

# **Novinky SQL Serveru 2005 v oblasti Business Intelligence**

---

Seminární práce na předmět Business Intelligence (4IT435)  
Vypracoval Borek Bernard, leden 2006

# 1 Abstrakt

Microsoft SQL Server 2005 je po mnoha letech další významnou verzí tohoto databázového stroje. Přináší velkou řadu zásadních novinek, z nichž některé se týkají oblasti business intelligence. Práce tyto novinky ve stručnosti popisuje.

## 2 Úvod

Kapitola s obecným úvodem je obvykle přeskakována, ale práce by bez ní nebyla kompletní. Nejprve je totiž potřeba stručně naznačit, k čemu firmy vůbec služby business intelligence potřebují.

Je již zcela běžné, že firmy pro evidování a podporu každodenních transakcí používají databáze, ať už se pod tímto pojmem myslí tabulka v excelu, jednoduchá Access aplikace nebo nejčastěji databáze v tradičním slova smyslu (OLTP databáze). Rovněž je typické, že data přicházejí nejen zevnitř, ale i z okolí podniku, takže je potřeba potýkat se s různými formáty dat a s dalšími nepříjemnostmi.

Aby se z údajů či dat mohly stát cenné informace, je potřeba něco víc než jen výkonné databáze umožňující zpracovávat X tisíc transakcí za překvapivě málo času. Proto přední dodavatelé databází neváhali a ke svým produktům přidávaly různé další služby, pro které se vžilo souhrnné označení business intelligence.

Jelikož je tato práce zaměřena na produkt Microsoft SQL Server, bude cílem práce popsat, jak se snaží tento produkt svým uživatelům přinést zajímavé informace o jejich podnikání a v čem jsou tyto metody nové.

## 3 Cíle SQL Serveru 2005 v oblasti BI

SQL Server 2000 přinesl velmi použitelnou platformu pro vytváření a konzumování služeb business intelligence. Hlavním cílem SQL Serveru 2005 je celý proces ještě zjednodušit a všechny nástroje lépe integrovat. Nejviditelnějším projevem je integrace shellu SQL Serveru s Visual Studiem, ale na pomoc přicházejí i přepracované nástroje pro ETL (z DTS se stávají Integration Services) a další komponenty SQL Serveru pro analýzu, reportování a data mining.

Obrázek vystihuje současné dilema mezi hodnotou BI nástroje a snadností jeho použití. Ambiciózním cílem SQL Serveru 2005, jak naznačuje šipka, je dosáhnout co nejlepší hodnoty získaných informací při snaze o zachování rozumné použitelnosti.

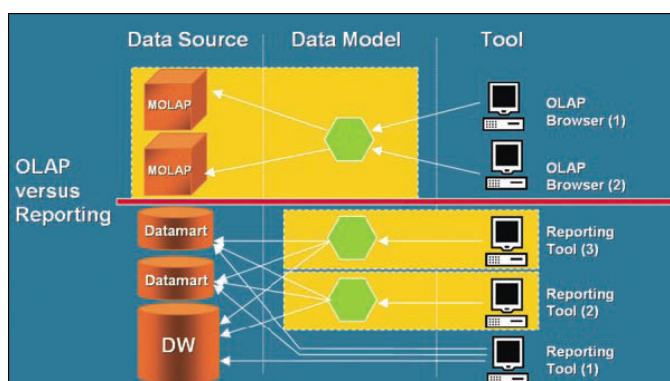


Obrázek 1 - Cíle SQL Serveru v oblasti BI

## 4 Základní filosofie BI v SQL Serveru 2005

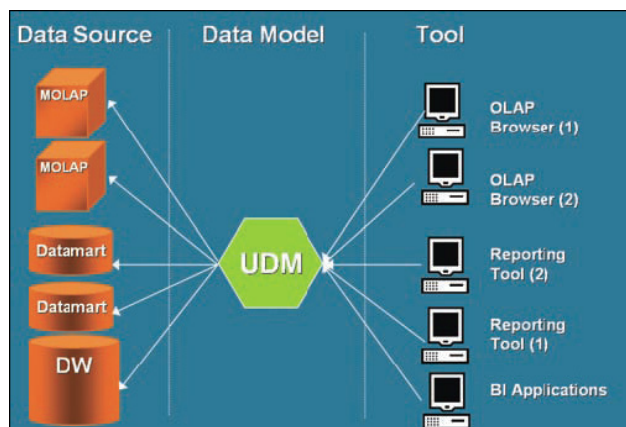
V SQL Serveru verze 2000 jsou poměrně ostře odděleny vrstvy databázových, analytických a reportovacích služeb, jak přehledně ukazuje obrázek 2. Hlavní nevýhodou této architektury je redundance údajů jednotlivých modelů, protože stejná data jsou uložena jak v relačních tak v multidimenzionálních databázích.

Nová verze 2005 celou základní architekturu mění za pomoci něčeho, co je nazváno Unified Dimensional Model (UDM) – viz obrázek 3. Všechny vrstvy business intelligence jsou sjednoceny do jednoho pohledu, který převzal to nejlepší z reportování a OLAP analýz.



Obrázek 2 - Model BI v SQL Serveru 2000

Typický postup v případě SQL Serveru 2000 byl ten, že data byla nejdříve získána z relačních tabulek nebo z datového skladu, po čemž došlo k vytvoření zcela nové datové struktury sloužící pro rychlé multidimenzionální dotazování (tato struktura má zkratku MOLAP – multidimensional on-line analytical processing).



Obrázek 3 - Unified Dimensional Model

Kromě redundance údajů navíc ještě rostou nároky na úložné místo – zvláště po přidání několika dimenzí mohou tyto nároky růst velmi rychle.

SQL Server 2005 tento problém řeší speciální keší, kde jsou uchovávány agregace nejčastěji pokládaných dotazů, takže

výkon trpí relativně málo a současně jsou překonány ty největší nedostatky.

## 5 Základní stavební kameny

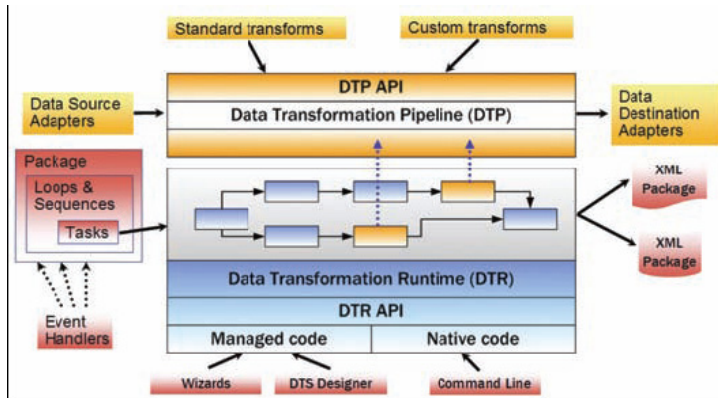
Služby business intelligence lze rozdělit do čtyř hlavních oblastí:

1. Integrované služby – slouží k integraci dat z různorodých datových zdrojů, zprostředkovávají holistický pohled na podnikání
2. Analytické služby – nabízí analytický pohled na nashromážděná data (OLAP a spol.)
3. Reportovací služby – vytváření reportů
4. Data mining

O každé oblasti bude stručně pojednáno.

## 6 Integration services

Integrační služby jsou „velkou novinkou“ SQL Serveru 2005 (viz [3]). Nahrazují Data Transformation Services známé z SQL Serveru 2000 a přinášejí několik zajímavých vylepšení.



Obrázek 4 - Architektura SSIS

Obrázek 4 zobrazuje základní architekturu SQL Server Integration Services (SSIS). Základním prvkem je DTP (Data Transformation Pipeline), která propojuje zdrojové a cílové datové adaptéry. V nižších vrstvách se odehrávají procesy pro písmenka E a T, tedy extrakce a transformace. Základem této vrstvy jsou úlohy, tzv. tasks, což jsou samostatně vykonatelné jednotky zpracování.

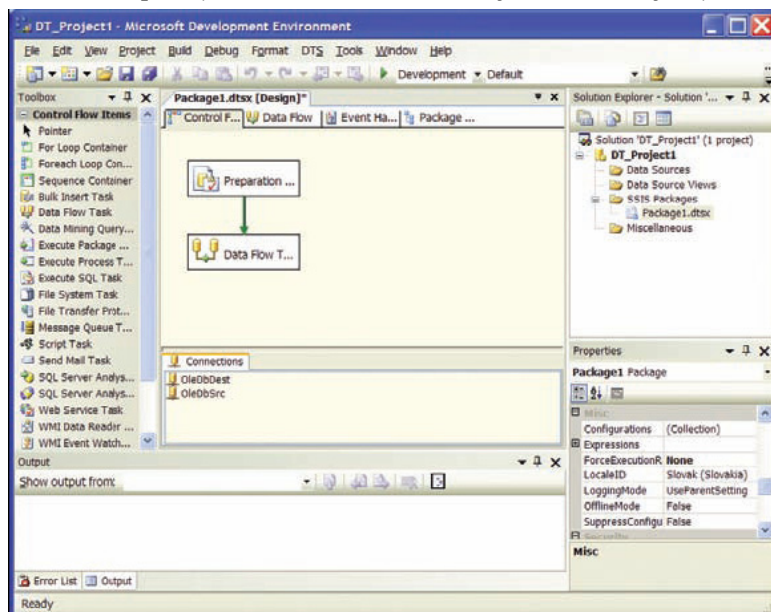
Úlohy lze řetězit (stejně jako v SQL Serveru 2000 zde existují workflow prvky jako on success nebo on failure).

Výrazného vylepšení se také dočkala výkonnost. V porovnání s verzí 2000 jsou běžné úlohy několikrát rychlejší (viz [1], str. 16) a s výhodou lze využít také x64 platformy, kde je možno počítat zhruba s dalším zhruba třináásobným růstem výkonu (viz tamtéž).

Jak už bylo zmíněno, celý SQL Server 2005 je nyní integrován do shellu Visual Studio, čímž pokračuje snaha Microsoftu dostat pod jednu střechu všechny zásadní vývojářské nástroje (ve Visual Studiu lze dnes kromě samotných vývojářských prostředí typu C++ nebo Visual Web Developer provozovat také právě SQL Server a navíc BizTalk Server). Pátý obrázek ukazuje, jak tato integrace vypadá.

Mezi základní etapy vývoje SSIS projektů patří 4 kroky, jak už napovídá solution explorer svými záložkami:

- definování datových zdrojů
- definování pohledů na datové zdroje
- SSIS balíčky
- různé další činnosti



Obrázek 5 - BI v prostředí Visual Studio

Zajímavým oknem je záložka Data Flow, která po spuštění projektu umožňuje vizuálně sledovat průběh vykonávání ETL. Celý výsledek je nakonec uložen do přehledného logu, aby bylo možno zjistit detaily vykonávání jednotlivých kroků.

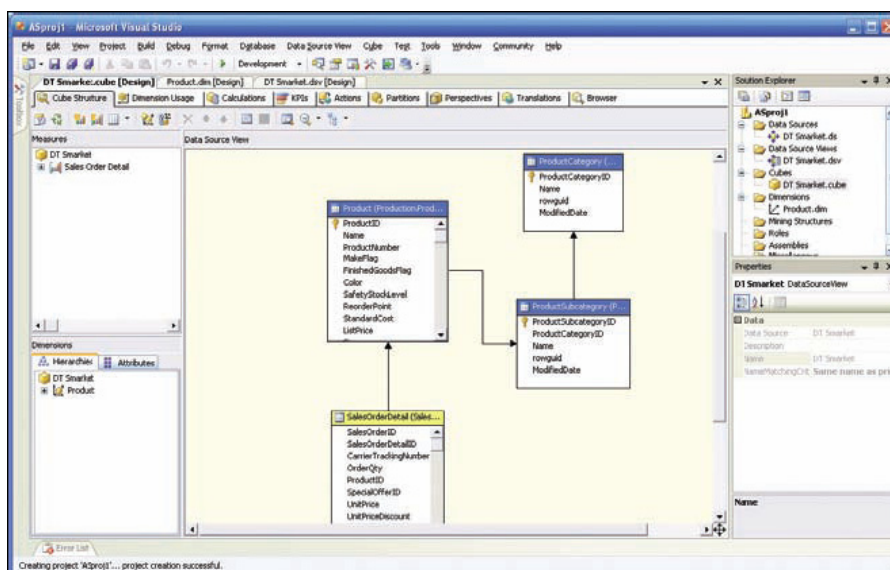
## 7 Analytické služby – OLAP analýza údajů

Také analytické služby jsou nyní plně integrovány do prostředí Visual Studia. Prostředí je vidět na obrázku 6. Velké množství záložek poukazuje na mnoho různých pohledů na vytváření datovou kostku.

Hlavní novinkou analytických služeb je to, že poskytují sjednocený a integrovaný model pro všechny ostatní služby business intelligence, od OLAP analýz přes KPI scorecards až po data mining.

Mezi zásadní vlastnosti patří:

- Unified Dimension Model (UDM), zmíněný již dříve. Jedná se o sémantický model, který definuje business entity, business logiku, kalkulace a metriky. UDM je centrálním místem, na které se obracejí všechny další služby BI.
- Proaktivní kešování umožňuje spojit dvě zatím protichůdné vlastnosti: real-time aktualizace a vysoký výkon OLAP dotazování. Místo toho, aby byly údaje kopírovány do nových multidimenzionálních struktur (jak to bylo u verze 2000), data jsou uložena centrálně a SQL Server automaticky udržuje komprimovanou a vysoce optimalizovanou keš nejčastěji požadovaných údajů a agregací. Při změně dat jsou tyto změny automaticky reflektovány v keši.
- Podpora webových služeb implementací standardu XML/A (XML for Analysis)
- Enterprise vlastnosti jako škálovatelnost, jednoduchá správa (pomocí nového SQL Server Management Studia) a produktivita (integrace s Visual Studiem, průvodci na všechno, integrace se systémy pro správu verzí, debugování atd.)



Obrázek 6 - Analytické služby ve Visual Studiu

## 8 Data mining

Data mining je oblastí business intelligence, která pomáhá odhalit určité vzory v množině údajů. Výraznou novinkou data miningu v SQL Serveru 2005, jak již bylo naznačeno, je integrace se zbytkem business intelligence platformy. Dolování dat proto vychází z dat získaných a připravených integračními službami (SSIS), obohacenými o multidimenzionální pohled analytickými službami (SSAS) a nakonec zveřejněnými třeba pomocí reportovacích služeb.

Po technologické stránce přináší SQL Server 2005 nové algoritmy pro data mining. Zde je přehled:

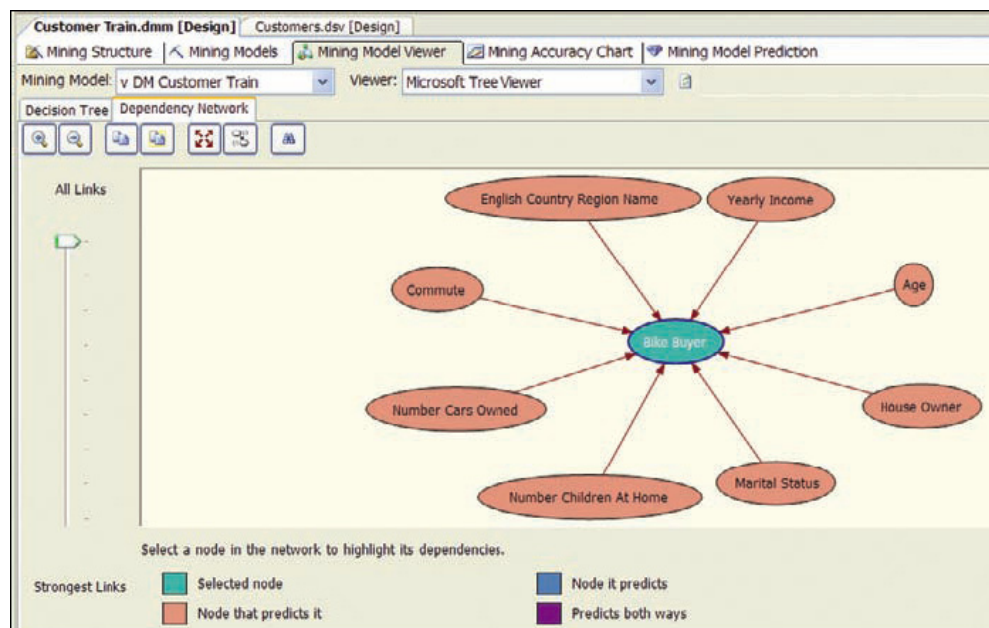


- Asociační pravidla – hledají se souvislosti přiřazování. Používá se např. při hledání odpovědi na otázku, jaké produkty kupují zákazníci současně. Matematicky se jedná o hledání korelací.
- Vícerozměrné shlukové diagramy – hledá shluky dat v dimenzionálním prostoru, např. zákaznické segmenty podle jednotlivých kritérií.
- Sekvenční shlukování je specifickým případem shlukování – zde záleží na pořadí případů.
- Nevyvážené rozhodovací stromy – odhaluje závislosti a vyhledává specifické vlastnosti, které slouží pro sestavení predikčního modelu na jednotlivých úrovních hierarchické struktury stromu. Podporuje predikci spojitých i diskrétních atributů.
- Analýza časových řad – slouží k hledání trendu a jeho predikci pro budoucnost.
- Neuronové sítě – nevychází ze statistických metod, ale snaží se rozpoznávat vzory a minimalizovat chyby
- Naive Bayes – je velmi rychlý a proto je vhodný na rychlou předběžnou analýzu, pomocí které lze vyloučit málo důležité atributy. Potom se většinou aplikuje náročnější algoritmus.

Na výběr je tedy bohatá paleta různých algoritmů, která by měla uspokojit poměrně rozličné požadavky uživatelů.

Pro vytvoření data miningového modelu existuje poměrně rozsáhlý průvodce, jehož výsledkem je miningový model prezentovaný ve Visual Studiu např.

v podobě zobrazené na obrázku 7.

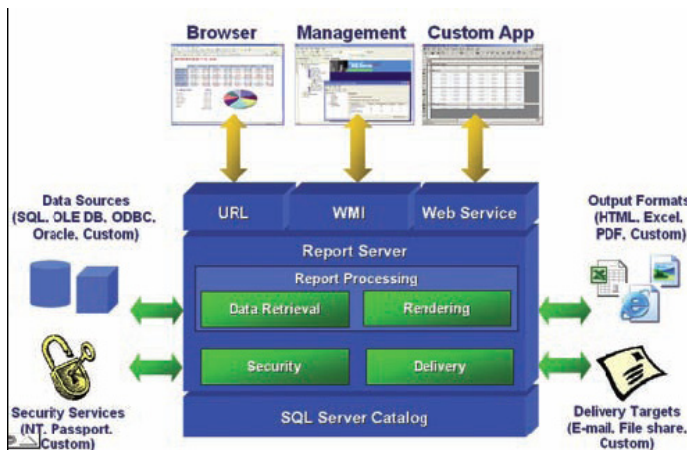


Obrázek 7 - Miningový model ve Visual Studiu

## 9 Reportovací služby

Reportovací služby jsou podle statistik nejpoužívanější oblastí business intelligence v SQL Serveru. Slouží k vytváření jak papírových, tak interaktivních webových prezentací provedených analýz.

Zasazení reportovacích služeb do kontextu jejich IT okolí a samotnou architekturu ukazuje obrázek 8. Základem je SQL Server Catalog, což je SQL databáze, jejíž údaje využívá Report server a taky si tam určité věci ukládá (metadata, definice reportů apod.).



**Obrázek 8 - Architektura a návaznost reportovacích služeb**

Samotný reportovací server je nestavový (podobně jako HTTP) a má nad sebou několik základních aplikačních rozhraní – URL, WMI (pro správu reportovacích služeb) a webové služby.

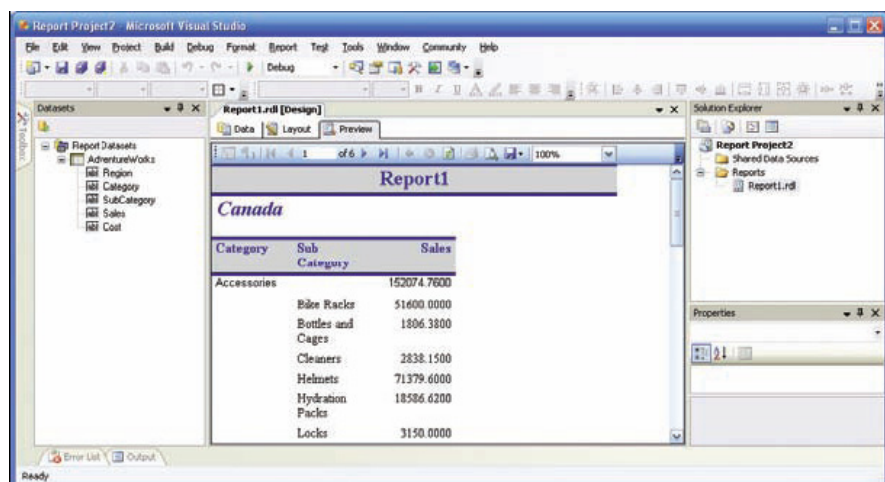
Pro vytvoření reportu, podobě jako u jiných oblastí BI, existuje podrobný průvodce. Report je potom navrhován v prostředí Visual Studio, což ukazuje obrázek 9.

Důležité je, že reportovací služby podporují úplný životní cyklus reportů.

Ten začíná vytvořením reportu, což je typicky práce pro vývojáře nebo specialistu. Podporován je standard RDL (Report Definition Language), takže lze pro authoring použít libovolný nástroj s podporou tohoto formátu.

Report je potom nutné spravovat, tj. zpřístupňovat ho nebo naopak, doručovat uživatelům, vést evidenci apod. Tyto činnosti jsou integrovány do Management Studia.

Doručování koncovým uživatelům podporuje dva scénáře: push i pull. Uživatelé si mohou reporty prohlížet buďto webovým prohlížečem nebo je mohou obdržet mailem.



**Obrázek 9 - Report ve Visual Studiu (preview)**

Klíčovou roli hraje bezpečnost, protože reporty mohou obsahovat cenná data o obchodování. Reportovací služby implementují flexibilní bezpečnostní model založený na rolích.

## 10 Shrnutí

Ve stručnosti byly představeny základní novinky SQL Serveru 2005 na poli business intelligence. Je znát, že se produkt tvořil dlouhé roky, takže beze změny nezůstala žádná oblast. Nejdůležitější konceptuální změnou je integrace UDM do centra architektury všech BI služeb, což umožňuje zamezit redundanci dat v relačních a multidimenzionálních datových strukturách při zachování vysokého výkonu OLAP analýz.

Mezi nejdůležitější změny pro vývojáře pak patří integrace do Visual Studia, což řadu věcí značně usnadňuje.

SQL Server 2005 tedy pro zákazníky BI rozhodně má co nabídnout.

## 11 Zdroje

- [1] Lacko, L.: *Business Intelligence v SQL Serveri 2005*. Microsoft ČR, 2005.
- [2] Microsoft SQL Server: *Business Intelligence Solutions*. Dostupné z WWW: <http://www.microsoft.com/sql/solutions/bi/default.aspx>
- [3] Kadlec, J.: *Top 10 new features in SQL Server 2005*. Dostupné z WWW: [http://searchsqlserver.techtarget.com/tip/1,289483,sid87\\_gci1137301,00.html](http://searchsqlserver.techtarget.com/tip/1,289483,sid87_gci1137301,00.html)
- [4] Microsoft: *SQL Server 2005 Analysis Services Datasheet*. Dostupné z WWW: <http://download.microsoft.com/download/9/9/2/99238f44-7769-4fb6-838f-d2bfc16ec5f2/SQLAnalysisServicesDatasheet.pdf>