

BUSINESS INTELLIGENCE

**VYSOKÁ ŠKOLA EKONOMIE A MANAGEMENTU
2011**

Business Intelligence

Doc. Ing. Jan Žižka, CSc.

Copyright ©Vysoká škola ekonomie a managementu 2011

Vydání první. Všechna práva vyhrazena.

ISBN: 978-80-86730-79-0

Vysoká škola ekonomie a managementu

www.vsem.cz

Žádná část této publikace nesmí být publikována ani šířena žádným způsobem a v žádné podobě bez výslovného svolení vydavatele.

Tato publikace neprošla redakční ani jazykovou úpravou.

Předmluva

Text ve skriptech se zaměřuje na stručný úvod do velmi moderní a intenzívně se rozvíjející oblasti zvané Business Intelligence, jejímž cílem je co nejhlouběji analyzovat rozsáhlá data a o výsledku podat zpravodajství vedoucím pracovníkům, kteří jsou zodpovědní za řízení a plánování. Znalost, kterou Business Intelligence z dat získá, pak napomáhá správnému rozhodování, plánování a řízení.

Skripta jsou primárně zaměřena na dvě oblasti:

1. Ukládání dat do datových skladů, které následně slouží jako zdroj dat analytických aplikací.
2. Analýzy obsahu datových skladů, a to zejména nejmodernějším způsobem prostřednictvím dolování znalosti z dat.

Skripta jsou zaměřena na studentky a studenty bakalářského stupně ekonomických oborů. Záměrem bylo podat zejména verbální vysvětlení základních a nejdůležitějších pojmů a uvést do terminologické oblasti Business Intelligence. Zájemci o širší či hlubší znalosti probíraného a velice rozsáhlého oboru mohou využít odkazy na literaturu, které obsahují většinou zcela nové knižní zdroje.

OBSAH

KAPITOLA 1: Business Intelligence.....	3
KAPITOLA 2: Data, informace, znalost, metaznalost.....	7
KAPITOLA 3: Přehled procesů Data Warehousing	11
3.1 Správa a využití dat	11
3.2. Datové sklady	12
3.3 Architektura datových skladů.....	14
3.3.1 Základní vícevrstvé architektury.....	15
3.3.2 Webová architektura.....	17
3.3.3 Požadavky na realizaci datového skladu	19
3.3.4 Alternativní architektury datových skladů	20
3.3.5 Faktory ovlivňující rozhodnutí o architektuře datového skladu	22
3.3.6 Která architektura je nejlepší?	23
3.4 Integrace dat a ETL proces.....	25
3.4.1 Integrace dat.....	26
3.4.2 ETL proces	27
3.4.3 Vytvoření datového skladu	30
3.4.4 Přístupy k budování datových skladů	33
KAPITOLA 4: Dimenzionální modelování dat v datovém skladu	39
4.1 Tabulky faktů.....	39
4.2 Tabulky dimenzí.....	40
4.3 Propojení tabulek dimenzí s tabulkami faktů.....	40
4.3.1 Schéma Star.....	40
4.3.2 Schéma Snowflake	41
4.4 Analýza dat v datovém skladu	42
4.4.1 Odlišnosti a spolupráce systémů OLAP a OLTP.....	42
4.4.2 Operace poskytované systémy OLAP.....	43
4.4.3 Různé variace OLAP.....	45
4.5 Vytvoření datového skladu	46
4.6 Masívní datové sklady a jejich rozšiřitelnost.....	51
4.7 Datové sklady v reálném čase	52
4.7.1 Proces RDW/ADW.....	53
4.7.2 Problém „více pravd“ v souvislosti s reálným časem.....	54
4.8 Administrace a bezpečnost datových skladů	55
4.8.1 Administrátor datového skladu a požadavky na něj	56
4.8.2 Bezpečnost a privátnost podnikových informací.....	57
4.9 Budoucnost datových skladů	58

4.9.1 Datové zdroje – získávání dat z různých rozptýlených zdrojů	59
4.9.2 Infrastruktura (architektura – hardware a software – zlepšování)	60
KAPITOLA 5: Dolování z dat pro business intelligence	65
5.1 Koncepty a definice dolování z dat	66
5.2 Taxonomie dat pro business intelligence	72
5.3 Jak dolování z dat pracuje?	77
5.4 Aplikace dolování z dat	79
5.5 Proces dolování z dat	82
5.6 Metody dolování z dat	87
5.6.1 Klasifikace	87
5.6.2 Odhad skutečné spolehlivosti klasifikačních modelů	88
5.6.3 Některé klasifikační techniky	90
5.6.4 Shlukovací analýza	94
5.6.5 Asociační techniky	95
5.7 Softwarové nástroje pro dolování z dat	96
Literatura	103
VZOROVÝ TEST	105

KAPITOLA 1: Business Intelligence

Úvod

První kapitola přináší základní informace o tom, co to vlastně anglický termín Business Intelligence znamená a jaký je jeho vztah k záležitostem ekonomiky. Zavádí se definice a nová terminologie s vysvětlením podstatných termínů.

Cíle kapitoly

Seznámit se

- ♣ s pojmem Business Intelligence
- ♣ s definicí oboru Business Intelligence
- ♣ s tím, co vlastně Business Intelligence dělá
- ♣ se odvětvím Business Intelligence ve vztahu k ekonomice

Business Intelligence, ve zkratce **BI**, je původní anglický název pro oblast, která se zabývá pokud možno co nejhlubší **analýzou dat** pocházejících z oblasti podnikání. Termín *Business Intelligence* je – částečně vážně a částečně snad i žertem – inspirován oblastí zdánlivě zcela odlišnou, která však ve skutečnosti dělá totéž, i když nejen s daty podnikatelskými v převážně ekonomickém prostředí: zpravodajskými službami, špionáží a kontrašpionáží. Názvy řady zpravodajských služeb v sobě slovo *intelligence* ve významu *zpravodajství* či *informace* obsahují, například *CIA* (Central Intelligence Agency v USA), *MI5* (Directorate of Military Intelligence, Section 5, UK), *ASIS* (Australian Secret Intelligence Service, Austrálie), *NZSIS* (New Zealand Security Intelligence Service, Nový Zéland), a celá řada dalších. Tyto služby, obdobně jako *BI*, jsou založeny na co nejrozsáhlejším sběru dat. Předpokládá se totiž, že data v sobě obsahují něco užitečného, co se dá použít (případně i zneužít) k různým cílům. Zmíněné cíle představují pro sběratele a vlastníka dat určitý užitek, a to podle konkrétního zaměření obsahu získaných dat nebo jejich vztahu k určitým skutečnostem.

V žádném případě však nelze tvrdit, že by **cílem BI** byla nějaká špionáž. I když se přirozeně něco na způsob špionáže v podnikatelské praxi objevuje (a proto vzniklo mj. příslušné odvětví oboru *bezpečnost dat*), *BI* je svým principem a vývojem zaměřeno především na to, jak se moderní společnosti, instituce, organizace a podniky **eticky a legálně** soustřeďují na shromažďování co největšího množství údajů o svých zákaznících, akcionářích, konkurentech, podnikatelském prostředí, a to z různých zdrojů potenciálně cenných informací, a dále zejména na to, jak z takto nahromaděných dat „vydolat“ libovolnou užitečnou a použitelnou znalost.

Lze tedy říci, že na *BI* je možno se dívat jako na určitý druh legálního a etického zpravodajství, které má za úkol v disponibilních datech – ať už svých vlastních veřejně přístupných (často na základě zákona) nebo naopak utajovaných (například obchodní tajemství), případně i veřejně dostupných od konkurentů – objevit něco přínosného, bez ohledu na to, zda se jedná o data z vlastního nebo konkurenčního podniku; přínosné je v konkurenčním prostředí samozřejmě obojí. V ekonomické oblasti, podobně jako třeba ve vojenské nebo sportovní, je účelem získat převahu nad soupeři-konkurenty. Jak tu převahu získat? Objevením užitečného obsahu skrytého v nashromážděných datech a jeho využitím pro zlepšení vlastních činností tak, aby překonaly konkurenci.

V tomto učebním textu budeme tedy nadále rozumět pod pojmem *Business Intelligence* jeho význam jako **podnikové zpravodajství**. Nemusí se jednat výlučně o podniky, to jest firmy, továrny, služby a podobně – uvedený typ zpravodajství je využitelný i v libovolných organizacích nebo institucích, včetně neziskového sektoru, a nakonec i v organizacích nebojících přímo s konkurencí, avšak snažících se přesto zlepšit svou průběžnou činnost a výsledky. *Business Intelligence* je obor využitelný jak na úrovni „nejnižší“, neboli na úrovni velmi malých firem, které mohou obsahovat i jediného člověka, až po úroveň „nejvyšší“, jako jsou státy, kontinenty, a podobně – i státní rozpočty a plánování musí vycházet z obsahu určitých konkrétních údajů a nepochybně by měly být optimalizovány. Volba měřítka není rozhodující; rozhodující je snaha o optimalizaci, tedy dosažení co nejlepších výsledků v libovolné oblasti, a to nejen ekonomické, přestože tato oblast původně stojí u zrodu *Business Intelligence*.

Business Intelligence jako obor nemá svou přesnou definici, což je v současnosti u mnoha nejrůznějších oborů zcela typické. Nejedná se v principu o exaktní záležitost, jde spíše o odvětví vzniklé z ryze praktických a pragmatických potřeb, které představují nejrůznější problematiku vyžadující nějaké pozitivní řešení, které přináší určité zlepšení vůči minulému a současnému stavu pro budoucnost. *BI* je odvětví, které je především ovlivněno praktickými a pragmatickými potřebami, a s tím, jak se tyto potřeby vlivem změn a vývoje v čase mění, se mění i obsah a zaměření vlastního *BI*. To, že se nejedná o exaktní záležitost, však rozhodně v žádném případě neznamená, že *BI* se zaměřuje jen na něco nepřesného, přibližného, kde na exaktnosti nezáleží. Naopak, čím exaktnější výsledky mohou být dosaženy, tím lépe. Proto *BI* využívá, pokud je to možné, veškeré exaktní disciplíny, které mohou pomoci dosáhnout cíle, především matematiku a její aplikace na reálný svět.

Business Intelligence si přesto, na základě současného stavu (rok 2011), můžeme slovně

definovat například následovně, více-méně značně neformálně:

Definice

Business Intelligence (BI) je zastřešující odborný *termín* pro určitou kombinaci vhodných softwarových i hardwarových nástrojů, databází, analytických prostředků a metod, aplikací a metodologií za účelem získání znalosti skryté v údajích popisujících reálný svět okolo nás, a to v oblastech spojených s podnikáním, ekonomikou a jimi ovlivňovanými institucemi libovolného charakteru.

V literatuře existují i jiné definice, které však sdílejí společný základ: **co nejhlubší analýza co největšího množství dat**. Zde uvedené vyjádření definice BI je vzhledem ke své relativně značné obsahové volnosti značně flexibilní v tom smyslu, že může pro různé osoby znamenat různé věci. Je to kladná i zároveň záporná stránka celé záležitosti. Záporné hledisko představuje mimo jiné skutečnost, že vzhledem k neexaktnosti definice vznikla záplava různých akronym (zkratkových názvů jako např. OLAP, MOLAP, WOLAP, apod.) a s nimi spojených trendových, módních výrazů (buzz-words), které často mohou mít různý výklad v závislosti na určitém kontextu a vyžadují další definice a vysvětlení. Na druhé straně tato flexibilita pozitivně podporuje různá konkrétní odlišná řešení, která vznikají jako odpovědi na často značně obtížné požadavky praxe v dynamickém prostředí reálného světa. BI jako obor nemá standardní jedinou východiskovou teorii, z které by se mohla různá řešení odvíjet, avšak za svou již relativně dlouhou dobu existence (několik desetiletí) jsou určité spolehlivé metodologie a nástroje k dispozici.

Definice

Proces Business Intelligence vychází z *transformace* vstupních dat na informaci, která je dále přeměněna na znalost použitelnou pro rozhodování zakončené vybraným typem akce.

Jinými slovy, hlavním **cílem BI je umožnit interaktivní přístup k datům, jejich netriviální zpracování, a nakonec vytvoření co nejlepších možností pro analytiky a řídicí pracovníky dosáhnout příslušných výsledků z analýz**. Analýzou historických (často tzv. „zděděných“) a současných dat včetně konkrétních situací a výkonů získávají

pracovníci, kteří jsou zodpovědní za rozhodování, mnohem hlubší a velmi cenný náhled na reálný stav řešené záležitosti, což jim umožňuje na základě lepší informovanosti provádět lepší rozhodování. Je vhodné zdůraznit, že veliké množství problémů reálného světa, k nimž patří i ekonomika, naneštěstí často odolává přesnému řešení pomocí exaktních metod (např. matematických modelů). Proto nebývá pro daný problém k dispozici vždy jednoznačná odpověď a rozhodování stojí na rozcestí, kde pokud možno ten správný směr je nutno odhadnout z toho, co je pro daný okamžik známo a k dispozici. Kvalitní odhad pak závisí na stupni kvality analýz, jejichž vstupem jsou právě ta zmíněná data.

Shrnutí kapitoly

- ⤴ Kapitola se zaměřila na zavedení a vysvětlení pojmu Business Intelligence jako na podnikové zpravodajství.
- ⤴ Cílem Business Intelligence je výrazné prohloubení analýzy podnikových dat za účelem zlepšení pozice podniku v silném konkurenčním boji.
- ⤴ Výsledky analýz jsou určeny pro vedoucí pracovníky, aby jim pomohly v obtížném procesu řízení, plánování a rozhodování.

Klíčová slova

- ⤴ Business Intelligence
- ⤴ analýza dat
- ⤴ podnikové zpravodajství