

# AUTOMATICKÉ DATOVÉ SKLADY JAKO SKLÁDAČKA

**Namodelovat si vlastní podobu celého business intelligence řešení, datového skladu nebo některé jeho části v jednoduchém nástroji nebo v tabulkovém procesoru – sen, který již není pouhým snem. Lze ho uskutečnit a celé řešení pak nechat vygenerovat ve zvolené databázové platformě a začít ho plnit daty. Řešení se nazývá automatizace datových skladů. Tedy proces automatizovaného vývoje a provozu datového skladu.**

Transformace BI (Business Intelligence) do této podoby přináší rychlou implementaci změn napříč celým datovým skladem – DWH (Data Warehouse) a připravuje dimenzionální datové modely pro různé byznys nasazení. Datové sklady lze stavět podobně jako skládačky. Na jeho architekturu, která nejčastěji mívá tři vrstvy (stage, jádro a data marty) není třeba nic měnit. Soustředit se je nutné na jádro datového skladu, které v sobě skrývá datový model, metadatový popis celého skladu a na nástroje, které umožní datový sklad vytvořit podle určitého vzoru nebo šablony.

## Modelování

Stejně jako stavba domu je i příprava automatizovaného budování DWH spojená s jeho modelováním. Nakonec vhodná vizualizace každého modelu také může zabránit nedorozuměním a nenaplněným očekáváním. Bez modelování by automatizace nebyla možná. Konkrétní datový model by měl vycházet z obchodního modelu společnosti, pro kterou je sklad budován. Ať už se navrhuje pro výrobní společnost, banku, pojišťovnu, leasingovou firmu, mobilního operátora či softwarehouse, jádro DWH bude vždy vypadat jinak. Z obchodního modelu vyplne, jestli centrem pozornosti v DWH budou klientské smlouvy, vyráběné produkty nebo realizované projekty.

Kolem těchto center v modelu potom definujeme košatou síť vzájemných vazeb a popisných dimenzí. Po celou dobu vytváření modelu máme na paměti fakta, která bude třeba uchovávat, a to zejména s ohledem na ukazatele (KPIs), které nám datový sklad bude poskytovat. Toto je v modelování velmi důležité – pokud je faktem např. částka pohledávky, je důležité vědět, jestli je vázána na konkrétní fakturu, smlouvu, nebo jestli můžeme v modelu sledovat pohledávku třeba pouze v detailu na jednotlivého klienta. Výhodou je, že dnes je k dispozici množství vzorových datových modelů pro jednotlivé obchodní modely.

## Metadata

Metadata jsou data o datech. Např. tabulka v reportu obsahuje nějaká konkrétní data (třeba o měsíčních tržbách po produktech). Metadata k této tabulce strukturovaně sdělí, co znamenají řádky (měsíce) a sloupcečky (produkty) této tabulky a co je uvnitř tabulky (tržby v tisících korun).

Na podobném jednoduchém principu je postaven celý komplexní metadatový model datového skladu. Popisuje nejen datový model jádra (a datových martů), ale také všechny potřebné

datové zdroje na vstupu, jejich namapování do tabulek datového modelu, které se pak převádí na jednotlivé atomizované datové transformace a umožňuje popsat jejich závislosti. Například nejprve do skladu načteme změny v produktovém katalogu a teprve potom můžeme načítat fakta o produktech.

## Mapování

Mapováním nazýváme sadu poměrně detailních pravidel, která přesně určují, jaká data budeme ukládat v jádru DWH. Přesněji řečeno, pro každý sloupec každé tabulky jádra musíme určit, z jakého zdroje budeme data načítat. Pro tento účel definujeme atomické datové transformace, kterých mohou být stovky a které mohou být od velice jednoduchých až po velice složité.

Jednoduché transformace je možné strojově generovat opět na základě metadat a složitější je nutné naprogramovat (typicky v SQL). Toto je jediný okamžik, kdy v automatizaci DWH vzniká ručně psaný kód. Důležité je, že tento kód opět uložíme společně s ostatními metadaty a při generování transformačních procedur jej kombinujeme s tím, co je uloženo v šablonách.



## Šablony

Vizualizace již ale hodně souvisí s výslednou podobou řešení. Tato podoba ale není zapsaná pouze v metadatech. Hodně záleží také na podstatných detailech, které jsou uloženy ve vzorech, které obsahují informace např. o konkrétních datových typech, způsobech ukládání historických údajů v různých typech tabulek, způsobech fyzického ukládání dat a mnoho dalších informací.

Šablony obsahují vzory pro všechny databázové objekty, ze kterých bude sklad postavený. Díky tomu jsou šablony specifické pro jednotlivé databázové platformy. Použitím šablon můžeme na základě stejného metadatového modelu vygenerovat stejný datový sklad nejprve pro MS SQL Server a následně jej přenést do prostředí Oracle nebo Teradata. Častější než změna celé databázové



Petr Hájek, Senior Advisor pro Informační Management společnosti Profinit

platformy ale bude třeba změna právě pravidel pro partitioning nebo historizaci dat.

## Generování

Když je hotový metadatový model a vybrané a odladěné šablony datového skladu, lze spustit tisk, nebo přesněji řečeno vygenerování celého řešení ve zvolené databázové prostředí. Mezi výsledkem budou strojově připravené databázové skripty, po jejichž nasazení a spuštění dostáváme skutečně kompletně celé řešení – databázové struktury (nyní ještě prázdné, bez dat) a procedury k jejich naplnění.

Napoprvé nás bude čekat ještě prvotní naplnění skladu (historickými) daty a potom už pravidelné spuštění procedur, pro načítání a aktualizaci nových dat.

## Změny

Skutečná síla automatizovaného přístupu k DWH se ale projeví ve chvíli, kdy je zapotřebí ve skladu provést zásadnější změny. Vytvoření nového data-martu nebo doplnění skladu o nová zdrojová data jsou nejčastěji se vyskytující změny.

Generátor pomocí metadat a šablon může dodat skripty, pomocí kterých se odroluje několik posledních historizačních operací, provede oprava dat a postupně načtou data nová.

## Závěrem

DWA (Data Warehouse Automation) není nutné aplikovat jen na celé datové sklady. Principy DWA lze použít třeba ve finančním nebo obchodním oddělení na chytrý návrh a údržbu vlastního data martu. Je nesporné, že automatizovaný a agilní přístup k tvorbě, provozu a rozvoji DWH přináší snadno měřitelné úspory peněz a času, a to mimo jiné tím, že je tolerantní k chybám v designu. Pokud první, druhá nebo další verze DWH nebyla navržena optimálně, je možné vygenerovat a znovu naplnit datový sklad, a to i s již existujícími daty. ■

Petr Hájek