

Životný cyklus dátových skladov*

Tomáš Tisovský

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií
xtisovskyt@stuba.sk

30. september 2015

Abstrakt

Témou tohto článku je proces tvorby dátových skladov v rámci oblasti Business Intelligence. Vzhľadom na pomerne vysokú komplexnosť DW/BI systémov je využitie agilných praktík v tejto oblasti nutnosťou pre efektívnu tvorbu daných projektov. Pri znalosti metód a postupov pri vývoji dátových skladov je práca na projekte organizovaná a môže dôjsť k výraznému ušetreniu času, či finančných prostriedkov. Článok má za cieľ predstaviť najlepšie techniky, praktiky a prístupy pre tento obor. Článok sa venuje rôznym etapám životného cyklu dátového skladu ako plánovanie a riadenie projektu, definícia a zber požiadaviek, technologická fáza, dátová fáza, aplikačná fáza a údržba. Tieto ciele článok spĺňa opísaním kompletného životného cyklu dátových skladov.

1 Úvod

Tento článok sa venuje životnému cyklu dátových skladov. Technológia dátových skladov predstavuje v súčasnosti jeden z najvýznamnejších trendov v rozvoji podnikových informačných systémov. Dátový sklad (Data Warehouse) možno definovať mnohými spôsobmi. Za základ však budeme považovať definície jedného zo zakladateľov DWH, Williama Inmona, „Dátový sklad je integrovaný, subjektovo orientovaný, stály a časovo rozlíšený súhrn dát, usporiadaný pre podporu potrieb manažmentu“ [4]

Životný cyklus prebieha v niekoľkých etapách. Poznanie etáp projektu je dôležité pre všetkých účastníkov projektu, teda manažérov, analytikov, návrhárov, či vývojárov na vykonanie správnych úloh v správny čas. Pri vytvorení takéhoto softvéru sa kladie hlavný dôraz na požiadavky užívateľov, iteratívnosť a dimenzionalitu v poňatí štruktúrovaných dát. Za štandard v tejto oblasti sa považuje životný cyklus od Ralpha Kimballa. Tento model využíva agilný prístup pre jeho vyššiu efektivitu a úspešnosť. Tento model je zobrazený v Kimball Lifecycle diagram (obr. nižšie). Tento diagram poskytuje celkový plán znázorňujúci postupnosť úloh na vysokej úrovni potrebných pre úspešné DW/BI projekty.

Model podľa Kimballa začína plánovaním. V tejto fáze je potrebné určiť rozsah projektu a potrebné zdroje. V tejto fáze sa začínajú aj riadiace povinnosti,

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2021/22, vedenie: Ing. Vladimír Mlynarovič, PhD.

ktoré pretrvávajú počas celého zvyšku projektu. Na fázu plánovania nadväzuje fáza definovania užívateľských požiadaviek. Táto fáza je pre projekt kľúčová pretože ovplyvňuje všetky nasledovné fázy projektu. Medzi fázou plánovania a fázou zberu požiadaviek je potreba úzkej vzájomnej spolupráce. Horná časť diagramu sa venuje technologickej stránke projektu kde prebieha technologická architektúra. Po dokončení architektúry sa vyberú vhodné nástroje na tvorbu softvéru. Stredná časť diagramu sa zaoberá dátovou stránkou projektu. V tejto fáze sa vývojári venujú dátam, operáciám s nimi. Vzniká tu multidimenzionálny model, ktorý je podstatou DW/BI projektov. Taktiež sa tu tvorí fyzický model a prebehne ETL proces(extract, transform, load). Spodná časť diagramu sa sústreďuje na výstupy pre užívateľov vo forme multidimenzionalnej aplikácie. V sekcii s názvom rast sú vyjadrené praktiky inkrementálnosti. Táto sekcia hovorí o tom, že pri každom prírastku dát by sa vývojári mali vracieť k plánu a držať sa požiadaviek užívateľov. Nasledujúca časť článku je venovaná jednotlivým etapám životného cyklu. [2]

2 Plánovanie a riadenie projektu

Dlhodobý cieľ projektu dátového skladu počíta nielen s jeho vybudovaním, ale definuje aj stratégiu správy dátového skladu, pričom počíta s dokumentáciou dátového skladu a školením jeho užívateľov. V tejto fáze sú definované aj základy architektúry podnikového dátového skladu. Behom fázy definície sa definuje rozsah a cieľ prírastkového vývoja. Vytvorí sa počiatočný prírastok, konceptuálny model, zdokumentujú sa zdroje dát a presne sa vymedzí rozsah kvality týchto dát. Je navrhnutá ako aj architektúra dátového skladu, tak aj architektúra technických prostriedkov. V tejto fáze máme najlepšiu príležitosť zamerať sa na pochopenie štruktúry operačných a externých zdrojov dát. Stanovia sa krátkodobé a dlhodobé obchodné ciele, pre podporu ktorých je dátový sklad budovaný. [1]

Proces riadenia projektov sa týka koordinácie ľudských, finančných a materiálnych zdrojov, je zameraný na dosiahnutie dopredu stanovených cieľov v danom rozsahu, čase, nákladoch, kvalite a spokojnosti účastníkov projektu. [5]

Tak ako pri každom projekte aj pri DW/BI projekte je kľúčový vedúci projektu. Pre projekt je tiež dôležitá úloha biznis analytika, ktorý by mal mať dobré povedomie o spolupráci IT a biznisu. Úlohou dátového analytika je analyzovať kvalitu dát, ich kvalitu a granularitu. Úlohou externého konzultanta je obvyčajne transfer poznatkov podniku, vyškolenie a riadenie ľudí, vedenie projektov. Konzultačné služby v oblasti dátových skladov sa spravidla týkajú výberu hardvéru a softvéru, návrhu architektúry a optimálnej zostavy softvérových technológií a zaistenie využitia najnovších informačných technológií v rôznych oblastiach podnikania. Konzultanti poskytujú služby tiež v oblasti systémovej integrácie, konvergenzie služieb a technológií, služby v oblasti on-line bezpečnosti, vytváranie webových stránok, podnikové informačné systémy. IT konzultant všeobecne pomáha firmám pochopiť, akým spôsobom môžu využiť technológie pre svoj prospech. [1]

Pre projekt je kľúčové vymedzenie nákladov:

- Hardvér - náklady na hardvér nie sú zanedbateľné, ale z hľadiska filozofie dátového skladu sa jedná o technické prostriedky, ktoré sú nahraditeľné.

Jedná sa o technické prostriedky, nie o dáta ktoré sú v nich uložené. Pre rýchly prístup k obrovskému množstvu dát je potrebné mať výkonné servery, alebo dátové centrá. [4]

- Softvér - Nástroje pre vytváranie dátových skladov a analýzu dát sú veľmi drahou záležitosťou. Čím ďalej viac sa presadzuje trend integrácie analytických služieb priamo do inštalácií databázových serverov. Poplatok za analytické služby je buď zahrnutý priamo v cene databázového serveru, alebo sa licenčné poplatky platia zvlášť. [3]

Taktiež je dôležité aj určenie benefitov, ktoré daný systém prinesie, teda určiť ktoré konkrétne rozhodovacie procesy v spoločnosti budú podporené.

3 Iná časť

Základným problémom je teda... Najprv sa pozrieme na nejaké vysvetlenie (časť 3.1), a potom na ešte nejaké (časť 3.1).¹

Môže sa zdať, že problém vlastne nejestvuje [?], ale bolo dokázané, že to tak nie je [?,?]. Napriek tomu, aj dnes na webe narazíme na všelijaké pochybné názory [?]. Dôležité veci možno *zdôrazniť kurzívou*.

3.1 Nejaké vysvetlenie

Niekedy treba uviesť zoznam:

- jedna vec
- druhá vec
 - x
 - y

Ten istý zoznam, len číslovaný:

1. jedna vec
2. druhá vec
 - (a) x
 - (b) y

3.2 Ešte nejaké vysvetlenie

Veľmi dôležitá poznámka. Niekedy je potrebné nadpisom označiť odsek. Text pokračuje hneď za nadpisom.

¹Niekedy môžete potrebovať aj poznámku pod čiarou.

4 Dôležitá časť

5 Ešte dôležitejšia časť

6 Záver

Literatúra

- [1] D. ARNOŠT. *Business Intelligence príručka manažera*. Praha : TATE International, s.r.o., 2007.
- [2] R. KIMBALL and M. ROSS. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Third Edition*. Indianapolis: John Wiley, 2013.
- [3] Ľ. LACKO. *Datové sklady analýza OLAP a dolovanie dat s príklady v SQL Serveru a Oracle*. Brno : Computer Press, 2003.
- [4] O. NOVOTNÝ, J. POUR, and D. SLÁNSKY. *Business Intelligence Jak využít bohatství ve vašich datech*. Praha : Grada Publishing, a.s., 2005.
- [5] M. TVRDÍKOVÁ. *Aplikace moderních informačních technologií v řízení firmy. Nástroje ke zvyšování kvality informačních systémů*. Praha : Grada Publishing, a.s., 2008.