

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačního inženýrství



Bakalářská práce

Archivace dat v datovém skladu

Jan Šnaur

© 2018 ČZU v Praze

Abstrakt (česky):

Bakalářská práce se zaměřuje na problematiku využití dostupných technologií pro vybudování datového skladu. Teoretická část práce stručně rozebírá význam jednotlivých komponent datového skladu. V praktické části práce je prováděna analýza technologií datových skladů od různých výrobců, které jsou bodově hodnoceny na základě požadavků fiktivní společnosti. V závěru práce jsou shrnuty výsledky analýzy jednotlivých výrobců a je provedeno porovnání. Z výsledků porovnání vyplývá, že konkurence v oblasti datových skladů a technologií s tím souvisejících je velmi vyrovnaná.

Klíčová slova: datové sklady, databáze, OLAP, business intelligence, dolování dat, technologie datových skladů

1 Úvod

Potřeba realizace datového skladu je v dnešní době u společností operující se svými daty ve smyslu provádění analýz a reportingu téměř nutností. Datové sklady se už netýkají jen velkých společností, které disponují velkým množstvím dat, ale i manažeři v malých a středních firmách potřebují svá rozhodnutí opřít o data z provozních systémů a o jejich analýzy. Drtivá většina společností uchovává svá data, ať už v klasických relačních databázích, nebo v úložištích založených na cloudové technologii. Součástí řešení datových skladů však není jen datové úložiště, ale celá řada dalších funkcí, které umožňují z těchto dat vytěžit cenné informace. Tyto informace pak mohou zefektivnit rozhodování v klíčových momentech a představovat tak možnou konkurenční výhodu. Existuje velké množství produktů od různých společností, které tato řešení nabízí, ne všechny však poskytují stejně kvalitní služby a aplikace pro realizaci datového skladu.

Tato bakalářská práce analyzuje a hodnotí možnosti realizace datového skladu podle různých poskytovatelů těchto řešení. Čerpá z dostupných informací jednotlivých výrobců a dále z informací nestranných společností dlouhodobě provádějící praktický výzkum, analýzu a testování datových skladů od distributorů těchto řešení. Tyto informační zdroje budou na příkladu fiktivní společnosti se stanovenými výběrovými kritérii dále analyzovány a bodovány, a na základě těchto výsledků bude vybrán výrobce s nejlepším hodnocením.

2 Cíle práce

Cílem bakalářské práce je analýza a hodnocení možností realizace datových skladů podle různých poskytovatelů těchto řešení. Čerpá z dostupných informací jednotlivých výrobců a dále z informací nestranných společností dlouhodobě provádějící praktický výzkum, analýzu a testování datových skladů od distributorů těchto řešení. Tyto informační zdroje budou na příkladu fiktivní společnosti se stanovenými výběrovými kritérii dále analyzovány a bodovány, a na základě těchto výsledků bude vybrán výrobce s nejlepším hodnocením.

3 Metodika práce

Analýza vhodných řešení bude prováděna na základě výsledků společností dlouhodobě provádějících výzkum a následné hodnocení v oblasti data warehousingu, konkrétně se jedná

o společnosti Gartner a Forrester uznávané autority v tomto oboru. Gartner vydal výsledky své analýzy v článku „Magic Quadrant for Data Management Solutions for Analytics“ (2017) a Forrester v obdobné práci „The Forrester Wave™: Big Data Warehouse, Q2 2017“ (2017). Firma nechce data warehouse stavět na službách založených na cloud technologiích, je pro ni velmi důležitá možnost budoucího rozšíření funkcí v podobě nových upgradů, technologie Hadoop umožňující zpracovávání dat velikostí pentabytů až exabytů. Dále chce integrovaný systém pro ukládání a přístup k datům podporující analytické funkce a SQL funkcionalitu.

4 Diskuze

Výsledné hodnocení jednotlivých společností ukazuje velkou vyrovnanost na trhu datových skladů, viz Tabulka 1. Téměř všechny analyzované společnosti splnily požadavky fiktivní firmy Satax s.r.o, která aktuálně eviduje data o zákaznících a objednávkách v informačním systému, jenž ukládá veškerá data do relační databáze.

Tabulka 1- Výsledné bodové hodnocení

Datový sklad nezaložený na Cloud technologii	Možnosti upgradu	Analytická podpora	Technologie Hadoop	Podpora SQL	Celkové hodnocení
Teradata	5	5	5	5	25
Oracle	5	5	5	5	25
Microsoft	5	5	5	5	25
IBM	5	5	5	5	25
HP	5	5	5	5	25
SAP	5	1	5	5	21
Amazon Web Services	1	1	5	5	17

Satax klade důraz na tradiční řešení datového skladu a nechce využívat cloudové služby, jež také umožňují realizaci datového skladu. Dále vyžaduje podporu technologie Hadoop z důvodu možného rozšíření společnosti a zvětšení datové základny, a dále podporu SQL dotazování. V těchto bodech vyhovělo všech sedm analyzovaných společností. V

neposlední řadě bylo požadováno, aby existovaly možnosti rozšíření datového skladu o další funkce, jako například Machine learning, jenž je součástí prediktivní analýzy pro zlepšování kvality svých aplikací. V tomto bodě nevyhověly společnosti SAP a Pivotal, jenž neposkytují dostatečné možnosti pro upgrade datového skladu o nové funkce. Satax vzhledem k důležitosti analýzy dat klade důraz na širokou podporu analytických funkcí, toto kritérium nesplňuje společnost Pivotal. Vzhledem ke shodnosti výsledků hodnocených kritérií u společností Teradata, Oracle, Microsoft, IBM a HP bude konečná volba jedné ze společností záležet primárně na cenách realizace jejich řešení. To musí být vykomunikováno s obchodním oddělením každého poskytovatele.

5 Závěr

Čím dál větší množství společností má potřebu využívat svá data pro lepší zacílení svých marketingových kampaní a také jako oporu pro rozhodování. Tuto výsadu měli dříve jen manažeři velkých společností. Tyto společnosti svá data spravovaly v datových skladech, jenž byly součástí business intelligence, která umožňovala těžit z dat potřebné informace. Tato skutečnost dnes není výsadou jen velkých společností, dostupná jsou již i technologická řešení pro malé a střední firmy, které chtějí využívat těchto technologií pro zlepšení a zvýšení konkurenceschopnosti.

Primárním cílem této práce bylo analyzovat a vyhodnotit možnosti realizace datového skladu podle různých poskytovatelů těchto řešení. Jako uživatel datového skladu byla zvolena fiktivní společnost jménem Satax, která si na základě svých požadavků stanovila kritéria výběru vhodného řešení datového skladu. Tato kritéria byla aplikována na informace od nestranných společností dlouhodobě provádějící praktický výzkum, analýzu a testování datových skladů od distributorů těchto řešení, a dále na informace poskytované jednotlivými společnostmi v rámci možností a technických specifikací svých produktů. Tato kritéria byla hodnocena bodovou škálou a výsledné hodnocení zobrazeno ve formě tabulky, viz Diskuze. V předcházejících kapitolách byla rozebírána také technologie datových skladů jako základ systémů business intelligence.

6 Základní použitá literatura

INMON, William H. 2002. Building the data warehouse. 3rd ed. New York: J. Wiley, xx, 412 p. ISBN 04-710-8130-2.

KIMBALL, Ralph a Margy ROSS. c2010. The Kimball Group reader: relentlessly practical tools for data warehousing and business intelligence. Indianapolis, IN: Wiley, xxiv, 718 p. ISBN 978-0-470-56310-6.

LACKO, Euboslav. 2011. 1001 tipů a triků pro SQL. Vyd. 1. Brno: Computer Press, 416 s. ISBN 978-80-251-3010-0.

NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. 2005. Business intelligence: jak využít bohatství ve vašich datech [online]. 1. vyd. Praha: Grada, 254 s. Management v informační společnosti. ISBN 80-247-1094-3.

On-line internetové zdroje