoval, že data má a dále si uvědomoval jejich potenciální význam. Lacko k tomu poznamenává [4, s.14]

„Data se stávají informacemi pokud:

- *máte data*
- *víte, že máte data*
- *víte, kde tyto data jsou uložena*
- *máme k datům přístup*
- *zdroji dat můžeme důvěřovat“*

Autor této práce na základě vlastních zkušeností k předchozí citaci poznamenává, že v současné době si pracovníci s rozhodovacími pravomocemi obvykle velmi dobře uvědomují, jaký význam a cenu data organizace mají. Bez kvalitních analytických výstupů je obtížné na úrovni středního i vyššího managementu složitější organizační strukturu řídit, zachytit vývojové trendy a přiměřeně na ně reagovat. Dalším příkladem využití analýz je oblast nákupu, kde je nutné, aby pracovníci, kteří jsou pověřeni nákupem zboží od dodavatelů, byli pro jednání o cenách dodávek vybaveni „spolehlivými“ výkazy s historií i predikcí prodejů sortimentních položek.

2. Od transakčních databází k analytickým

V současnosti téměř každý podnikající subjekt využívá při svých aktivitách nějaký typ softwarového systému, který mu umožňuje prováděné činnosti evidovat případně organizovat. Příkladem takového systému může být účetnictví, skladové hospodářství, fakturace, výrobní agendy, ale také agendy pro evidenci neziskových činností nebo práce a aktivit ve zdravotnictví, ve státní správě a dalších oblastech. V odborné literatuře označují tyto systémy různí autoři pojmem „*Transakční databáze*“ [9, s.22] nebo „*OLTP (On Line Transaction Processing)*“ [4, s.18] a také „*Produkční systémy*“ [6, s.28]. V tomto textu bude použito označení „*OLTP*“ a „*Produkční systémy*“.

Při evidenci sledovaných aktivit obvykle vzniká větší nebo velké množství dat a to dílem z pohledu dalšího analytické zpracování bezcenné, ale také velmi cenné, které ale zůstává bez využití, protože data nejsou uložena ve formě vhodné pro získávání informací. A tak může nastat situace, kdy údaje existují, ale řídicí pracovník se k informacím, které by mu pomohly činnost řízené organizace zlepšovat nebo optimalizovat, nemusí dostat. [3, s.21]

2.1 Role a možnosti produkčních systémů

Oblast využití transakčních databází je velmi pestrá, téměř neomezená, a odpovídá širokému spektru lidských činností a aktivit. Lze konstatovat, že ke každé lidské činnosti nebo aktivitě je možné vytvořit databázi OLTP, která bude průběh této činnosti nebo aktivity v určité míře detailu zaznamenávat.

K údajům v OLTP databázi přistupuje ve stejném čase více uživatelů, kteří údaje z databáze čtou nebo do ní zapisují tak, jak je to nutné k zaznamenání jejich aktivity. Současné softwarové systémy používané na úrovni OLTP obvykle poskytují možnosti (často velmi komplexní), jak z provozních dat získat znalosti potřebné pro vyhodnocení ekonomiky a řízení. OLTP poskytují sestavy, ze kterých lze hodnotit výnosnost obchodních a výrobních činností s ohledem na čas, sortiment, druh činnosti, původce a podle mnoha dalších kritérií. Z pohledu zpracování větších objemů dat nahromaděných za určitý čas nebo dat z více produkčních systému však tento způsob vyhodnocení dat není dobrým řešením [3, s.23].

Možnosti, které OLTP v oblasti vyhodnocení činnosti poskytuje, obvykle postačují, pokud je třeba řídit jednu, dvě nebo několik málo provozoven. Pokud se začne jednat o řízení většího systému, například celé sítě obchodních nebo výrobních provozoven a středisek v počtu desítek a stovek, pak tento způsob řízení neposkytuje optimální výsledky, protože neposkytuje uživateli analýz potřebnou abstrakci nad celým systémem a komfort.

Architektura databází systémů OLTP je pro analytické účely obvykle nevhodná. Struktura údajů v databázích je většinou vysoce detailní a komplexní, navržená pro vysoký výkon při zpracování vysokého počtu transakcí při současném přístupu mnoha uživatelů. Tabulky v databázi respektují pravidla tzv. normálních forem, kdy jsou z dat odstraněny redundance, databáze obsahují velké množství číselníků, vazeb mezi tabulkami. Taková struktura dat není pro analytické účely optimální [4, s.19].

Další důvody, proč jsou OLTP databáze nevhodné pro analytické účely, shrnuje Požár [8, s.188] následovně:

- OLTP systémy často z kapacitních důvodů udržují pouze data za určité období, (např. kalendářní rok), kdy se data minulých období odsunují do archívů.
- Nemožnost zpracování dat z různých systémů. Ve společnosti je obvykle provozováno více různorodých produkčních systémů, data vznikají i mimo produkční systémy nebo pocházejí z externích zdrojů.
- Zpracování dat z produkčních systémů do analytických výstupů může mít nepříznivý dopad na výkonnost a odezvu produkčních systémů.

2.2 Business intelligence jako řešení

Možné řešení popsané situace, kdy potřeby řízení a optimalizace provozu složitějších podnikových nebo jiných organizačních struktur narážejí na limity produkčních systémů, nabízí nasazení technologie označované jako Business intelligence (dále BI).

Definice BI

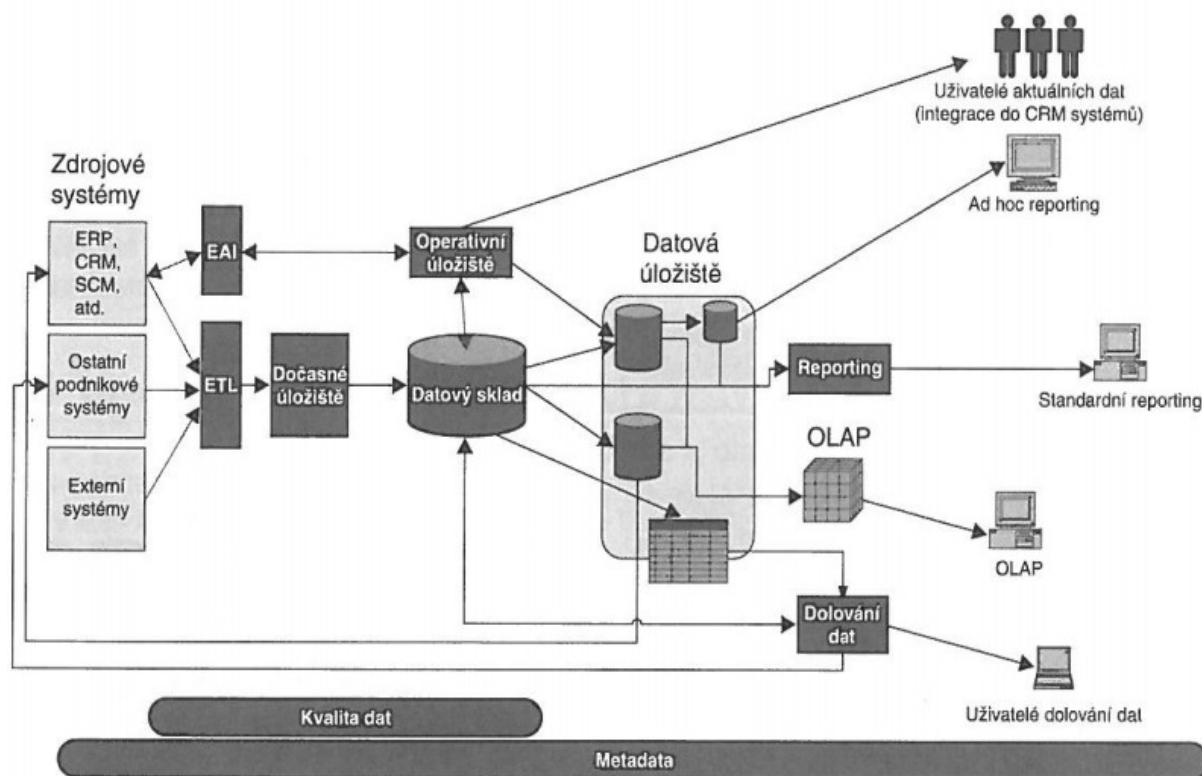
Novotný definuje BI následovně [6, s.19]

„Business intelligence je sada procesů, aplikací a technologií, jejichž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na principech multidimenzionálních pohledů na podniková data.“

Do nástrojů a aplikací BI je možné zahrnout

- Produkční systémy jako zdroj dat
- ETL jako transformační nástroje pro získávání dat z produkčních systémů
- Integrované nástroje (EAI – Enterprise Application Integration)
- Dočasná úložiště (DSA – Data Staging Area)
- Operativní úložiště (ODS – Operational Data Store)
- Datové sklady (DWH – Data Warehouses)
- Datová tržiště (DMA – Data Marts)
- Nástroje pro analýzu dat OLAP (On Line Analytical Processing)
- Reporting
- Dolování dat (Data Mining)
- a další

V schématu na Obr.2 je zachycena architektura, kterou je možné považovat za obecnou, výchozí koncepci, tak jak se během vývoje v oblasti BI ustálila. Uspořádání a použití jednotlivých komponent v konkrétních případech řešení BI se může měnit podle situace, požadavků a potřeb zákazníka (organizace) [1, s.227]. Schéma obsahuje všechny hlavní komponenty BI řešení a vazby mezi nimi.



Obr. 2 Hlavní komponenty BI a jejich vazby [6, s.28]

V realizovaných řešeních BI nemusí být některé komponenty zastoupeny, např. u jednodušších implementací nemusí být realizována dočasná nebo datová úložiště, mohou být zahrnuty jen některé analytické aplikace.

Na schématu figurují (kromě jiných) komponenty *ETL*, *Datový sklad*, které zajišťují oblast získání dat z produkčních systémů a jejich uložení pro další zpracování, a komponenty *OLAP* a *Reporting*, které patří do skupiny aplikací poskytujících funkce a nástroje pro přípravu a zpracování dat do analytických výstupů. Bližšímu popisu významu a účelu těchto komponent se práce v dalším textu věnuje.

2.3 Datový sklad

Základní a obvyklou součástí řešení BI je komponenta označovaná jako „*Datový sklad*“. Datový sklad plní roli hlavní datové základny, do které jsou pravidelně doplňována data z produkčních systémů resp. mohou být sehrávány i další zdroje dat, které jsou důležité pro analytické nástroje. Analytické části infrastruktury BI (OLAP, reporting, dolování dat) navazují na datový sklad přímo nebo zprostředkovaně prostřednictvím datových tržišť.

Jako datové tržiště se označuje vybraná část dat datového skladu, která v rozsáhlejších nebo komplikovanějších implementacích slouží specifickým potřebám některých typů uživatelů nebo oddělení.

Definici pojmu „Datový sklad“ poprvé (v roce 1991) zformuloval William Harvey Inmon jako: „*Datový sklad je integrovaný, subjektivě orientovaný, stálý a časově rozlišený souhrn dat, uspořádaný pro podporu potřeb managementu.*“ [1, s. 229]

Jiná spíše „obchodní“ definice je formulována takto [4, s. 39] :

„Datový sklad musí poskytovat aktuální a přesné odpovědi na dotazy a musí je produkovat v co nejkratším čase.“

Pojmy z Inmonovy definice lze interpretovat následovně [6, s.32] :

- **Subjektová orientace** - data jsou rozdělována podle jejich typu, ne podle aplikace, místa kde nebo způsobu jak vznikla.
- **Integrovanost** - data jsou ukládána za celou organizaci, a ne pouze v rámci jednotlivých oddělení.
- **Stálost** - datové sklady jsou koncipovány jako "Read Only", což znamená, že zde žádná data nevznikají ručním pořízením, ale pouze procesy ETL, data nelze žádnými uživatelskými nástroji měnit.
- **Časová rozlišenost** – pro provedení analýz za určitá období, je nutné, aby byla data do datového skladu uložena včetně informací časového rozlišení.

Marakas charakteristiky datového skladu doplňuje [5, s.46]:

- **Sumarizace, souhrny** – provozní data jsou mapována do formátu používaného pro rozhodování
- **Normalizace** – respektování normálních forem není v datovém skladu dodržováno, data mohou být a často jsou uložena redundantně
- **Velký objem** - sady dat v datovém skladu jsou obvykle velmi velké
- **Metadata** – datový sklad obsahuje metadata
- **Zdroje dat** – jsou sehrávána data z interních i externích neintegrovaných produkčních systémů a zdrojů

V následující tabulce jsou popsány základní charakteristiky produkční databáze a datového skladu.

Tabulka č.1 Základní charakteristiky produkční databáze a datového skladu [6, s.33]

	Produkční databáze	Datový sklad
Funkce	Zpracování transakcí, obchodních operací	Podpora rozhodování
Data	Orientovaná na proces, aktuální hodnoty	Předmětně orientovaná, obvykle sumarizovaná, historická i aktuální
Užití	Strukturovaně, opakovaně, požadována rychlá odezva	K analytickým účelům, <i>ad-hoc</i> nebo opakující se reporty, analýzy
Procesy	OLTP, obchodní operace	OLAP, reporting, dolování dat

Datový sklad je ucelenou databází, ve které je organizace dat přizpůsobena a optimalizována pro dotazování a analýzu dat. Cílem je poskytnout čitelné, organizované informace z produkčních databází a dalších interních nebo externích zdrojů, které je možné prostřednictvím navazujících analytických nástrojů využít k řízení firmy nebo instituce. [10, s.108] Datový sklad je organizovaný podle jiných pravidel než produkční systémy, data se mohou vyskytovat redundantně, v nenormalizované podobě.

2.4 Informační systém BI s datovým skladem

Informační systém BI by měl poskytnout kvalitní analytické výstupy a poskytnout abstrakci nad celým sledovaným a řízeným systémem při zachování komfortu při obsluze.

Abstrakcí nad systémem je myšlen přístup, kdy při získávání dat není třeba brát ohled na typ produkčního systému, na místo uložení dat, geografickou polohu provozoven. Důležité je to, jakým způsobem ovlivňují a zapadají jednotlivé části celku do výsledného celkového obrazu (dat) .

Komfortem se myšlena taková úroveň přístupu k informacím a znalostem, kdy analytické výstupy jsou dostupné v reálném nebo v krátkém čase na běžném zařízení typu PC s běžnou mírou znalosti obsluhy výpočetní techniky a aplikací.

Budování informačního systému s datovým skladem se zdá být velmi komplikovaným projektem v oblasti informatiky. Lacko [4, s.41] tento náhled vidí mnohem jednodušeji a komentuje ho následovně :

"Problém je formulovatelný do jednoho krátkého odstavce. Stačí získat data z operačního prostředí a po úpravě je zavést do datového skladu. Nad těmito daty vykonat analýzy a jejich výsledky zpřístupnit uživatelům, tedy manažerům a analytikům."

V následující tabulce jsou shrnuty a popsány rozdíly mezi OLTP databází a datovým skladem.

Tabulka č.2 Základní rozdíly databází OLTP a datového skladu [4, s.40]

Vlastnost	Produkční databáze (OLTP)	Datový sklad
Čas odezvy	Zlomky sekund až sekundy	Sekundy až hodiny
Operace	DML ¹ (Data Manipulation Language) CRUD ²	Primárně jen čtení
Původ dat	30-60 dní	Série snímků za časový úsek
Organizace dat	Podle aplikace	Podle předmětu a času
Velikost	Malá až velká	Velká až velmi velká
Zdroje dat	Operační, interní	Operační, interní, externí
Činnosti	Procesy	Analýza

V tabulce je znovu poznamenáno, že data jsou v datovém skladu organizována podle předmětu a času. Tato organizace dat poskytuje oproti OLTP mnohem širší možnosti pro analytické zpracování dat.

2.5 Extrakce, transformace a přenos

V předchozím textu je popsáno, že

- Data se v pravidelných intervalech přenášejí z produkčního prostředí do databáze datového skladu.

¹ DML je zkratka pro *Data Manipulation Language*, jedná se jazyk pro provádění operací a dotazování nad daty v SQL databázích

² CRUD je zkratka pro základní operace nad tabulkou databáze *Create, Read, Update, Delete*

- Struktura databází OLTP a datového skladu je rozdílná.
- Aplikace OLTP mohou být různorodé strukturou dat i technologicky.

Z těchto důvodů přenos není a ani nemůže být realizován přímo v tom smyslu, že by se data z databáze OLTP pouze označila časem sehrání a zapsala se do datového skladu. Přenos dat je netriviální proces, který je realizován prostřednictvím vrstvy ETL, která v procesu přenosu dat provádí následující operace [4, s.75].

Extrakce

Jedná se o schopnost převzít (získat) data z produkčních systémů často různého charakteru, formátu nebo struktury. Po ukončení této fáze jsou data ve formátu struktur pro extrahovaná data

Transformace

Transformace je postupná řada operací a úkonů pro čištění, ověření validity, integraci a časové označení dat. Důvodem pro provedení těchto úkonů je stav, kdy data často nejsou po extrakci z produkčních systémů pro načtení do datového skladu dostatečně připravena.

Obvyklým problémem je neúplnost dat, nesoulad dat z různých zdrojů, rozdíl ve formátech. Všechny tyto nedostatky a nečistoty je třeba odstranit. Je nutné provést kontroly, doplnění nebo změny v datech, konverze na stejné formáty, odstranění duplicit a jiných nekonzistencí, proběhne celková konsolidace dat [10, s.104].

Přenos

Přenos je proces, kdy se extrahovaná a transformovaná data vkládají do fyzického prostoru datového skladu.

2.6 Analytické nástroje

Analytické nástroje řešení BI jsou komponenty, které zajišťují zpracování dat uložených v datovém skladu do požadovaných analytických výstupů pro podporu rozhodování a optimalizaci činností organizace. V obecné architektuře BI (viz *Obr. 2 Hlavní komponenty BI a jejich vazby*) k nim řadíme nástroje *OLAP*, *Reporting* a systémy pro *Dolování dat*. Práce

se v dalším textu podrobněji zabývá nástroji OLAP a krátce je popsán i účel a funkce reportingových systémů.

2.6.1 OLAP

Pojem OLAP zavedl Dr. E. F. Codd s účelem popsat nástroje, postupy a technologie zpracování dat uložených v datovém skladu do podoby, která je srozumitelná a vhodná pro koncové uživatele, obvykle manažery a řídicí pracovníky.

Historická poznámka - Pojem OLAP dříve a dnes

Pro systémy pro práci s produkčními a analytickými daty se v osmdesátých letech minulého století používal název OLAP. V pozdější době byl zaveden pojem BI a pojem OLAP se nyní používá pro analytické nástroje a systémy pro práci s multidimenzionální databází. [6, s.21]

Jednou z existujících definic OLAP [4, s.175] je *„OLAP je volně definovaná řada principů, které poskytují dimenzionální rámec pro podporu rozhodování.“*

OLAP databáze je soubor nástrojů a dat, které jsou dostupné v jedné nebo více OLAP datových kostkách. Datová kostka, narozdíl od datového skladu, obsahuje data předzpracovaná podle určených (definovaných) struktur dimenzí a jejich případných kombinací. [6, s.33]

Technologie OLAP doplňuje dříve používané statické sestavy a při pohledu na data dává uživatelům možnost aktivně s datovou strukturou pracovat. Uživatel může volit míru detailu zobrazení rozbalením nebo sbalením podrobností sestavy, může volit výběr dat omezením dimenzí, může také nahlížet na data z různé perspektivy např. z pohledu času, regionu, produktu, apod. OLAP kostka v podstatě představuje souhrn mnoha sestav v jedné [2, s.30].

Pro budování struktur OLAP se využívá technologie multidimenzionální databází. OLAP databáze je připojena na zdroj dat, kterým jsou obvykle tabulky datového skladu. Může být také připojena i na jiné zdroje dat. Může jít například o napojení přímo na produkční systém, o napojení na zdroj dat jiný než relační databázový systém (tabulky MS Excel) nebo zdrojem dat může být jiná OLAP databáze [6, s.171]. V dalším textu bude uvažována varianta OLAP připojená na datový sklad.

OLAP kostka je datová struktura, která obsahuje obsahuje dva základní druhy dat :

- tabulku faktů, která obsahují metriky kvalifikované podle dimenzí
- dimenze

Metriky jako součást tabulky faktů mohou být [2, s. 156]

- aditivní, lze je agregovat podle všech dimenzí např. prodané množství zboží
- semiaditivní, lze je agregovat jen podle některých dimenzí
- neaditivní, nelze je agregovat např. procento sazby dph, pohlaví pacienta

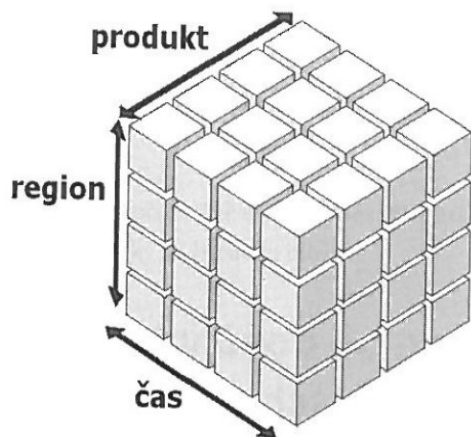
Metriky je dále možné členit podle zdroje na [2, s. 160]

- získané, jsou metriky které jsou převzaty z produkčního nebo jiného systému tak, jak jsou např. prodejní cena zboží v konkrétní transakci
- odvozené, jsou metriky, které vznikly až v datovém skladu např. denní počty prodejů

Lacko dělení na fakta a dimenze popisuje [4, s. 172] takto:

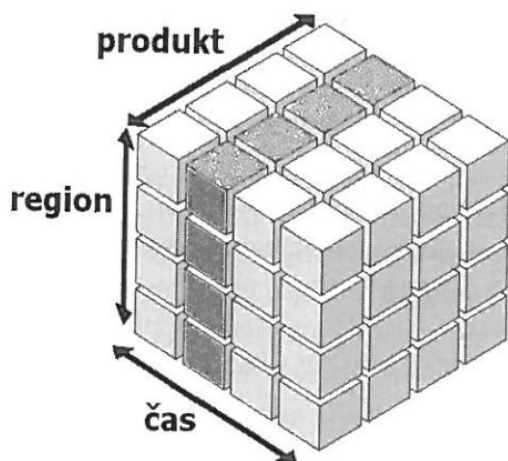
„Fakta jsou numerické měrné jednotky obchodování. Dimenze obsahují logicky nebo organizačně hierarchicky uspořádaná data. Jsou to vlastně textové popisy obchodování.“
Tabulka faktů obsahuje velký objem dat a je obvykle největší tabulkou databáze. Jedná se o záznamy realizovaných transakcí. Tabulka faktů může být jedna anebo jich může být více. Tabulka dimenzí je obvykle menší než tabulka faktů. „Tabulky dimenzí vysvětlují všechna „proč“ a „jak“, pokud jde o obchodování a transakce prvků.“ (Lacko, 2009, s.177)

OLAP kostka je multidimenzionálním datová struktura, která je výsledkem agregace a analýzy obvykle velkého množství dat. OLAP kostka má několik dimenzí, jejichž počet není omezený. Jako příklad je možné uvažovat dimenze *čas, region, produkt*. Uvažovanou OLAP kostku si lze představit jako 3D objekt (kostku), v níž jsou data (metriky) logicky situována v průnicích dimenzí.

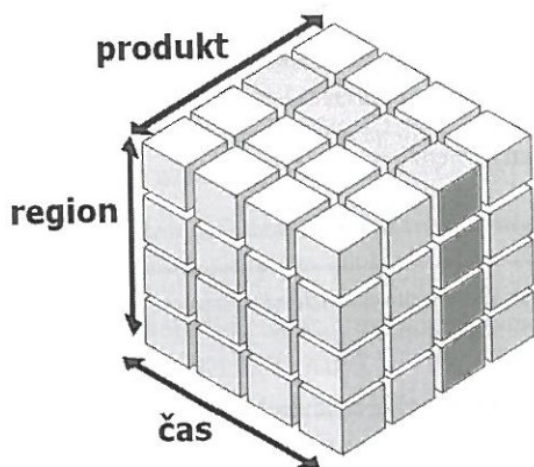


Obr. 3: Příklad OLAP kostky [4, s.173]

Analýzu dat za vybrané časové období s omezením dimenze *čas* resp. dimenze *produkt* lze modelovat následovně.



Obr. 4: Příklad OLAP kostky [4, s.173], analýza časového období podle produktu a regionu



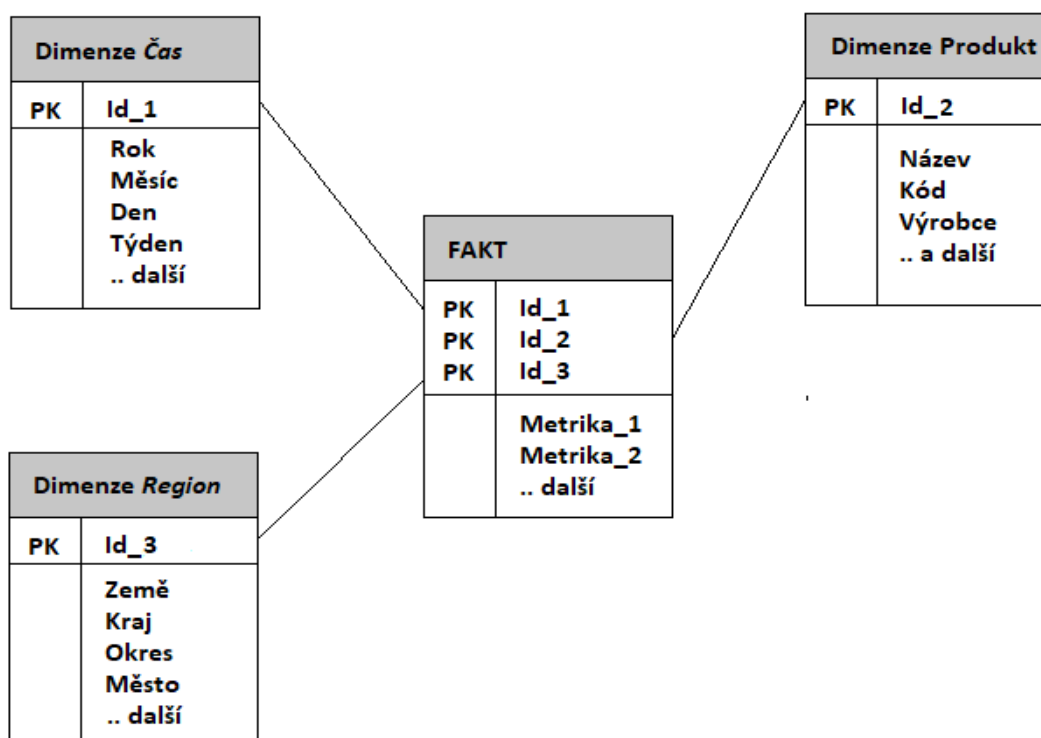
Obr. 5: Příklad OLAP kostky [4, s.174], analýza produktu podle času a regionu

2.6.2 Schémata dimenzí

OLAP krychle vznikne na základě modelu, který má určité topologické uspořádání, které se označuje jako *schéma*. Nejčastěji používaným typem schématu je hvězdicové schéma (star schéma) a schéma „sněhové vločky“ (snowflake schéma). [3, s.114]

Hvězdicové schéma je tvořeno tabulkou faktů a samostatnými tabulkami pro každou dimenzi. Tabulka faktů obsahuje cizí klíče, které určují prostřednictvím primárních klíčů tabulek dimenzí vztah faktů k dimenzi. Schéma nemá normalizované dimenze a neobsahuje vazby mezi tabulkami dimenzí. Výhodami schématu jsou snadná pochopitelnost a vysoký výkon při dotazování, nevýhodou pak obtížnější tvorba modelu, kdy některé dimenze (např. „produkt“) mají spíše strukturovaný charakter.

OLAP kostku s dimenzemi *čas*, *produkt*, *region* lze znázornit následujícím schématem.

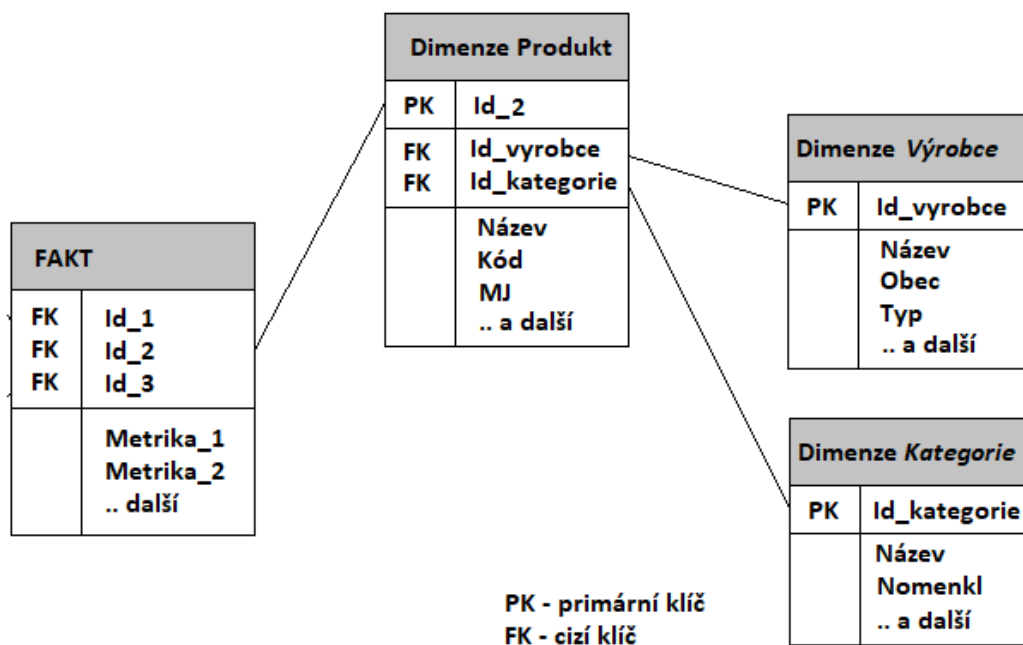


Obr. 6: Schéma dat pro analýzu „hvězda“ [zdroj: vlastní]

Schéma „sněhové vločky“ obsahuje dimenze složené z relačně propojených normalizovaných tabulek. Schéma poskytuje nižší výkon při dotazování. Ve fázi modelování dat schéma poskytuje pro návrh některých typů dimenzí vhodnější řešení např. v dimenzi

produkt může být vnořeno mnoho dalších členění (výrobce, zařazení do nomenklatur podnikových nebo oborových, apod.), které jsou přirozenou součástí údajů o produktu.

OLAP kostku lze v dimenzi *produkt* rozvinout a znázornit následujícím schématem.



Obr. 7: Schéma dat pro analýzu „sněhová vločka“ [zdroj: vlastní]

Schéma „Soubor hvězd“ - constellation

Tvrdíková [10, s.104] používá pojem schéma „Soubor hvězd“, které je kombinací obou předchozích schémat, kdy je do modelu začleněno více tabulek faktů, k nim jsou definovány dimenze, které jsou tabulkami faktů sdíleny.

2.6.3 Dělení OLAP podle uložení dat

Technologie OLAP je možné podle typu uložení dat a způsobu zpracování dělit [10, s.114]

- ROLAP (Relational OLAP), data pro analýzu uložena v relační podobě, multidimenzionální dotazy se automaticky překládají na SQL příkazy, umožňuje realizovat přístup k datům přímo do produkčních databází

- MOLAP (Multidimensional OLAP), data jsou uložena ve specifické multidimenzionálních databázi, výsledky agregace a sumarizace dat jsou již vypočítány a uloženy v databázi, přístup k datům je velmi rychlý
- HOLAP (Hybrid OLAP) hybridní řešení, kombinace MOLAP a ROLAP
- DOLAP (Desktop OLAP) jedná se o řešení, kdy se z centrální datové kostky stáhne požadovaná část kostky na lokální počítač a další zkoumání probíhá na d staženou částí

2.6.4 Reporting

Reportingový systém je aplikace, která poskytuje analytické výstupy prostřednictvím dotazování do databází (datového skladu) v rámci standardních rozhraní těchto databází (obvykle DML např. Transact-SQL) . Způsob jakým jsou dotazy formulovány, parametrizovány a předávány může být řešen na různé úrovni, obvykle je pro tyto činnosti připraveno uživatelské rozhraní, které umožní komfortně dotaz zadat (včetně uložení pro opakované použití), předat ke zpracování a získat výsledek. Některé implementace reportingu je možné ve flexibilitě a možnostech dotazování srovnávat s aplikacemi OLAP.

3. Popis prostředí pro plánovanou implementaci BI

V této kapitole je popsáno prostředí, ve kterém proběhne implementace infrastruktury BI včetně využití technologie OLAP. Popis prostředí odpovídá stavu ve fázi přípravy a návrhu implementace.

3.1 Struktura obchodního řetězce APC

Síť APC je tvořena 31 firmami, které provozují 55 středisek typu lékárna, prodejna zdravotnického materiálu, distribuční pracoviště, e-shop. Pro všechny typy středisek dále bude používán společný termín „*provozovna*“.

3.2 Produkční systém

V síti APC je pro zajištění obchodních aktivit používán jednotný informační systém RADIX (dále bude používán termín „*SW RADIX*“). Jedná se o informační systém realizovaný na relační databázové platformě *Microsoft VisualFoxPro 9.0*.

SW RADIX je představitelem OLTP systému. Slouží především k

- provádění a zaznamenání obchodních transakcí spojených s činností provozovny (výdejů léku, zdravotnického materiálu a ostatního zboží)
- evidenci stavu skladů
- vyúčtování péče poskytnuté pacientům a zákazníkům na portálech zdravotních pojišťoven
- evidenci plateb hotovostních, platebními kartami a poukázkami provedených v souvislosti s prodejem zboží realizovaném prostřednictvím maloobchodních výdejních míst (pokladen), případně plateb za zboží dodané na faktury
- ke komunikaci s úložišti elektronických dokumentů a hlášení (CÚeR „*Centrální úložiště elektronických receptů*“, ROB „*Registr obyvatel*“, ÚZIS „*Ústav zdravotnických informací a statistiky*“, EET „*Elektronická evidence tržeb*“ a další
- evidence objednávek přijatých od odběratelů a stavů jejich vykrytí
- evidence objednávek vystavených k dodavatelům
- fakturaci zboží dodaného odběratelům – zdravotnickým zařízením a ostatním odběratelům
- evidenci receptů a poukazů

Popsané provozní a obchodní činnosti jsou zaznamenávány v 82 samostatných databázích. Důvodem pro vyšší počet produkčních databází než je počet provozoven, je skutečnost, že některé provozovny mají pro různý typ činností zavedené různé databáze např. na jedné provozovně mohou být samostatně vedeny agendy e-shopu, zásobování nemocnic, lékárny.

Hodnocení

Používání jednotného informačního systému SW RADIX je z pohledu budování infrastruktury BI výhodou. Procesy ETL je možné realizovat a udržovat pouze nad jedním typem provozních dat. Při přípravě analytických databází lze vycházet z jednoho modelu dat produkčního systému.

3.3 Jednotné číselníky

V rámci přípravy datových struktur a zdrojů pro budování infrastruktury BI je nutné zajistit existenci jednotných číselníků důležitých pro definici a tvorbu dimenzí. Jedná se o číselník sortimentu, obchodních partnerů a uživatelů OLTP. Dimenze vybudované z těchto datových zdrojů budou využity v multidimenzionálních OLAP databázích i v reportingovém systému. [6, s.117)

Situace v oblasti číselníků je v rámci sítě APC je následující:

a/ Neexistuje jednotný číselník sortimentu

Sortimentní položky vznikají na provozovnách podle potřeby a bez jakékoli koordinace tzn. na každé provozně nezávisle. Pro pořízení se využívají některé veřejně nebo komerčně dostupné číselníky (seznam cen a úhrad léčiv SCAU [13], seznam cen a úhrad zdravotnického materiálu ZPSCAU [14], sortimentní číselník PDK [11] a další). Využití těchto datových zdrojů je nepovinné.

b/ Neexistuje jednotný číselník odběratelů a dodavatelů

Záznamy o obchodních partnerech vznikají podobným způsobem jako sortimentní položky, tzn. nezávisle a bez koordinace mezi provozovnami.

c/ Existuje jednotný číselník uživatelů SW RADIX

Číselník uživatelů (dále „CUZIV“) je v celé síti veden jako centrální agenda s pravidelnou publikací změn ve směru z centrály na provozovnu.

Hodnocení

Situace popsaná v bodech a/ a b/ je problémem, který je nutné odstranit.

3.4 Výstupy, reporty, analýzy

Podklady pro podporu rozhodování jsou v současnosti připravovány podle následujícího scénáře

- podklady (sestavy a exportní soubory) jsou generovány na provozovně ze SW RADIX a to pro účely provozovny nebo jsou postoupeny k dalšímu zpracování
- vybrané podklady jsou dále ručně nebo automatizovaně upravovány nebo sehrávány podle potřeb uživatelů do reportů
- reporty jsou rozesílány jednotlivým uživatelům a oddělením nebo jsou uloženy na sdílená úložiště pro případné pozdější použití

Tento stav popisuje Lacko následovně [4, s.75]

„K dispozici jsou různé sestavy, dokumenty, a data z primárních transakčních systémů OLTP (Online Transaction Processing).“

„Zpravidla i v těchto případech používá zákazník pro získávání informací pro podporu rozhodování nějaké provizorní řešení, například jsou generovány výstupní sestavy z transakčních systémů, a tyto jsou potom buď ručně, nebo pomocí automatizačních prvků software typu Office zpracovány do podkladů poskytovaných manažerům pro podporu rozhodování.“

Hodnocení

Současný způsob přípravy a distribuce reportů neposkytuje uživatelům dostatečný komfort a flexibilitu. Změny v podkladech je třeba projednat, odsouhlasit a realizovat v rámci a podle zavedených pravidel tzn. že většinou není možné provést operativně změnu. Zároveň má současný stav velké nároky na lidské zdroje a koordinaci a je náchylný ke vzniku chyb a nepřesností. Jednoznačně lze stav hodnotit jako nevyhovující.

3.5 Uživatelé

Uživatelem současných výstupů je

- personál na provozovnách - má přístup jen k datům provozovny
- vedoucí provozovny(en) - má přístup k datům provozoven, které řídí
- vedení společnosti, manažeři - mají přístup ke všem reportům generovaným z pokladů zaslaných z provozoven
- oddělení nákupu, zásobování, marketingu - má přístup k vybraným reportům generovaným z pokladů zaslaných z provozoven

Uživatelé analytických nástrojů budou členěni do stejných skupin, přístup k jednotlivým typům dat a oblastí bude autorizován podobným způsobem jako je to zavedeno v současnosti.

3.6 Konektivita

Nastavení procesů ETL bude provedeno tak, že v časovém rámci přibližně 3 hodin v noci bude každý den probíhat sehrání dat provozoven do datového skladu. Pro účely pravidelného sehrávání dat provozoven do datového skladu je současné připojení k internetu na straně provozoven i infrastruktury v centrále dostatečně kapacitní.

3.7 Sehrávání dat, aktualizace číselníků

Na provozovnách APC je v současnosti pravidelně spouštěna mezi 01:00 hod až 04:00 hod noční procedura, do které jsou zařazeny úkony pro kontroly a zálohování dat, aktualizace centrálních číselníků.

V souvislosti se stavem zabezpečení a možnostmi přístupu je možné aktivovat (spustit) noční proceduru pouze na straně provozovny. Z bezpečnostních důvodů není možné a nepředpokládá se, že by data provozovny byla dostupná nebo spuštění jakýchkoli procedur nebo úkonů proběhlo z vnějšího prostředí tzn. aktivita by vycházela odjinud než z provozovny. Pro spuštění procesu ETL se předpokládá využití noční procedury.

4. Příprava na implementaci BI v APC

Tato kapitole popisuje kroky, které je nutné provést k odstranění překážek, které by implementaci BI v řetězci APC mohly bránit nebo vytvářet podmínky pro možný neúspěch.

4.1 Sjedení číselníků

Velkou nevýhodou či přímo problémem jsou neexistující společné číselníky sortimentu a obchodních partnerů. Tyto číselníky budou vstupovat do procesů souvisejících s analytickým zpracováním dat jako důležitá součást ve formě dimenzí, proto je nutné před implementací BI tento klíčový problém vyřešit [6, s.117]

Problém bude řešen zavedením jednotného sortimentu (dále „JSO“) a zavedením centrálního číselníku obchodních partnerů (dále „COP“). JSO a COP bude poskytován ve formě verzovaných aktualizací prostřednictvím publikačního systému CENTDATA (REST API aplikace provozovaná APC). Aktualizace JSO a COP budou poskytnuty jednak provozním a zároveň nahrávány do datového skladu. Aktualizace bude probíhat v rámci noční aktualizací procedury.

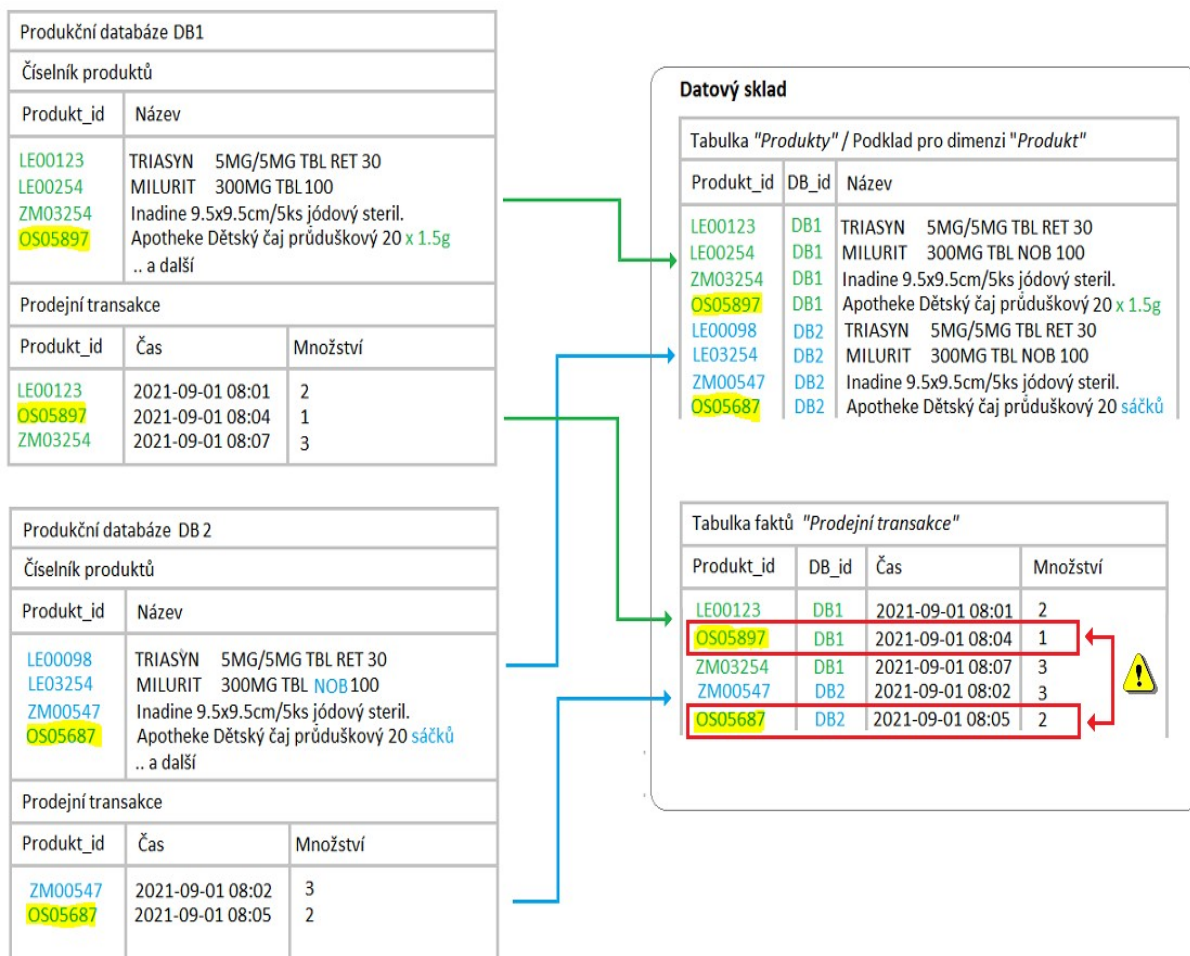
4.1.1 Zavedení JSO

Níže na obrázku 8 je modelově zachycen stav před zavedením JSO. Je patrné, že chybějící koordinace v zakládání a použití sortimentních položek mezi produkčními systémy neumožní identitu „stejně“ sortimentní položky z různých produkčních databází v datovém skladu správně interpretovat. Příklad ukazuje, že v datovém skladu není možné spolehlivě zjistit, že položky „*Apotheke dětský čaj průduškový*“ evidované v DB1 jako produkt_id= OS05897 a v DB2 jako produkt_id=OS05687 jsou totožné. Analýzu prodejních transakcí tedy nelze správně provést.

K odstranění tohoto problému je třeba zavést JSO, tzn. z jednoho místa přesně a úplně řídit a koordinovat zakládání a správu sortimentních položek pro produkční systémy i navazující aplikace BI.

Zavedení JSO za běhu produkčních systémů je nelehký, organizačně i časově náročný úkol. Jako řešení je navržen níže popsáný postup, který umožní zavádět JSO na

provozových postupně v rámci 2-3 měsíců s předpokládaným souběhem provozoven v různé fázi zavádění.



Obr. 8: Situace před zavedením JSO [zdroj: vlastní]

Zavedení jednotného sortimentu proběhne ve čtyřech fázích a to

Fáze 1/ Příprava číselníku JSO

Příprava bude provedena formou výběru sortimentních položek (u nichž se předpokládá, že s nimi bude obchodováno nebo se již obchoduje) a to z různých zdrojů (produkční databáze, číselníky spravované SUKL, číselníky poskytované na komerční bázi např. PDK, produktové číselníky dodavatelů a výrobců). Nad vybranými zdroji proběhne kompilace (výběr, spojení, odstranění duplicit) do jednoho verzovaného zdroje dat JSO.

Fáze 2/ Napárování současného sortimentu na JSO

V rámci této fáze proběhne jednoznačné a úplné napárování sortimentních položek v produkčních systémech na JSO.

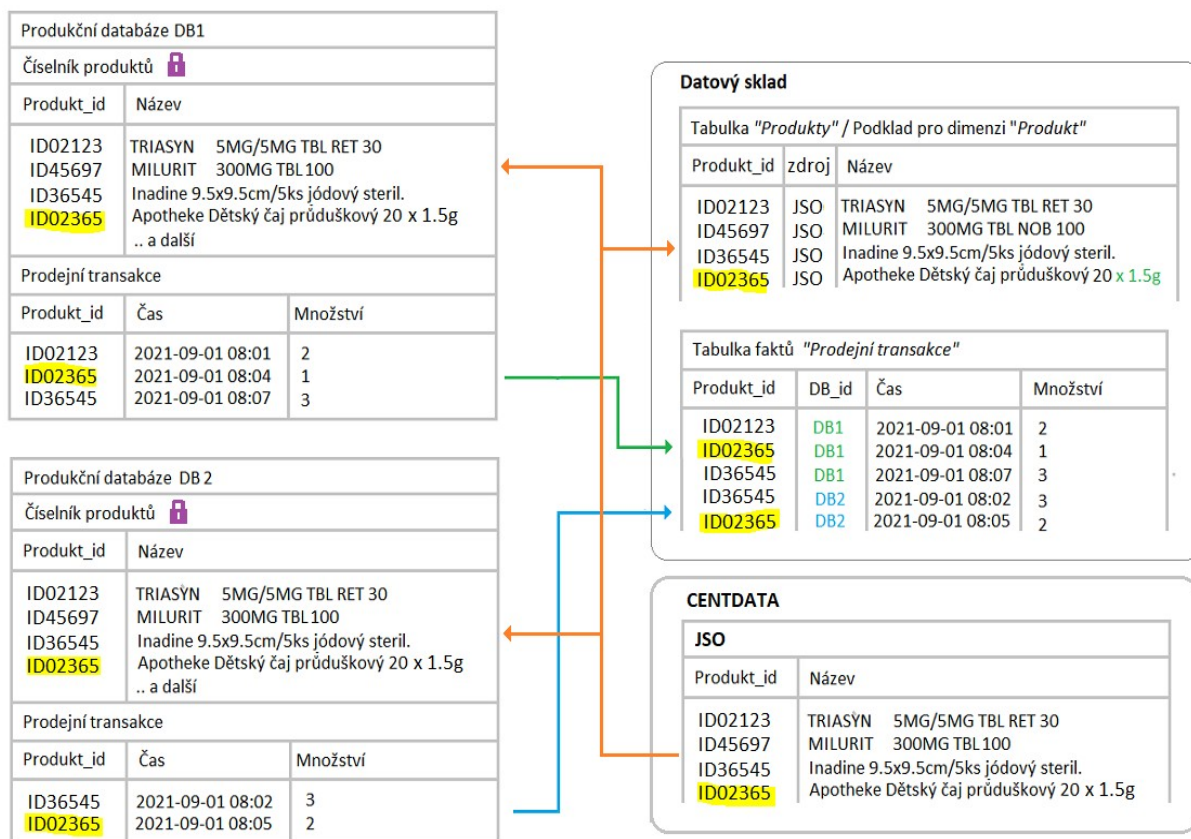
Fáze 3/ Uzavření sortimentu na provozovně

Jedná se o provedení změny, kdy provozovna již nadále nemá možnost samostatně sortimentní položky v produkčním systému pořídít, upravovat (centrálně řízené údaje) nebo rušit. Provozovna může žádat pověřenou osobu (správce JSO) o zalistování nebo změnu sortimentní položky. Údaje, které mají lokální charakter (např. normy skladových zásob, lokální cenové kampaně, apod.) nejsou součástí JSO a provozovna je upravovat může.

Fáze 4/ Spuštění aktualizací služby JSO

Číselník JSO bude nadále udržovat správce JSO, který průběžně sortimentní položky doplňuje resp. upravuje podle požadavků z provozoven a změny následně publikuje prostřednictvím aktualizací JSO službou CENTDATA.

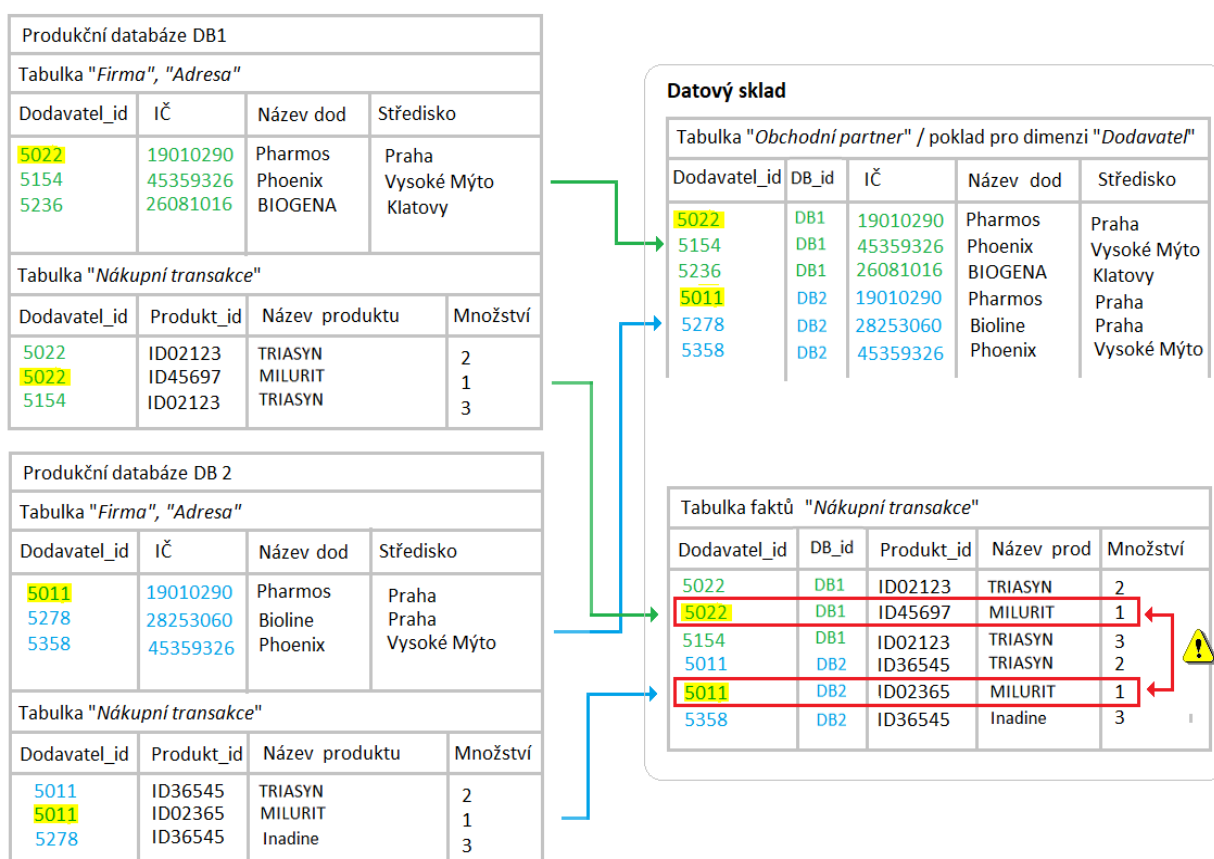
Stav po zavedení JSO je znázorněn v obrázku 9. Sortimentní položka „*Apotheke dětský čaj průduškový*“ je ve všech systémech evidovaná jako produkt_id=ID02365. Analýzu prodejních transakcí lze nyní správně provést. U produkčních databázích piktogram zámku indikuje, že zakládání a správa sortimentních položek již není dostupná. Oranžové šipky ukazují, že zdrojem dat pro číselník produktů je nyní výhradně JSO.



Obr. 9: Situace po zavedení JSO [zdroj: vlastní]

4.1.2 Zavedení COP

Schéma na obrázku 10 zachycuje stav před zavedením COP, zvýrazněné oblasti modelově ukazují příklad problému. Podobně jako u JSO chybějící koordinace mezi produkčními systémy při zakládání a použití záznamů o obchodních partnerech neumožňuje transakce z různých produkčních databází v datovém skladu správně interpretovat. Příklad ukazuje, že v datovém skladu není možné nákupní transakce sortimentní položky „MILURIT“ od obchodního partnera s dodavatel_id=5022 v DB1 a dodavatel_id=5011 v DB2 považovat spolehlivě za dodávku od stejného dodavatele. Řešení problému pomocí údaje IČ nelze použít, protože někteří obchodní partneři IČ nemají a u některých je pod jedním IČ evidováno více středisek v různých kombinacích zavedení v jednotlivých OLTP. Analýzu nákupních transakcí tedy nelze správně provést.



Obr. 10 : Stav před zavedením COP [zdroj: vlastní]

Řešením problému je zavedení COP, tzn. na jednom místě přesně a úplně zakládat a spravovat záznamy o obchodních partnerech pro produkční systémy i navazující aplikace BI.

Zavedení COP znamená provést úplnou náhradu číselníků obchodních partnerů v produkčních systémech za COP. Jako řešení je navržen níže popsany postup.

Zavedení centrálního číselníku obchodních partnerů proběhne ve třech fázích a to

Fáze 1/ Příprava číselníku COP

Příprava COP bude provedena formou sehrání obchodních partnerů z produkčních systémů do jednoho verzovaného zdroje dat. Do COP budou sehráni obchodní partneři, u kterých je evidována nějaká obchodní aktivita od 1.1.2019

V rámci přípravy

- budou vyřešeny duplicity, jedná se řešení situace, kdy jsou někteří obchodní partneři zavedení na více provozovnách
- bude provedena kontrola na správnost a úplnost údajů v registru ARES, proběhne případná korekce nebo doplnění údajů

Fáze 2/ Uzavření číselníku obchodních partnerů na provozovně

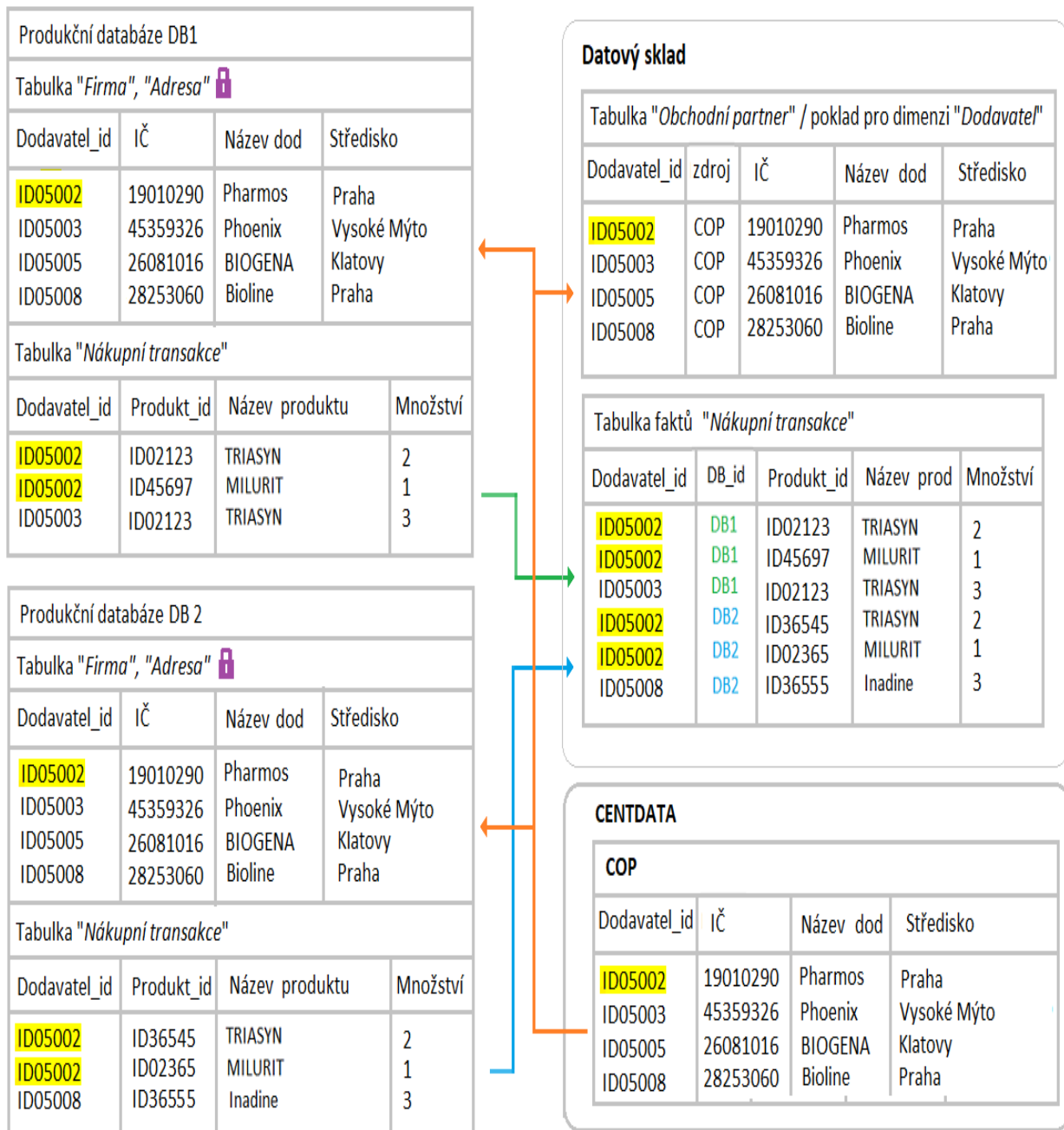
Podobně jako u JSO se jedná o provedení změny, kdy provozovna již nadále nemá možnost samostatně obchodní partnery v produkčním systému pořídit, upravovat (centrálně řízené údaje) nebo rušit. Údaje, které mají lokální charakter (např. obchodní podmínky, poznámky, apod.) nejsou součástí COP a provozovna je upravovat může.

V rámci této fáze proběhne úplná výměna dat o obchodních partnerech v produkčních systémech a to záznamy v COP. Časově je možné fázi provést na přelomu kalendářního roku, kdy po úkonech spojených s ukončením provozu roku nejsou v produkčních systémech doklady, které by na původní záznamy obchodních partnerů odkazovaly.

Fáze 3/ Spuštění aktualizací služby COP

Číselník COP bude nadále udržovat správce COP, který průběžně obchodní partnery doplňuje resp. upravuje podle požadavků z provozoven a změny následně publikuje prostřednictvím aktualizací službou CENTDATA.

Stav po zavedení COP je znázorněn na obrázku 11. Obchodní partner „PHARMOS“ je ve všech systémech evidován jako partner_id=ID05002. Analýzu nákupních transakcí lze nyní správně provést. U produkčních databázích piktogram zámku indikuje, že zakládání a správa obchodních partnerů již není dostupná. Oranžové šipky ukazují, že zdrojem dat je nyní výhradně COP.



Obr. 11 : Stav po zavedení COP [zdroj: vlastní]

4.2 Mapování požadavků

Při přípravě analytických výstupů je třeba důkladně zmapovat v současnosti používané reporty a analytické výstupy tak, aby bylo možné každý jednotlivý současný report nahradit odpovídajícím reportem v novém řešení BI.

V tomto procesu je třeba provést inventarizaci a odsouhlasení reportů, které umožní jednak upřesnit strukturu a obsah reportů pro jejich realizaci v BI a také vyřadit reporty, které jsou sice v současnosti připravovány, ale již nejsou používány. Provedení inventarizace a odsouhlasení proběhne formou postupného projednání, kdy pověřená

osoba projde s každým relevantním uživatelem množinu „jeho“ reportů. Tak vznikne výčet reportů k realizaci v novém řešení BI.

4.3 Souběh starého a nového řešení

Předpokládá se, že po nasazení a rozběhu řešení BI bude původní systém přípravy reportů a analytických výstupů nadále běžet. Důvodem je možnost provádět průběžné ověřování validity nových výstupů proti původním výstupům. Validace také proběhne v rámci akceptace řešení BI, ale pro úplné ověření funkčnosti nového systému je třeba mít možnost výsledky alespoň 6 měsíců komparovat. Souběh zároveň umožní důkladné seznámení uživatelů s novým systémem BI bez ohrožení obchodních aktivit nebo jiných činností.

5. Volba řešení BI

Kapitola obsahuje informace o zvoleném řešení BI pro implementaci v řetězci APC včetně popisu důvodů, které vedly k výběru konkrétní varianty.

5.1 Volba řešení z pohledu časového a organizačního

Při budování datového skladu je velmi důležitým krokem výběr správné (nejvhodnější) metody. Do úvahy je třeba vzít organizační strukturu, finanční možnosti, personální kapacity a další aspekty. Metody budování datového skladu a infrastruktury OLAP lze [4, s.44] z pohledu časového a organizačního rozdělit na následující varianty:

- **Metoda velkého třesku**

Metoda uvažuje o implementaci datového skladu v rámci jednoho projektu. Výhodou je, že realizace proběhne rychleji v čase, a obvykle ve 3 etapách

- Analýza požadavků
- Realizace datového skladu
- Realizace přístupu k datům

- **Přírůstková metoda (evoluční)**

Metoda předpokládá, že datový sklad bude budován po etapách, kdy každá realizovaná část bude zapadat jako součást celkové plánované architektury. Metoda má výhodu v tom, že během budování je možné lépe reagovat na případné změny v požadavcích. Další výhodou je, že obvykle je možné rychleji předložit a ověřit „nějaký“ výsledek např. pro potřebu managementu

Volba metody v APC

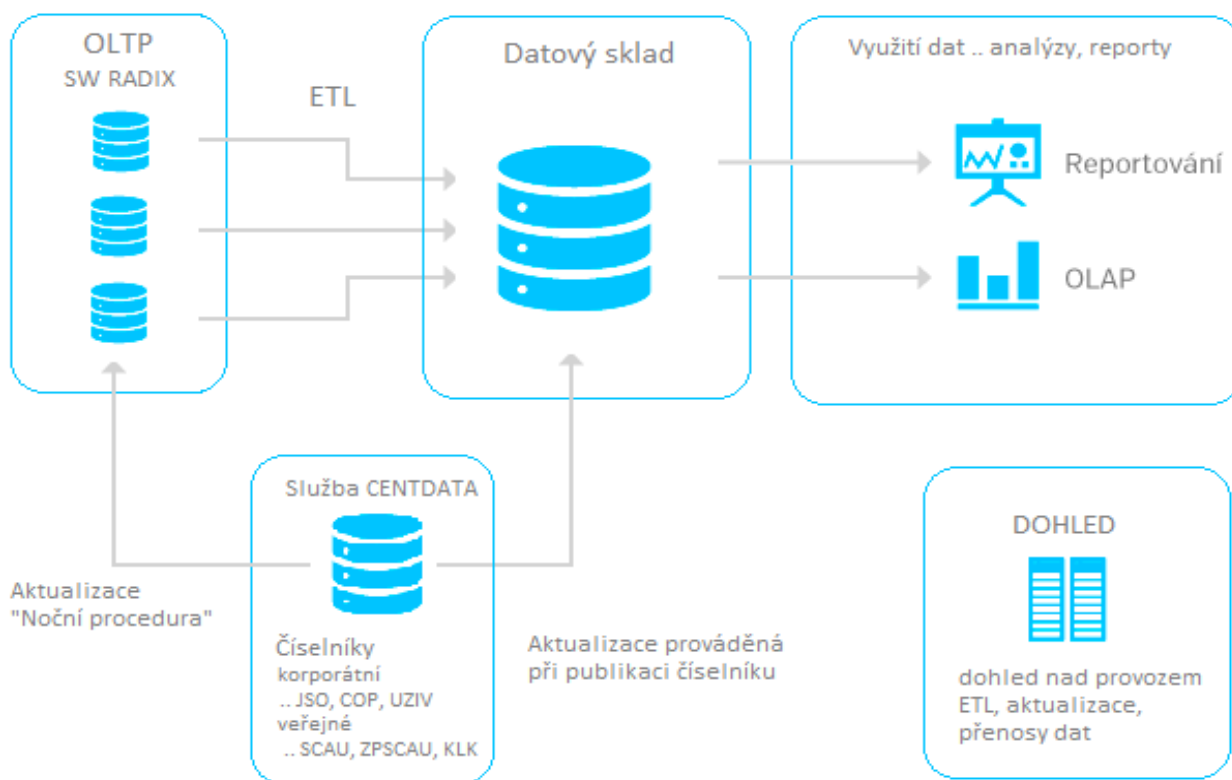
Z pohledu rozsahu požadavků a jejich případných změn lze konstatovat, že požadavky jsou určeny, důkladně prodiskutovány a podrobeny zkoumání v rámci provedení „mapování požadavků“, konzultace a odsouhlasení s vedením APC i IT oddělením APC. Z časového hlediska je požadována rychlá realizace v horizontu 3-4 měsíců. Volbu metody významně ovlivňuje i fakt, že budou zpracovávána data z jednoho typu produkčního systému.

Po uvážení výše popsaných faktů je pro realizaci vybrána „*Metoda velkého třesku*“.

5.2 Volba řešení z pohledu architektury

Pro realizaci infrastruktury pro podporu rozhodování v řetězci APC je zvolena varianta vybudování infrastruktury BI s datovým skladem, využití technologie OLAP a v budoucnosti případně implementovat reportingový systém. V systému bude figurovat i současně provozovaná služba CENTDATA. Nově bude zavedena služba DOHLED.

V následujícím schématu je zachycena zvolená architektura řešení BI v řetězci APC.



Obr. 12: Architektura řešení BI v řetězci APC [zdroj: vlastní]

5.3 Procesy ETL

Procesy ETL jsou řešeny v jednotlivých částech následovně

- **Extrakce + Transformace** probíhá na provozovně v rutinách, které budou součástí speciálně připraveného modulu **RX_uploader**, který provede přípravu a kontrolu dat a následně vytvoří aktualizací balíček dat, který je odeslán ke zpracování
- **Loading** probíhá přímo načtením balíčku dat do datového skladu tzn. nebude realizováno dočasné úložiště

5.3.1 Provedení ETL

Časově je provedení procesu ETL zařazeno do pravidelné „noční procedury“. Po provedení sehrání dat do datového skladu bude odeslána notifikace o výsledku tohoto úkonu do aplikace DOHLED, která slouží ke sledování provozu infrastruktury BI. V případě neúspěchu nebo jiného identifikovaného problému je navíc odeslána informace na IT oddělení tak, aby bylo možné na vzniklý problém operativně reagovat.

5.3.2 Rozsah dat - časově

Rozsah dat sehrávaných v rámci ETL procesů z OLTP do datového skladu bude omezen na aktuální období, kterým je poslední měsíc. Při volbě tohoto omezení se vychází z faktu, že ke změnám v uzavřených obdobích (měsících) nedochází a není třeba je opakovaně sehrávat. V rámci aktuálního měsíce ke změně v provedených obchodních transakcích dojít může. Při měsíční uzávěrce, která je na provozovně prováděna za ukončený měsíc na počátku nového měsíce je provedeno poslední sehrání uzavřeného měsíce do datového skladu a dále do ETL vstupuje pouze aktuální měsíc.

5.3.3 Rozsah dat – obsahově

Rozsah dat sehrávaných z OLTP v rámci ETL procesů bude obsahem přizpůsoben potřebám analytických výstupů a reportů. Datový balíček odesílaný z produkční databáze noční procedurou budou obsahovat pouze údaje potřebné k zachycení transakcí v databázi faktů tzn. údaje o vazbách do dimenzí a metriky.

5.3.4 Ochrana osobních a citlivých dat

Produkční systémy SW RADIX v transakčních databázích evidují i citlivé údaje. Jedná se o kombinace údajů *rodné_číslo*, *jméno_příjmení_pacienta*, *vydaný_léčivý_přípravek* nebo *zdravotnický_prostředek*, *diagnóza* a další. Provozovny APC jsou zdravotnickými zařízeními a legislativa evidenci kombinací těchto typů údajů v informačních systémech těchto zařízení povoluje.

Do datového skladu nebudou nahrávána taková data nebo kombinace dat, které by mohly vést ke spojení konkrétní osoby, vydaného zboží a diagnózy. Pro potřeby některých analýz je nutné některé vyjmenované údaje v datovém skladu evidovat. Příkladem může být potřeba vyhodnotit vydané léčivé přípravky podle diagnóz, věku nebo pohlaví.

Vždy bude postupováno tak, že už při přípravě dat (Extrakce a Transformace) pro odeslání transakcí do datového skladu prostřednictvím noční procedury proběhne anonymizace dat formou nahrazení posledních 4 číslic údaje *rodné_číslo* hashem (předpokládá se využití funkce CRC), kdy vstupem pro výpočet je nahrazovaná část. Jiné údaje, které by mohly vést ke ztotožnění pacienta např. *jméno* a *příjmení*, se neodesílají. Tak bude zajištěna možnost data analyzovat a zároveň dodržena pravidla pro ochranu osobních a citlivých dat.

5.4 Služba CENTDATA

Důležitou součástí budované infrastruktury jsou zdroje dat, které bude poskytovat služba CENTDATA. Služba publikuje data produkčním systémům SW RADIX a data budou nahrávána a využívána i v datovém skladu.

Vybrané tabulky do datového skladu takto publikovaných a nahraných dat využijí struktury OLAP při definici a tvorbě dimenzí. Tím je zajištěno, že data klíčová pro tvorbu dimenzí v OLAP databázi jsou v datovém skladu a pro práci v produkčních systémech tožná a v konzistentním stavu.

Služba CENTDATA publikuje korporátní číselníky a definice

- Jednotný sortiment JSO
- Číselník obchodních partnerů COP
- Číselník uživatelů CUZIV
- Definice cenotvorby

a veřejně přístupné číselníky

- Seznam cen a úhrad léčivých přípravků SCAU [13]
- Seznam cen a úhrad zdravotnického materiálu ZPSCAU [14]
- Seznam registrovaných léčivých přípravků KLK [12]
- Číselník anatomicko-terapeutických skupin ATC [15]

Veřejně přístupné číselníky jsou do aktualizací poskytovaných službou CENTDATA zahrnuty z důvodu zajištění dostupnosti a jednotného přístupu k aktualizacím.

5.5 Datový sklad

Zdrojem dat datového skladu jsou

- produkční systémy OLTP provozoven tzn. SW RADIX
- centrálně spravované číselníky a definice poskytované službou CENTDATA

Vzhledem k homogenitě produkčního prostředí (1 typ OLTP), struktury a rozsahu zpracovávaných dat, nebude budovaná infrastruktura BI obsahovat dočasná úložiště. Proces ETL bude ve výše popsaném rozsahu a čase nahrávat data přímo do datového skladu.

5.6 OLAP databáze

OLAP databáze řetězce APC bude realizována technologií MOLAP (Multidimensional On Line Analytical processing).

Jako zdroj dat pro OLAP databáze budou použity výhradně tabulky uložené v datovém skladu. V podrobnější kapitole níže v textu je návrh tabulky faktů a dimenzí a popis některých detailů navrhovaného řešení.

5.7 Reporting

Předpokládá se, že v budoucnu proběhne implementace nástrojů pro reporting. Jako datová základna bude použit budovaný datový sklad. Je uvažováno o *standardním* i *ad-hoc* reportingu [6, s.34].

Standardní reporting bude velmi pravděpodobně poskytovat výstupy (z pohledu obsahu), které bude možné získat i z OLAP databází, proto ve fázi přípravy i při jeho implementaci bude možné použít návrhy, znalosti i zkušenosti z implementace OLAP. Předpokládá se, že reporting pokryje potřeby uživatelů, kdy není třeba nebo není možné (např. uživatel není pro takovou práci proškolen) pracovat s OLAP kostkou, ale kdy postačuje pravidelný výkaz. Příkladem může být „nekomplikovaný“ výkaz pro měsíční vyhodnocení výkonů zaměstnanců pro vedoucího skupiny provozoven.

Ad-hoc reporting bude zaměřen na řešení potřeb uživatelů v různých speciálních, jednorázových nebo jinak specifických případech. Reporty bude na základě požadavků uživatelů poskytovat pověřená osoba (datový analytik). Příprava *ad-hoc* reportů bude obvykle probíhat prostřednictvím dotazů do datového skladu formulovaných v některém DML např. Transact-SQL.

5.8 Aplikace DOHLED

Aplikace DOHLED bude nezbytnou součástí budovaného řešení BI. V produkčním provozu složitějších systémů je nutné mít k dispozici nástroje a nastaveny postupy, které umožní operativně reagovat na vzniklé nestandardní provozní situace nebo problémy. V oblasti dohledu nad provozem se jedná především o efektivní předávání informace od místa vzniku problému ke kompetentnímu řešiteli a o důslednou evidenci nastalých problémů a provozních situací.

6. Návrh dimenzí a tabulky faktů

Předpokládá se, že finální infrastruktura BI bude tvořena z více samostatných resp. sdílených definic dimenzí a tabulek faktů. To přirozeně plyne z faktu, že jednotlivá oddělení mají své specifické nároky na analýzu produkčních dat, mohou mít jinou míru abstrakce nad daty i jiné nároky na volby pohledů na data.

Tato kapitola obsahuje návrh některých dimenzí a tabulky faktů obchodních transakcí typu

- *Pokladní prodej*, takto se označují prodeje a výdeje zboží zákazníkům a pacientům v rámci maloobchodního prodeje přes pokladnu
- *Výdej na žádanku*, takto se označuje prodej zboží s následnou fakturací lékařům, zdravotnickým zařízením, ostatním organizacím a soukromým osobám

Tyto obchodní transakce jsou souhrnně označovány jako „*Prodejní transakce*“.

6.1 Návrh dimenzí „*Prodejních transakcí*“

K dimenzionálnímu modelování je možné přistoupit, pokud jsou shromážděny požadavky uživatelů na analytické výstupy a zmapovány produkční systémy [6, s.108].

Návrh dimenze určuje následující charakteristiky [6, s.109]:

- Identifikace dimenze
- Plný název dimenze
- Hierarchická struktura dimenze
- Prvky dimenze
- Počty prvků dimenze na hierarchických úrovních, pokud je lze určit
- Zdroj dat dimenze, např. tabulka nebo tabulky datového skladu

Plánované řešení BI bude pro analýzu prodejních transakcí obsahovat následující dimenze:

- Čas – základní dimenze pro analýzu dat
- Produkt – důležitá strukturovaná dimenze, umožňuje analyzovat transakce podle jednotlivých produktů, typů zboží, skupin korporátní nomenklatury, výrobce, nomenklatury ATC

- Provozovna – dimenze pro sledování místa provedení prodejních transakcí
- Dodavatel – dodavatel zboží až na úroveň střediska dodavatele
- Odběratel – odběratel zboží až na úroveň střediska odběratele
- Prodejní kanál – typ prodeje a výdeje
- Předepisující lékař – analýza výdeje léků na recept a zdravotnického materiálu na poukaz
- Expedient – uživatel produkčního systému, který provedl prodejní transakci
- Zákazník – prodejní transakce typu *Prodej přes pokladnu* je standardně anonymní, v případě použití karty věrnostního systému se transakce přiřazuje k věrnostnímu kontu, obvykle není třeba zkoumat transakce až na úroveň konkrétního zákazníka, ale je třeba zkoumat celkové objemy prodejů podle rozdělení
 - a) transakce mimo věrnostní systém,
 - b) transakce v rámci věrnostního systému
 - c) transakce podrobně podle typu věrnostního systému

a případné další

V rámci této práce jsou pro podrobnější popis vybrány dimenze *Čas*, *Produkt*, *Prodejní kanál* a *Expedient*.

Tabulka č.3 Návrh dimenze *Čas*

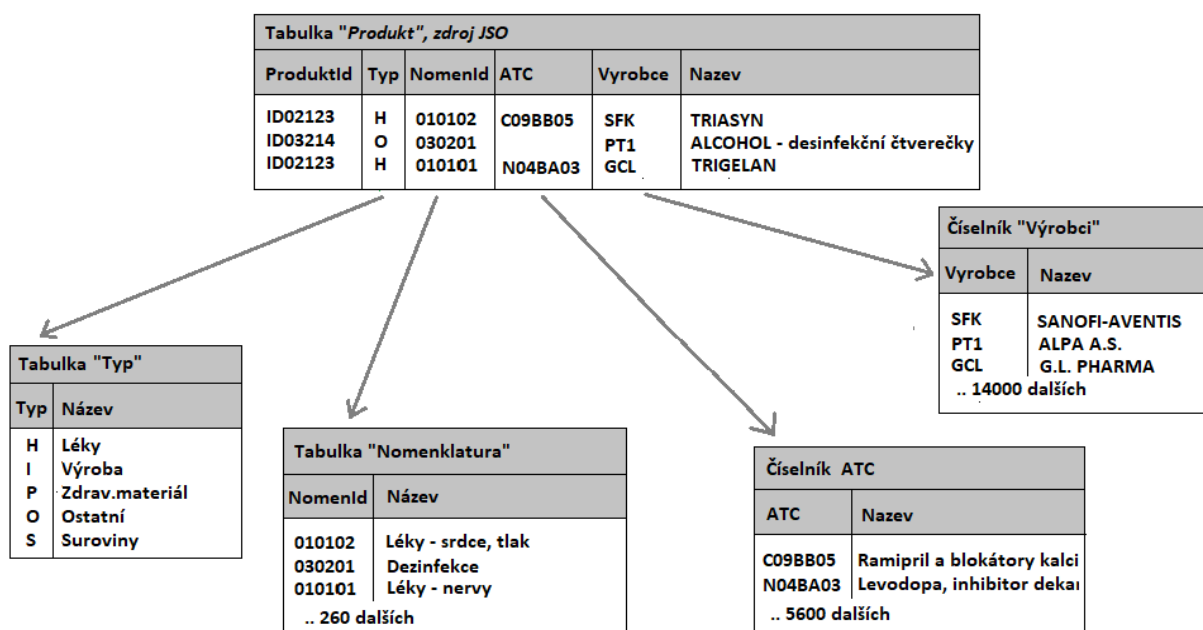
Identifikace	Dim_Cas
Plný název	Čas
Hierarchie	Rok, Měsíc, Den, Hodina, Minuta Zařazení hierarchické úrovně <i>Hodina</i> , <i>Minuta</i> je nutné pro potřeby analýzy transakcí <i>Prodej přes pokladnu</i> , kdy je ve vybraných výstupech sledováno vytížení provozoven podle denní doby. Analýza slouží k pravidelné kontrole obrátu ve vztahu k otevírací době pro její případnou úpravu.
Schéma	Hvězdicovité

Dimenze *Čas* není reprezentována tabulkou, na kterou by bylo možné se prostřednictvím konkrétní hodnoty odkazovat, není to třeba. Klíčem dimenze v tabulce faktů je strukturovaný údaj, který má význam a který svojí hodnotou odkazuje na časovou osu v požadované míře detailu rozlišení.

Tabulka č.4 Návrh dimenze *Produkt*

Identifikace	Dim_Produkt
Plný název	Produkt
Hierarchie	<p>Produkt – sortimentní položka, zboží</p> <p>Typ – základní členění sortimentu na 5 typů (léky, zdravotnický materiál, výrobu v laboratoři, suroviny a ostatní zboží)</p> <p>Nomenklatura – zařazení produktu do korporátní nomenklatury</p> <p>ATC skupina - určuje příslušnost léku k anatomicko-terapeutické skupině ATC, umožňuje např. vybrat jen léky pro léčbu diabetu nebo vysokého krevního tlaku, apod, je určena veřejně dostupnými číselníky</p> <p>Výrobce – číselník výrobců</p>
Prvky	Položky číselníku JSO
Zdroj	Datový sklad, číselník JSO a tabulky Vyrobcce, Nomenklatura, ATC
Schéma	„ <i>Sněhová vločka</i> “

Dimenze „*Produkt*“ je jedním z klíčových hodnotících hledisek analytického zpracování dat prodejních transakcí. Dimenze je strukturovaná a je tvořena několika relačně propojenými tabulkami. Její datová reprezentace je zachycena v následujícím schématu.



Obr. 13: Datová reprezentace dimenze *Produkt*

Tabulka č.5 Návrh dimenze *Prodejní kanál*

Identifikace	Dim_ProdejniKanal
Plný název	Prodejní kanál
Hierarchie	Prodejní kanál, 1 úroveň
Prvky	<i>Výdej přes pokladnu</i> REC, SREC – na recept a storno výdeje na recept POU, SPOU – na poukaz a storno výdeje na poukaz VP, SVP – volný prodej a storno volného prodeje <i>Výdej na žádanku</i> ZAD, SZAD – výdej, vratky výdeje, následná fakturace
Zdroj dat	Datový sklad, tabulka „ <i>Prodejní kanál</i> “
Schéma	Hvězdicovité

Tabulka č.6 Návrh dimenze *Expedient*

Identifikace	Dim_Expedient
Plný název	Expedující
Hierarchie	Expedient, 1 úroveň
Prvky	Uživatelé produkčního systému
Zdroj	Datový sklad, číselník CUZIV
Schéma	Hvězdicovité

6.2 Návrh tabulky faktů „*Prodejních transakcí*“

Tabulka faktů obsahuje logicky související hodnoty metrik (ukazatelů). Sloupce tabulky faktů jsou klíčovými položkami, které realizují vazbu do dimenzí, nebo hodnotami metrik. Záznamy tabulky faktů jsou přiřazovány k nejnižší úrovni detailu ve strukturách dimenzí. [6, s.115].

6.2.1 Granularita tabulky faktů

Granularitou je určena podrobnost údajů uložených v tabulce faktů a přímo závisí na podrobnosti dimenzí souvisejících s touto tabulkou faktů. Nízká úroveň granularity neumožňuje pracovat s daty detailně.

Obecná doporučení pro řešení úrovně granularity v datovém skladu jsou [6, s.116]

- data ukládat s nejvyšší možnou granularitou
- transformovat granularitu dat z různých zdrojů na stejnou úroveň

V návrhu dimenzí je u dimenze *Čas* určeno, že součástí dimenze je i hodina a minuta, zároveň je zavedena (kromě jiných) i dimenze *Expedient, Prodejní kanál*. Z takto navržených dimenzí plyne, že není možné fakta o prodejních transakcích v této části řešení agregovat, tzn. musí být zachována vysoká granularita a to až na úroveň jednotlivé transakce.

6.2.2 Zdroje dat v produkčním systému

Zdrojem dat pro tabulku faktů *Prodejní transakce* jsou v produkční databázi tabulky *Vydejka* záznamy typu *Expediční případ*, tabulka *Skladové pohyby* a tabulka *Pokladna*.

Pro jinou tabulku faktů, v níž bude probíhat analýza jiné oblasti aktivit provozoven např. pro činnosti spojené s *Výrobou v laboratoři* budou zdrojem dat v produkční databázi tabulky *Vydejka* záznamy typu *Výroba v laboratoři* a tabulka *Skladové pohyby*, tzn. data z jedné zdrojové tabulky produkčního systému mohou figurovat v tabulkách faktů řešících výstupy pro různé oblasti, a to v různém rozsahu i míře abstrakce.

6.2.3 Struktura tabulky faktů

Následující tabulka obsahuje strukturu tabulky faktů prodejních transakcí.

Tabulka č.7 Návrh tabulky faktů pro analýzu prodejních transakcí

Dim/Metrika	Název	Popis	Příklad / Data					
			každý sloupec reprezentuje 1 záznam tabulky faktů					
DIM	ProvozovnaId	FK číselníku provozoven	618	606	606	621	621	618
DIM	Cas	Čas transakce	20211009	20211009	20211009	20211009	20211009	20211009

			0831	0833	0829	0832	0832	0828
DIM	ProduktId	FK do JSO	1023	2056	1245	5645	5645	3265
DIM	Expedient	FK do číselníku CUZIV	221	228	228	195	195	136
DIM	ProdejniKanal	Prodejní kanál	REC	REC	POU	ZAD	ZAD	VP
DIM	OdberatelId	FK do číselníku COP	null	null	null	1033	1033	null
DIM	DodavatelId	FK do číselníku COP	1002	1008	1008	1002	1002	1008
DIM	Rodne_cislo	Rodné číslo pacienta 6-míst zleva + hash zbytku	660702 +hash	730906 +hash	891205 +hash			
Metrika neaditivní	CisloPor	Interní číslo obchodního případu	PO21618 05462	PO21618 023142	PO21606 003654	ZA21621 00452	ZA21621 00452	PO21618 02354
Metrika aditivní	Mnozstvi	Vydané množství	2	3	1	2	1	10
Metrika neaditivní	SazbaDPH	Sazba DPH %	10	10	15	21	21	0
Metrika neaditivní	JednPCbDph	Jednotková prodejní cena bez DPH	100	200	100	200	200	100
Metrika neaditivní	JednPCsDph	Jednotková prodejní cena s DPH	110	220	115	242	121	100
Metrika aditivní	CelkPCbDph	Celková prodejní cena bez DPH	200	600	100	400	200	1000
Metrika aditivní	CelkPCsDph	Celková prodejní cena s DPH	220	660	115	484	242	1000
Metrika neaditivní	JednDoplatek	Jednotkový doplatek pa- cienta	50	20	65	null	null	0
Metrika aditivní	CelkDoplatek	Celkový doplatek pacienta	50	60	65	null	null	1000
Metrika neaditivní	JednUhrZP	Jednotková úhrada zdravotní pojišťovny	60	200	50	null	null	0
Metrika aditivní	CelkUhrZP	Celková úhrada zdravotní pojišťovny	60	600	50	null	null	1000

6.3 Další řešené oblasti

Do finálního řešení BI se předpokládá zavedení dimenzí a tabulek faktů pro zkoumání dalších oblastí provozu a to především:

- Elektronické a listinné recepty, recepty na psychotropní látky a poukazy, problematika bude řešena včetně vykazování a fakturace péče na zdravotní pojišťovny

- Příjem zboží
 - Fakturace za dodané zboží
 - Výroba v laboratoři
- a další

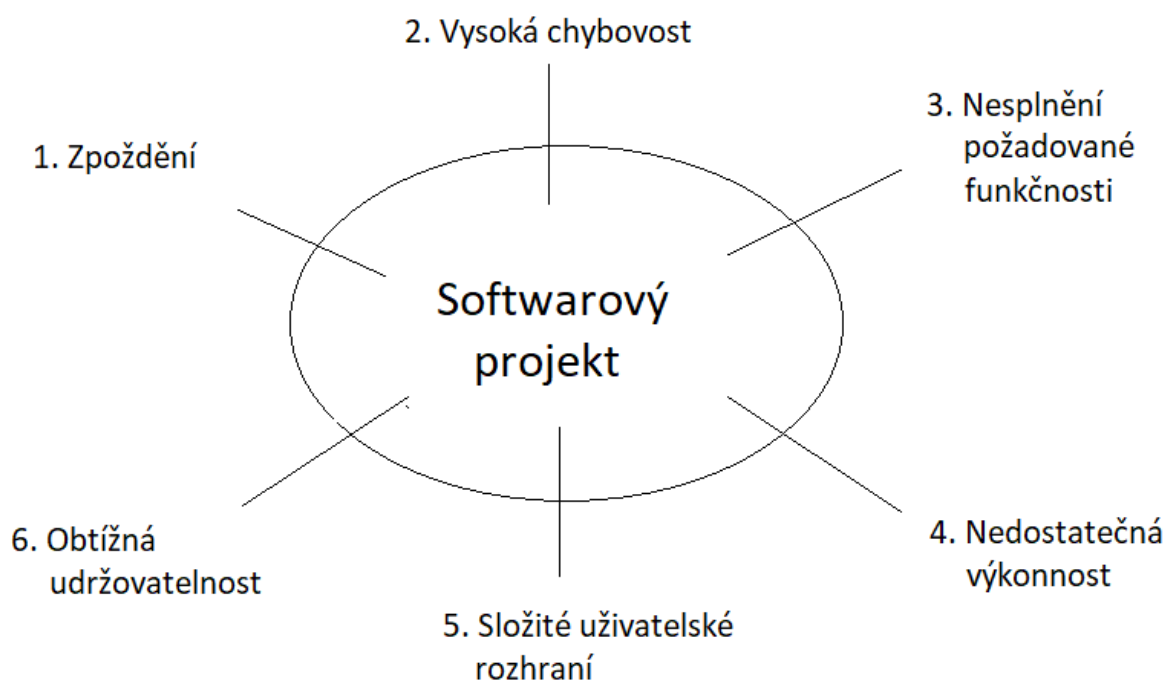
Je pravděpodobné, že dimenze definované pro jednu oblast bude možné použít i v jiných oblastech tzn. dimenze bude společná pro více tabulek faktů (příkladem je dimenze *produkt*), a také je možné, že tabulka faktů jedné oblasti bude sloužit jako zdroj dat v oblasti jiné.

Struktura řešení bude odpovídat spíše než samostatně stojícím tabulkám faktů se samostatnými dimenzemi schématu, které popisuje Tvrdíková jako schéma „*Soubor hvězd*“ – constellation [10, s.104], kdy do modelu bude začleněno více tabulek faktů, k nim jsou definovány dimenze, které jsou tabulkami faktů sdíleny.

7. Rizika a problémy implementace

V kapitole jsou popsána některá rizika a problémy, které se mohou při implementaci vyskytnout. Je zde i navrženo, jak vzniklou situaci řešit, případně zda je riziko vzniku problému reálné.

Na obrázku 14 jsou zobrazeny nejčastější problémy softwarových projektů tak, jak je zformuloval Paleta [7, s.15]



Obr. 14 : Nejčastější problémy softwarových projektů

V dalším textu je řešení některých příčin potenciálního neúspěchu projektu BI podrobněji vysvětleno.

Implementace metodou „Velkého třesku“ (potenciální problém „1. Zpoždění“)

Pro implementaci BI byla zvolena metoda „Velkého třesku“, kdy vybudování řešení proběhne v řádu měsíců. Pokud budou úspěšně ukončeny přípravné kroky (sjednocení číselníků JSO, COP) jako nutná podmínka dalšího postupu realizace projektu, zpoždění na straně implementace se nepředpokládá.

Nesoulad mezi daty v produkčních systémech, datovém skladu a výstupy z BI (potenciální problém 2. *Vysoká chybovost*, 3. *Nesplnění požadované funkčnosti*)

Nutnou součástí implementace systému je validace výstupů z OLAP tak, aby bylo ověřeno, že procesy ETL správně data extrahují, transformují a načítají do datového skladu a výstupy OLAP odpovídají skutečnosti (transakcím) v produkčních systémech.

Validace výstupů bude nutnou součástí akceptace řešení BI na straně uživatele a bude probíhat i po náběhu nového systém během souběhu nového a původního systému přípravy analytických výstupů.

Pro provedení kontroly je možné použít dvě doplňující se metody

- a) Porovnání obsahem stejných nebo podobných výstupů z produkčních systémů a OLAP
- b) Provést kontrolní agregované i neagregované dotazy prostřednictvím
 - T-SQL nebo jiného DML v produkčním systému a v datovém skladu
 - MDX³ v OLAP databázi

Metodu a) bude možné použít ke komparaci reportů nejen v rámci přípravy řešení, ale také po dobu přibližně 6 měsíců při souběhu původního systému pro přípravu reportů a nového řešení BI.

Problém s odezvou, provoz systému ve vztahu k velikosti dat (potenciální problém 4. *Nedostatečná výkonnost*)

Při návrhu datových struktur datového skladu, tabulky faktů a dimenzí pro OLAP, výběru hardwarových i softwarových komponent je třeba věnovat velkou pozornost správné volbě všech částí budované infrastruktury a to s ohledem na předpoklad, že systém bude velmi pravděpodobně fungovat mnoho let, počet prodejních a skladových transakcí provozoven v jednom roce je přibližně 5 000 000, tzn. velikost dat bude mít při nevhodném návrhu na odezvu velmi pravděpodobně negativní vliv.

Problém s kvalitou dat, historická data [2, s.120]

(problém 2. *Vysoká chybovost*, 3. *Nesplnění požadované funkčnosti*)

3 MDX znamená *MultiDimensional eXpressions*. Jedná se o dotazovací jazyk do OLAP kostek, využívá technologii analytických služeb SQL Serveru.

Kvalitu dat produkčních systémech SW RADIX je možné hodnotit jako dostatečnou. Extrakce a transformace dat bude probíhat v produkčním systému při přípravě dat na sehrání do datového skladu. Zde je možné zachytit a přímo korigovat nebo notifikovat správci systému prostřednictvím aplikace DOHLED případný problém.

Předpokládá se, že do budovaného řešení proběhne sehrání i dat historických za roky 2019 a 2020 tak, aby bylo možné již nad daty současného roku vyhodnotit případné trendy a provést komparace výsledků obchodních činností za různá období. Laberge u historických dat upozorňuje [2, s.94] na možné problémy s jejich kvalitou. V řešeném případě je již 10 let používán produkční systém SW RADIX. Historická data má IT oddělení řetězce APC plně pod kontrolou včetně možnosti jejich případné korekce nebo doplnění.

Uživatelské rozhraní

(problém 5. *Složitě uživatelské rozhraní*)

Pro prezentaci analytických výstupů je v současnosti zvolen MS Excel, který je v řetězci APC používán jako standardní součást softwarového vybavení. V budoucnosti je plánován vývoj webové aplikace pro prezentaci výsledků analýz dat.

8. Implementace, volba platformy a dodavatele, finální řešení

8.1 Úvod

Implementace technologie BI v konkrétní firemní struktuře může probíhat mnoha způsoby, které reflektují na kontext organizace tzn. na konkrétní podmínky organizační, provozní, technologické, personální a další. V rámci implementace je nutné řešit a volit varianty řešení různých oblastí, mezi základní je možné zařadit oblast volby dodavatele řešení, volby dle umístění datového skladu, volby dle dohledu nad aplikací BI.

Volba dle dodavatele řešení

- 1) interní, oddělení vývoje a IT, konzultace s externí autoritou
- 2) externí dodavatel, úzká spolupráce s oddělením vývoje a IT
- 3) externí dodavatel, plně v jeho kompetenci, schéma zadání→dodané řešení

Volba dle umístění datového skladu

- a) interní infrastruktura, provozováno APC (HW a SW, hosting, cloud)
- b) externí infrastruktura, provozovaná externím dodavatelem

Volba dle dohledu nad aplikací BI

- i) interní, oddělení IT
- ii) externí, dodavatel

Poznámka : Výčty jistě neobsahují všechny možnosti.

Počáteční úvahy a diskuse o využití technologie BI v řetězci APC, které byly přirozeně následované průzkumem a seznámením se s možnými variantami řešení, byly postupně rozpracovány a vyústily do rozhodnutí implementaci realizovat ve variantě 1) + a) + i) tzn. nasadit interní kapacity oddělení vývoje a IT pro návrh, přípravu a vývoj s případnou pomocí vybrané konzultační firmy a provozovat BI na interní hardwarové a softwarové infrastruktuře s dohledem interního oddělení IT nad provozem.

V současné době jsou pro implementaci BI dostupné různé softwarové nástroje, systémy a platformy (dále jen „platformy“). Platformy poskytují široké možnosti v oblasti analýzy dat, vizualizace výsledků analýz a reportingu. Konektivita k datům je řešena vel-

mi komplexně, zdrojem dat mohou být SQL databáze, tabulky MS Excel, textové soubory, podnikové informační systémy a další.

V konkrétních případech řešení BI lze systémy i kombinovat např. datový sklad může být řešen na platformě *Oracle Database* a jako analytické nástroje použity funkce platformy *SQL Server Business Intelligence*.

8.2 Popis a základní charakteristika vybraných platform

Jako platformy pro implementaci BI ve variantě 1)+a)+i), byly uvažovány 2 tradiční systémy BI *Microsoft SQL Server Business Intelligence* a *Oracle Business Intelligence* a zástupce self-service BI *QlikView*.

8.2.1 SQL Server Business Intelligence

SQL Server Business Intelligence je platforma vyvíjená firmou Microsoft

Platforma umožňuje řešit všechny části implementace BI [4, str. 21].

- **Integrační služby** poskytuje komponenta *SQL Server Integration Services* („*SSIS*“), která umožňuje získat data z různých typů systémů, v případě potřeby je transformovat a integrovat
- **Datový sklad** (úložiště, repozitář dat) realizuje relační databázový server *SQL Server DBMS*
- **Analytické služby** poskytuje komponenta *SQL Server Analysis Services* („*SSAS*“)
- **Reportovací služby** poskytuje komponenta *SQL Server Reporting Services* („*SSRS*“), zpřístupňuje uživatelům data a výsledky analýz ve formě reportů

Pro práci s jednotlivými částmi řešení BI jsou dostupné nástroje :

SQL Server Management Studio („*SSMS*“) je IDE pro návrh, implementaci a provoz datového skladu a pro návrh, implementaci a využití *SSIS*, *SSAS*, *SSRS*

Business Intelligence Development Studio („*BIDS*“) je IDE pro práci s *SSAS* a *SSRS*.

8.2.2 Oracle Business Intelligence

Oracle Business Intelligence je platforma, která umožňuje řešit jednotlivé oblasti implementace BI

- **Integrační služby** pro vytváření, správu a údržbu procesů integrace dat produkčních systémů poskytuje nástroj **Oracle Data Integrator**
- **Datový sklad** realizuje relační databázový server **Oracle Database**
- **Analytické služby** poskytuje multidimenzionální analytický engine **Oracle OLAP**
- **Reportovací služby** jsou realizovány systémem **Oracle Reports Services**

8.2.3 QlikView

QlikView je velmi rozšířená platforma pro analýzu dat, je zástupcem self-service nástrojů. Oproti tradičním nástrojům BI, které fungují na principu vytváření vybraných a konkrétních reportů pro určité skupiny uživatelů, platforma QlikView vytváří obecný model reportů, který si každá skupina sama přizpůsobí vlastním potřebám. Samotná platforma QlikView v sobě zahrnuje tři komponenty

- **QlikView Desktop** slouží k extrakci dat a vytvoření obsahu
- **QlikView Server** umožňuje provádět in-memory analýzy, umožňuje komunikaci mezi QlikView uživatelem a serverem
- **QlikView Publisher** zabezpečuje přímé načítání datových zdrojů prostřednictvím definovaných spojení

8.2.4 Porovnání, výběr platformy

Všechny platformy porovnávané v této kapitole umožňují vybudovat řešení BI v plánované struktuře (viz kapitola 5.2) a variantě implementace 1)+a)+i) . V porovnání nejsou uvažovány finanční aspekty, je hodnocena pouze vhodnost platformy ke splnění plánovaných cílů.

Při volbě platformy pro realizaci BI vstupují do rozhodování následující předpoklady :

- Ve velké míře bude používána připravená množina vybraných reportů, u kterých bude možné nastavit výběr dat prostřednictvím zadaných hodnot dimenzí a nastavit míru abstrakce a detailu výstupu.
- Frekvence zavádění nových reportů nebo potřeba přípravy inovativních pohledů na data nebude častá.
- Realizaci změn (úpravy stávajících i nových výstupů) bude mít plně v kompetenci IT oddělení, nepředpokládá se účast uživatelů, proto do úvahy přicházejí spíše tradiční platformy BI.

- Aplikace i serverové systémy používané a provozované v řetězci APC jsou (kromě několika výjimek) z portfolia produktů firmy Microsoft a to aplikace běžné, licencované (MS Windows, MS Office, MS SQL Server a další) i aplikace vyvíjené (.NET Core). Produkty Microsoft jsou v APC preferované, zkušenosti s provozem a správou jsou na dobré úrovni.

Popsané okolnosti významně ovlivňují výběr platformy pro implementaci. Lze konstatovat, že v situaci, kdy nebyly nalezeny důvody pro vyřazení *SQL Server Business Intelligence* z výběru, preference produktů Microsoft jasně určuje volbu této platformy.

8.3 Jak to všechno dopadlo, finální řešení implementace BI

Postupem času byla některá původní zadání, podmínky a předpoklady s ohledem na organizační, technologické, časové i finanční aspekty změněny a původní schéma (kapitola 5.2 a na Obr. 12) implementace BI s realizací analytických nástrojů technologií OLAP a s výhledovou implementací reportingového systému přehodnoceno. V upravené variantě je jako hlavní a jediný analytický nástroj zvolen reportingový systém, přičemž o využití nástrojů OLAP je uvažováno v budoucnosti. Zároveň došlo ke změně a výběru varianty řešení 2)+b)+ii) tzn. pro návrh, přípravu a vývoj řešení BI využít externího dodavatele, provozovat řešení na hardwarové a softwarové infrastruktuře externího dodavatele, dohled je realizován dodavatelem řešení

Důvody pro změnu lze stručně popsat takto

- Umístění datového skladu není v případě zajištění výkonnostních požadavků, provozních a bezpečnostních hledisek rozhodující
- Realizace interními kapacitami nemusí vést k požadovanému výsledku, již ve fázi návrhu a přípravy bylo nutné řešit různé nejasné situace a procesy
- Byl nalezen externí dodavatel, který byl shledán vhodným (finančně, kompetenčně, kapacitně)
- Proběhly prezentace a ukázky řešení reportingového systému v maloobchodním prostředí srovnatelného typu, které byly shledány jako plně postačující požadavkům a potřebám uživatelů

8.3.1 Výběr finálního řešení a jeho realizace

Na základě popsané změny varianty bylo k implementaci vybráno řešení BI dodavatele REPORTING.CZ, s.r.o., který poskytuje aplikaci pokrývající celý reportovací cyklus :

- Získání dat z produkčních systémů, jejich konsolidaci, vyčištění a transformaci, nahrání dat do datového skladu
- Uložení dat do datového skladu, který je provozován v rámci infrastruktury dodavatele
- Prezencí koncovým uživatelům prostřednictvím webové aplikace REPORTING.CZ

Důležitým aspektem pro volbu řešení *Reporting.cz* byla skutečnost, že bylo možné využít většinu předchozí práce, výsledků a výstupů z kroků provedených v rámci přípravy na původně plánovanou variantu 1)+a)+i) a to :

- Zavedení jednotných číselníků JSO, COP
- Seznam reportů z provedené inventarizace včetně detailů a dopřesnění
- Návrh dimenzí a tabulek faktů
- Služby aplikace CENTDATA pro publikaci datových zdrojů
- Scénář a část nástrojů na sehrávání dat prostřednictvím „noční procedury“
- Řešení *Reporting.cz* je nasazeno a provozováno u více než 600 korporací [16]

Následující tabulka obsahuje porovnání varianty původně plánované a realizované.

Tabulka č.8 Porovnání varianty řešení BI původně plánované a realizované

	Původní řešení varianta 1)+a)+i)	Realizované řešení varianta 2)+b)+ii) Reporting.cz
Proces ETL	Interní Aplikace: RX_uploader Spuštění - provozovna	Externí - Reporting Aplikace: Uploader Spuštění - provozovna
Provoz datového skladu	Interní Infrastruktura: APC Technologie: MS SQL server	Externí Infrastruktura: Reporting.cz Technologie: v kompetenci dodavatele

Služba CENDATA	Infrastruktura: APC	Infrastruktura: APC
Prezentace analytických výstupů a reportů	MS Excel	Webová aplikace dodavatele Tabulkový formát, grafy + možné exporty do XLS, PDF, CSV

V realizovaném řešení se procedury a úkony, které zajišťuje APC pro provoz BI, zredukovaly a došlo celkově k jejich významnému zjednodušení. Zároveň byly podstatným způsobem omezeny personální i další infrastrukturní hardwarové i softwarové kapacity.

Oblasti provozu aplikace BI jsou řešeny následovně:

- **Procesy ETL**

Dodavatel poskytl modul *Reporting Uploader* pro zajištění procesu ETL, modul je konfigurován a spravován kapacitami dodavatele

APC zajišťuje spuštění modulu na provozovně a řeší konektivitu směrem k infrastruktuře dodavatele, dále poskytuje informace o změnách ve struktuře dat nebo významu údajů produkčních databází (jedná se o plánovaný vývoj aplikace SW RADIX, realizaci legislativních změn, apod.) tak, aby dodavatel mohl změny vyhodnotit a přiměřeně reagovat

- **Datový sklad**

Dodavatel poskytuje a zajišťuje provoz a správu datového skladu na vlastní infrastruktuře

APC je od zajištění provozu a správy datového skladu plně „odstíněna“

- **Analytické a reportovací služby**

Dodavatel poskytuje rozhraní (webovou aplikaci REPORTING.CZ), které zajišťuje přístup k analytickým výstupům, umožňuje definovat a připravit reporty podle zavedených dimenzí a ukazatelů, definice reportů je možné uložit, opakovaně využívat a sdílet, dodavatel poskytuje i služby pro řešení *ad-hoc* reportů,

provoz, správa a rozvoj aplikace je plně v kompetenci dodavatele

APC využívá webovou aplikaci REPORTING.CZ, po nastavení přístupových práv a zaškolení mohou uživatelé aplikaci používat

- **Služba CENDATA**

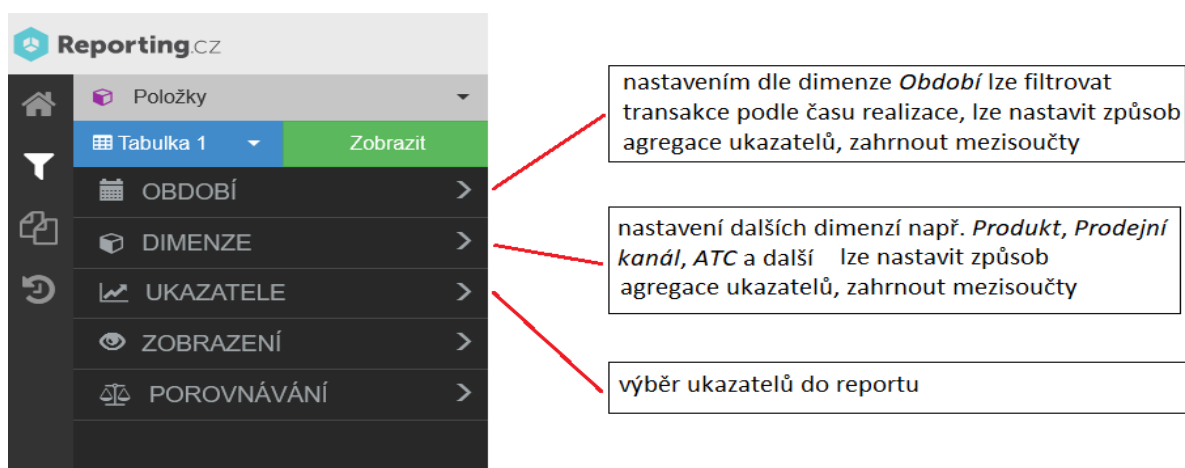
Dodavatel využívá službou poskytované zdroje dat pro zajištění provozu BI

APC zajišťuje provoz služby, která poskytuje zdroje dat prostřednictvím verzovaných publikací pro provoz produkčních systémů a infrastruktury BI dodavatele, do poskytovaných datových zdrojů jsou zahrnuty i veřejně dostupné zdroje, důvodem je jednotný způsob publikace a zajištění dostupnosti

8.3.2 Aplikace REPORTING.CZ

V následujícím textu jsou krátce popsány vybrané části aplikace REPORTING.CZ, jsou použity fragmenty formulářů aplikace a je naznačen způsob práce při přípravě reportů.

Základní rozcestník pro přípravu reportů je zobrazen na následujícím obrázku.



Obr. 15: REPORTING.CZ rozcestník definice reportu

V následujícím obrázcích jsou zobrazeny fragmenty aplikace pro nastavení dimenze *Období*, *Nomenklatury* a *ATC skupin*.

Nastavení dle dimenze "Období"

ovlivňuje i zahrnutí mezisoučtů

mezisoučet za rok 2021 nebude zahrnut

mezisoučet za 1Q 2021 bude zahrnut

mezisoučty za měsíce 1+2+3/2021 budou zahrnuty

Obr. 16: REPORTING.CZ nastavení dle dimenze „Období“ [zdroj: vlastní]

Reporting.CZ

Položky

Tabulka 1 Zobrazit

OBDOBÍ

Měsíce: I.Q 2021, II.Q 2021, III....

Vyhledat

1 2 3 4

2021

I.Q 2021

leden 2021

únor 2021

březen 2021

II.Q 2021

III.Q 2021

IV.Q 2021

2020

Reporting.CZ

Položky

Tabulka 1 Zobrazit

DIMENZE

Typ lékárny

Obchoduje

ATC

Vyhledat

1 2 3 4 5 6

<<All>>

<<Unknown>>

A Trávicí trakt

B Krev a krevet

C Kardiovasku

D Dermatologi

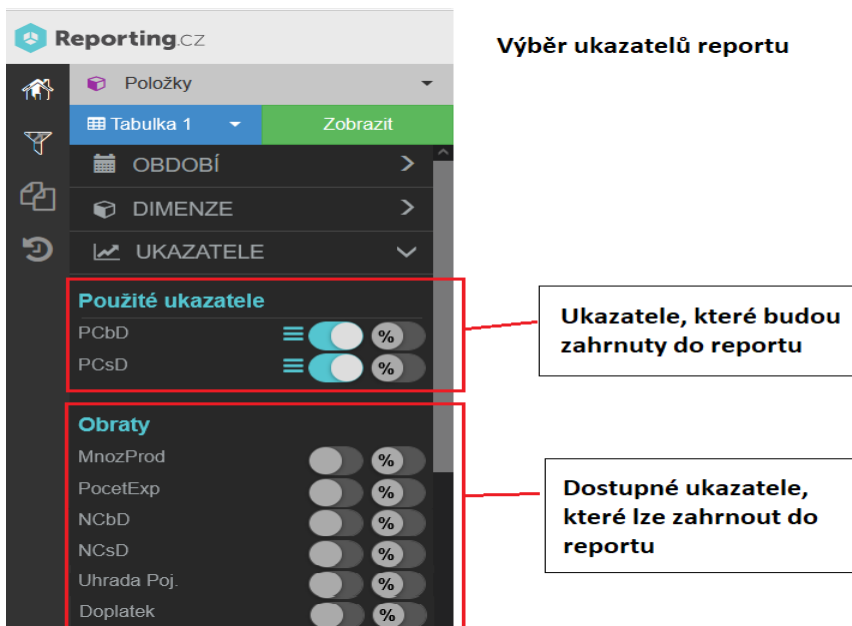
G Urogenitální

H SYSTÉMOV

J Antiinfektiva

Obr. 17 : REPORTING.CZ nastavení reportů dle dimenze *Nomenklatura* a *ATC skupina*

Výběr ukazatelů určuje jaké údaje vstupují do zpracování dat a jakým způsobem jsou data agregována, zda je zobrazena nominální nebo procentní reprezentace ukazatele.



Obr. 18 : REPORTING.CZ výběr ukazatelů reportu

Příklad výsledného reportu

Po volbě a nastavení dimenzí a výběru ukazatelů je možné tlačítkem Zobrazit spustit dotaz a report podle zadaných parametrů zobrazit. Formu výstupu (prezentace dat) je možné zvolit tlačítkem Tabulka . Výstup je možné exportovat ve formátu XLS, PDF, CSV.

Lékárna	I.Q 2021	II.Q 2021	III.Q 2021
All	572 775,64	289 596,95	637 074,48
Svatská lékárna, s.r.o.	84 076,09	8 350,23	81 776,14
Lékárna, s.r.o.	26 926,93	7 139,11	62 058,14
Těchovická lékárna, s.r.o.	42 481,89	18 730,94	98 885,99
Lékárna Brno, s.r.o.	7 756,50	811,62	55 718,91
Inter, s.r.o.	52 101,57	9 056,20	55 923,65
Lékárna U sv. Pavla, s.r.o.	41 355,45	70 979,17	28 000,00
Lékárna KAMI, s.r.o.	15 256,87	59 925,25	26 219,36
Vokovická lékárna, s.r.o.	18 979,89	32 339,57	32 378,19
Lékárna Trio, s.r.o.	8 454,02	4 900,95	22 275,86
Lékárna Soud, s.r.o.	9 853,88	13 122,57	76 905,30
Lékárna Poliklinika	2 249,96	53 651,23	67 270,66
Lékárna Pharm+	74 729,27	65 259,58	13 758,80
Art, s.r.o.	59 921,78	26 219,36	2 995,63

Obr. 19 : REPORTING.CZ příklad reportu v tabulkové formě

9. Shrnutí výsledků

Cílem práce bylo navrhnout a implementovat infrastrukturu pro vyhodnocování dat v maloobchodní síti lékáren, prodejen zdravotnického materiálu a distribučních pracovišť a to z pohledu ekonomických, sortimentních, geografických, personálních a dalších charakteristik.

Práce popisuje možnosti a některé aspekty nasazení technologie OLAP v prostředí konkrétní maloobchodní sítě. Jsou zachyceny motivace pro nasazení této technologie, principy na nichž je technologie vybudovaná a obvyklá struktura implementovaných řešení. Práce se věnuje také popisu prostředí maloobchodní sítě, kde se nasazení technologie OLAP plánovalo, dále krokům, které bylo nutné v současné infrastruktuře před implementací provést, a volbou varianty řešení budovaného systému. Jsou zmíněny i některé potenciální problémy implementace BI.

V úvodní části kapitoly o implementaci jsou popsány důvody, které vedly ke změně od původních varianty řešení BI (technologie OLAP, implementace v rámci interní infrastruktury a vlastními kapacitami) k variantě realizované a v současnosti provozované (reportingový systém, řešení dodané externím dodavatelem). Jsou zachyceny i některé podrobnosti a srovnání řešení původně plánovaného a řešení realizovaného.

Kapitoly práce ve většině popisují kroky projektu implementace BI před jejich skutečnou realizací. Kapitola 8.3 v závěru naopak popisuje variantu řešení realizovanou a již uvedenou do produkčního provozu. Realizovaná varianta řešení BI se v produkčním provozu plně osvědčila. Rozsah, kvalita a flexibilita analytických výstupů je na straně uživatelů hodnocena jednoznačně kladně. Z pohledu realizace lze projekt implementace BI v řetězci APC považovat za úspěšný.

10. Závěr

Vypracování práce probíhalo postupně v několika fázích, které přibližně odpovídaly etapám v rámci implementace BI v APC. Kroky implementace reálného projektu a kroky při vypracování práce jakoby se doplňovaly. Velkým přínosem byla možnost přímo v návrhu řešení uplatnit a využít vědomosti získané při studiu literatury a tvorbě práce, a zároveň i možnost vyhledat v literatuře postupy řešení některých nejasných situací nebo potenciálních problémů z praxe.

Po dokončení práce, kterému předcházelo uvedení projektu, který je v práci pojednán, do produkčního provozu, je patrné, že implementaci BI je třeba pečlivě plánovat včetně všech přípravných kroků. Chybné provedení nebo neprovedení všech nutných kroků může vést k nezdaru nebo pozastavení realizace aplikace BI na poměrně dlouhou dobu.

11. Seznam použité literatury

Tištěné zdroje

- [1] GÁLA, Libor, Jan POUR a Zuzana ŠEDIVÁ. ČESKÁ SPOLEČNOST PRO SYSTÉMOVOU INTEGRACI. Podniková informatika. 2., přeprac. a aktualiz. vyd. Praha: Grada, 2009, 496 s. : il. ; 25 cm. ISBN 978-80-247-2615-1.
- [2] LABERGE, Robert. Datové sklady: agilní metody a business intelligence. Brno: Computer Press, 2012. ISBN 9788025137291.
- [3] LACKO, Ľuboslav. Databáze: datové sklady, OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle. Brno: Computer Press, 2003. ISBN 80-7226-969-0.
- [4] LACKO, Ľuboslav. Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby. Brno: Computer Press, 2009. ISBN 978-80-251-2887-9
- [5] MARAKAS, George M. *Modern data warehousing, mining, and visualization: core concepts*. Upper Saddle River, NJ: Prentice Hall, 2003. ISBN 0-13-101459-5.
- [6] 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.
- [7] PALETA, Petr. *Co programátory ve škole neučí, aneb, Softwarové inženýrství v reálné praxi*. Brno: Computer Press, 2003. ISBN 80-251-0073-1.
- [8] POŽÁR, Josef. Manažerská informatika. Plzeň: Vydavatelství a nakladatelství Aleš Čeněk, 2010. ISBN 9788073802769.
- [9] RUD, Olivia Parr. Data Mining: praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníků (CRM). Praha: Computer Press, 2001. Databáze. ISBN 80-7226-577-6.
- [10] TVRDÍKOVÁ, Milena. Aplikace moderních informačních technologií v řízení firmy: nástroje ke zvyšování kvality informačních systémů. Praha: Grada, 2008. Management v informační společnosti. ISBN 9788024727288.

Internetové zdroje

[11] PHARMDATA. Význam PDK. Pharmdata.cz [online]. ©1997-2021 [cit. 2021-10-01].

Dostupné z: https://www.pharmdata.cz/ciselnik_pdk.htm

[12] STÁTNÍ ÚŘAD PRO KONTROLU LÉČIV. Číselník KLK. Sukl.cz [online]. 2021 [cit. 2021-10-01].

Dostupné z: <https://www.sukl.cz/ciselnik-klk?highlightWords=klk>

[13] STÁTNÍ ÚŘAD PRO KONTROLU LÉČIV. Seznam cen a úhrad LP/PZLÚ. Sukl.cz [online]. 2021 [cit. 2021-10-01].

Dostupné z: <https://www.sukl.cz/sukl/seznam-leciv-a-pzlu-hrazenych-ze-zdrav-pojisteni>

[14] STÁTNÍ ÚŘAD PRO KONTROLU LÉČIV. Seznam cen a úhrad ZP. Sukl.cz [online]. 2021 [cit. 2021-10-01].

Dostupné z: <https://www.sukl.cz/sukl/seznam-zdravotnickych-prostredku-hrazenych-na-poukaz>

[15] STÁTNÍ ÚŘAD PRO KONTROLU LÉČIV. Číselník ATC. Sukl.cz [online]. 2021 [cit. 2021-10-01].

Dostupné z: <https://www.sukl.cz/sukl/pomocne-ciselniky-k-seznamu-080401>

[16] REPORTING.CZ. reporting.cz [online]. 2021 [cit. 2021-10-01].

Dostupné z: <https://www.reporting.cz/>

12. Seznam obrázků

Obr. 1 Diagram toku.....	2
Obr. 2 Hlavní komponenty BI a jejich vazby.....	6
Obr. 3 Příklad OLAP kostky.....	13
Obr. 4 Příklad OLAP kostky, analýza časového období podle produktu a regionu.....	13
Obr. 5 Příklad OLAP kostky, analýza produktu podle času a regionu.....	13
Obr. 6 Schéma dat pro analýzu „Hvězda“.....	14
Obr. 7 Schéma dat pro analýzu „Sněhová vločka“.....	15
Obr. 8 Situace před zavedením JSO.....	22
Obr. 9 Situace po zavedení JSO.....	23
Obr. 10 : Stav před zavedením COP.....	24
Obr. 11 : Stav po zavedení COP.....	26
Obr. 12: Architektura řešení BI v řetězci APC.....	29
Obr. 13: Datová reprezentace dimenze „Produkt“.....	36
Obr. 14 : Nejčastější problémy softwarových projektů.....	41
Obr. 15: REPORTING.CZ rozcestník definice reportu.....	50
Obr. 16: REPORTING.CZ nastavení dle dimenze „Období“.....	51
Obr. 17 : REPORTING.CZ nastavení reportů dle dimenze „Nomenklatura“ a „ATC“.....	51
Obr. 18 : REPORTING.CZ výběr ukazatelů reportu.....	52
Obr. 19 : REPORTING.CZ příklad reportu v tabulkové formě.....	52

13. Seznam tabulek

Tabulka č.1 Základní charakteristiky produkční databáze a datového skladu.....	8
Tabulka č.2 Základní rozdíly databází OLTP a datového skladu.....	9
Tabulka č.3 Návrh dimenze „Čas“.....	35
Tabulka č.4 Návrh dimenze „Produkt“.....	36
Tabulka č.5 Návrh dimenze „Prodejní kanál“.....	37
Tabulka č.6 Návrh dimenze „Expedient“.....	37
Tabulka č.7 Návrh tabulky faktů pro analýzu prodejních transakcí.....	38
Tabulka č.8 Porovnání varianty řešení BI původně plánované a realizované.....	48

14. Zadání práce (kopie)

Zadání bakalářské práce

Autor: Jaromír Šplíchal

Studium: I14549

Studijní program: B1802 Aplikovaná informatika

Studijní obor: Aplikovaná informatika

Název bakalářské práce: **Využití technologie OLAP v prostředí maloobchodní sítě**
Název bakalářské práce Use of OLAP technology in retail business
AJ:

Cíl, metody, literatura, předpoklady:

Cíl: Navrhnout a implementovat infrastrukturu pro vyhodnocování dat v maloobchodní síti lékáren, prodejen zdravotnického materiálu a distribučních pracovišť a to z pohledu ekonomických, sortimentních, geografických, personálních a dalších charakteristik. Osnova: 1. Úvod 2. Motivace k využití OLAP v maloobchodní síti lékáren, prodejen zdravotnického materiálu a distribučních pracovišť 3. Teoretická východiska 4. Návrh řešení, návrh báze faktů a dimenzí 5. Implementace a nasazení navrženého řešení do produkce 6. Možné problémy při implementaci a nasazení do produkce a návrh opatření k jejich eliminaci 7. Závěr

-

Garantující pracoviště: Katedra informatiky a kvantitativních metod,
Fakulta informatiky a managementu

Vedoucí práce: doc. Mgr. Tomáš Kozel, Ph.D.

Datum zadání závěrečné práce: 14.1.2015