

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačního inženýrství



Teze bakalářská práce

Integrace dat v Business Intelligence

Roman Reiter

© 2019 ČZU v Praze

Souhrn:

V práci je popsán a definován pojem Business Intelligence, vývoj BI na světovém a českém trhu, dále popisuje architekturu BI, definuje zdrojové systémy, datový sklad s jeho vrstvami, ETL procesem. Práce se zabývá představením metodiky Data Vault, pro modelování datových skladů a jsou identifikovány výhody a nevýhody plynoucí pro organizace z jeho implementace. V praktické části byla implementována metodika Data Vault na modelu, volně dostupné datové sady, popsán ETL nástroj Oracle Data Integrator, pomoci, něhož byl následně představen způsob integrace dat. Závěrečná kapitola literární rešerše se zabývá metodikou BPMN pro modelování business byznys procesů a popisem grafických prvků využívaných v modelech.

Klíčová slova:

Business Intelligence, Data Vault, znalostní modul, ETL (extract, transform, load), Data Warehouse (DWH), Business Process Model and Notation (BPMN), Oracle Data Integrator (ODI)

Cíl a metodika práce:

Téma Bakalářské práce je zaměřeno na integraci dat v Business Intelligence. Hlavním cílem práce je návrh procesu integrace dat z datového úložiště do konsolidovaného datového skladu v podnikové praxi. Dílčí cíle bakalářské práce jsou:

- Vytvořit kritickou literární rešerši k problematice Business Intelligence.
- Identifikovat nové trendy v oblasti Business Intelligence.
- Provést deskripci nástroje pro integraci dat.
- Syntetizovat výsledky práce, formulovat závěry a doporučení pro podnikovou praxi.

Metodika řešení dané problematiky se skládá ze studia a analýzy odborných publikací, informačních zdrojů a dokumentace. Z těchto zdrojů jsou zapracované klíčové oblasti Business Intelligence, včetně současného stavu trhu s BI platformami.

Pro účel vlastní práce byl vybrán nástroj: Oracle Data Integrator. Výběr zmíněného nástroje byl proveden na základě studia odborných publikací a na postavení zástupců BI řešení na světovém trhu v roce 2019 dle magického kvadrantu od společnosti Gartner.

Řešení vlastní práce, ve které je realizována implementace BI řešení je dosaženo znalostí prostřednictvím vybraného ETL nástroje. Ve své práci jsem využil volně dostupné databázové řešení od společnosti Oracle, a to konkrétně Oracle HR Schema.

Hlavním trendem, kterým se autor této práce věnuje ve své odborné praxi je aplikování principů modelování pomocí metodiky Data Vault. Pro praktickou ukázkou jsem aplikoval principy metodiky na výše zmíněném modelu HR Schema od společnosti Oracle.

Dosažení hlavního cíle mé bakalářské práce je realizováno nastudováním standardu pro modelování business procesů, konkrétně pomocí metodiky Business Process Model and Notation a návrhem procesu integrace dat prostřednictvím notace BPMN.

Teoretická část:

V první kapitole se autor věnuje vymezení pojmu Business Intelligence, vývoji BI, představení BI systému a jeho řešení, procesu transformace surových dat. Kapitola se dále zabývala jednotlivým komponentám BI řešení, kde jsem představil jednotlivé datové zdroje, definoval datový sklad s jeho vrstvami, vysvětlil principy ETL. Poslední část kapitoly jsem popsal, pro splnění dílčího cíle, a to konkrétně identifikovat trendy v oblasti BI, mezi které patří metodika Agilního vývoje, zabývající se interaktivním způsobem řízení projektů, vysvětlil jsem význam dodávek produktů v malých iteracích zákazníkovi, důležitost spolupráce a zpětné vazby od zákazníka, popsal manifest definující principy agilního vývoje datových skladů, role agilního týmu. V poslední řadě popisují trend, konkrétně metodiku Data Vault pro modelování datových skladů. Na rozdíl od agilního vývoje zaměřením na řízení projektů, se Data Vault zaměřuje přímo na budování datového skladu. V kapitole popisují principy, pravidla pro modelování datového skladu, jeho využití z pohledu integrace, historizace a výhody spojené s agilním přístupem k řízení procesů. V poslední kapitole teoretické části, jsem věnoval vysvětlení metodiky Business Process Model, která se stala standardem pro modelování procesů, popsal grafické prvky využívané v metodice.

Praktická část:

Výsledkem práce je jasná implementace Oracle HR Schema testovací databáze do nového modelu podle vzoru Data Vault. Data Vault ukazuje ze dekompozice na malé části zrychluje vývoj. Díky menším částem se dá vývoj nejen urychlit ale také lépe adaptovat na změny v zadání. Menší části znamenají i menší model, a tak je zde jasné zrychlení i

s učením nového vzoru. Defaultně nám model umožňuje držet důležitá business data a zachovávat historii k veškerým změnám jejich parametrů. Toto umožňuje také lépe chápat business procesy.

Díky možnosti vyvíjet systém v malých částech, iteracích. Je zde jasný potenciál pro agilní vývoj tak jak je vidět v samotné práci. To ovšem znamená že každá iterace, která bude vytvářet skupinu menších částí, rapidně zvýší jejich počet. Toto může být nevýhodou dané implementace. Z teoretického uvažování je jasné, že se nebude hodit pro všechny typy projektů, např. pro ty, které mají malý počet entit, zdrojových systémů a nemají potřebu se v dlouhém časovém horizontu rozvíjet.

Důležité v této části tedy je si uvědomit na jaký projekt je Data Vault nasazován a zvážit i jiné možnosti. Z teoretického uvažování je jasné, že se nebude hodit pro všechny typy projektů, např. pro ty, které mají malý počet entit, zdrojových systémů a nemají potřebu se v dlouhém časovém horizontu rozvíjet.

Závěr:

Na začátku praktické části se věnuji dílčím cílům praktické části, které byly splněny prostřednictvím zhotovení datového modelu dle metodiky Data Vault, na základě teoretických poznatků a zkušeností získaných v praxi a implementace mapování prostřednictvím ETL zvoleného nástroje Oracle Data Integrator. Hlavním cílem práce, bylo navrhnout proces integrace dat. Tento cíl byl realizován prostřednictvím studie notace BPMN, z vlastní zkušenosti a z teoretického uvažování nad nejlepším možným postupem získání všech důležitých částí pro kvalitní BI systém.

Použita literatura:

- [1] NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. *Business intelligence: jak využít bohatství ve vašich datech*. Praha: Grada, 2005. Management v informační společnosti. ISBN 80-247-1094-3.
- [2] POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. *Business intelligence v podnikové praxi*. Praha: Professional Publishing, 2012. ISBN ISBN978-80-7431-065-2.
- [3] GÁLA, Libor, Jan POUR a Prokop TOMAN. *Podniková informatika: počítačové aplikace v podnikové a mezipodnikové praxi, technologie informačních systémů, řízení*

a rozvoj podnikové informatiky. Praha: Grada, 2006. Management v informační společnosti. ISBN 80-247-1278-4.

- [4] KIMBALL, Ralph a Joe CASETRA. *Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data*. John Wiley, 2011. ISBN 978-0-7645-6757-5.
- [5] ŠTOCHOVÁ, Zuzana a Eduard KUNCE. *Agilní metody řízení projektu*. Brno: Computer Press, 2014. ISBN 978-80-251-4196-6.
- [6] INMON, W.H. a Daniel LINSTEDT. *DATA ARCHITECTURE: A PRIMER FOR THE DATA SCIENTIST Big Data, Data Warehouse and Data Vault*. 225 Wyman Street, Waltham, MA: Morgan Kaufmann, 2015. ISBN 978-0-12-802044-9.
- [7] LINSTEDT, Daniel. Data Vault Basics. *Www.danlinstedt.com* [online]. 2015 [cit. 2019-11-15]. Dostupné z: <https://danlinstedt.com/solutions-2/data-vault-basics/>
- [8] HULTGREN, Hans. *Modeling the agile data warehouse with data vault*. New Hamilton, 2012. ISBN 978-0615723082.
- [9] VITOUŠEK, Michal. Data Vault 2.0: Alternativní datový model je budoucnost v realizaci datových skladů, tvrdí odborníci ze Sophia Solutions. *Www.systemonline.cz* [online]. 2018 [cit. 2019-11-14]. Dostupné z: <https://www.systemonline.cz/clanky/data-vault-2.0-alternativni-datovy-model-datoveho-skladu.htm>
- [10] SCHNIDER, Dani. Data Vault Partitioning Strategies: White Paper. *Www.trivadis.com* [online]. [cit. 2019-11-14]. Dostupné z: https://danischnider.files.wordpress.com/2018/02/wp_data_vault_partitioning_strategies_v2.pdf
- [11] GRAZIANO, Kent. Agile Modeling: Not an Option Anymore. *Agile Modeling: Not an Option Anymore* [online]. 2019 [cit. 2019-11-15]. Dostupné z: <https://www.vertabelo.com/blog/data-vault-series-agile-modeling-not-an-option-anymore/>

- [12] SOLOMKA, Petr a Hana KLÍČOVÁ. *Informační systémy v podnikové praxi*. 2. aktualizované a rozšířené vydání. Brno: Computer Press, 2010. ISBN 978-80-251-2878-7.
- [13] MIQUEL, Laura Hofman. Oracle® Fusion Middleware Developer's Guide for Oracle Data Integrator: 11g Release 1 (11.1.1). *Www.docs.oracle.com* [online]. ORACLE, 2010 [cit. 2019-11-15]. Dostupné z: https://docs.oracle.com/cd/E17904_01/integrate.1111/e12643.pdf
- [14] Entity Relationship Diagram. *Www.smartdraw.com* [online]. SmartDraw [cit. 2019-11-24]. Dostupné z: <https://www.smartdraw.com/entity-relationship-diagram/>
- [15] Modelování databází. *Www.root.cz* [online]. 2002 [cit. 2019-11-21]. Dostupné z: <https://www.root.cz/clanky/modelovani-databazi/>
- [16] Business Process Model And Notation. *OMG Object Management Group* [online]. Needham, MA, 2011 [cit. 2019-11-15]. Dostupné z: <https://www.omg.org/spec/BPMN/2.0/>
- [17] Magic Quadrant for Data Management Solutions for Analytics. *Www.gartner.com* [online]. Gartner, 2018 [cit. 2019-11-20]. Dostupné z: <https://www.gartner.com/doc/reprints?id=1-3U1LC65&ct=170222&st=sb>
- [18] Magic Quadrant for Data Integration Tools and the 2019. *Www.informatica.com* [online]. Gartner, 2019 [cit. 2019-11-20]. Dostupné z: <https://www.informatica.com/data-integration-magic-quadrant.html#fbid=L-3epJQ8XIe>
- [20] Business Process Model And Notation. *OMG Object Management Group* [online]. Needham, MA, 2011 [cit. 2019-11-15]. Dostupné z: <https://www.omg.org/spec/BPMN/2.0/>
- [21] HLAVA, Tomáš. Fáze a úrovně provádění testů. *Testovanisoftware.cz* [online]. 2011 [cit. 2019-11-26]. Dostupné z: <http://testovanisoftware.cz/tag/fat/#fat>
- [22] Oracle Database Online Documentation 11g Release 2 (11.2). *Www.oracle.com* [online]. Oracle [cit. 2019-11-29]. Dostupné z: https://docs.oracle.com/cd/E11882_01/server.112/e40540/tablecls.htm#CNCPT010