

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“ [2]

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 projekov. 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. [1]

2 Nejaká časť

Z obr. 1 je všetko jasné.

Aj text môže byť prezentovaný ako obrázok. Stane sa z neho označný plávajúci objekt. Po vytvorení diagramu zrušte znak % pred príkazom `\includegraphics` označte tento riadok ako komentár (tiež pomocou znaku %).

Obr. 1: Rozhodujúci argument.

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:

¹Niekedy môžete potrebovať aj poznámku pod čiarou.

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.

4 Dôležitá časť

5 Ešte dôležitejšia časť

6 Záver

Literatúra

- [1] R. KIMBALL and M. ROSS. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Third Edition*. Indianapolis: John Wiley, 2013.
- [2] O. NOVOTNÝ, J. POUR, and D. SLÁNSKY. *Business Intelligence Jak využit bohatství ve vašich datech*. Praha : Grada Publishing, a.s., 2005.