

Karel Martinec, Jan Pospíšil¹

Datové tržiště nákladní přepravy

Klíčová slova: *DTNP, nákladní přeprava, datové tržiště (data warehouse), statistika*

Úvod

Železniční nákladní přeprava je dosti složitým procesem a její detailní inženýrský popis je proces neméně obtížný.

Z celkového pohledu na nákladní přepravu pokrývá Datové tržiště nákladní přepravy (DTNP) oblast vyhodnocení a statistiky uskutečněných přeprav. DTNP je jedním z modulů informačního systému pro Podporu rozhodování managementu ČD (IS PRM) a nabízí kvalitní, ucelené, formálně sjednocené a v reálném čase dostupné informace o nákladní dopravě. Na jednom místě se tedy uživatelům nabízejí aktuální i historická data umožňující náročné dotazování, analýzy, tvorbu časových řad a trendů.

DTNP vznikalo na přelomu let 1999 a 2000 jako nástupce původního systému, který dávkově (jednou měsíčně bez zpětné kontroly a oprav) zpracovával data získaná přes diskety z jednotlivých stanic a následně z nich vytvářel pevně dané textové výstupy. Při přechodu zpracování z původního systému na nový způsob s využitím technologie datových skladů byl kladen důraz především na rozvoj funkcionalit a možnost tvorby libovolných dotazů.

¹ Karel Martinec, Ing., 1965, Vysoká škola dopravy a spojů Žilina, ČD-Telematika, a. s.
Jan Pospíšil, Ing., 1976, Univerzita Pardubice, Dopravní fakulta Jana Pernera, ČD-Telematika, a. s.

Životní cyklus informací o přepravě a jeho obraz v DTNP

Celý životní cyklus přepravy vozové zásilky od podeje zásilky do okamžiku dodeje je pokryt sledováním v primárním informačním systému Centrální nákladní pokladna (CNP). CNP je tedy základním zdrojem informací pro DTNP. Dalším systémem, který je nezbytný k fungování DTNP je informační systém Odúčtovny přepravních tržeb (OPT) zajišťující tzv. audit informací o přepravě získaných prostřednictvím CNP.

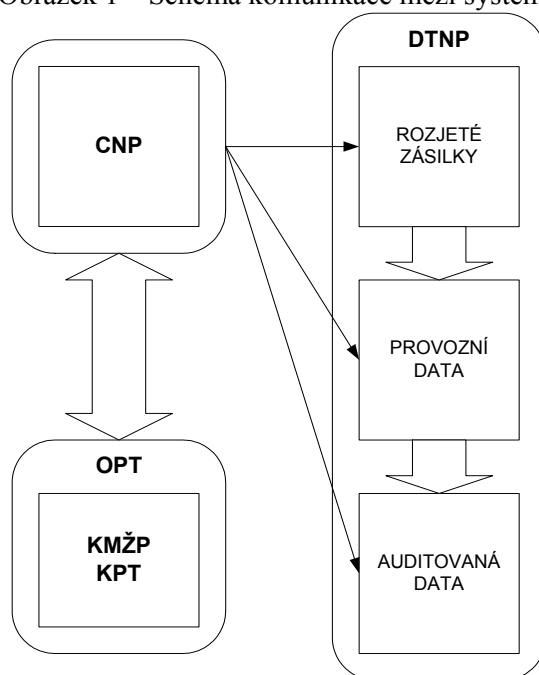
Hlavním dokumentem z něhož se získávají data do primárního informačního systému je nákladní list.

Základní prvky životního cyklu dat v DTNP:

- podej nebo vstup zásilky
 - základní údaje o zásilce jsou zpracovány v CNP a následně předány do DTNP jako zásilky „rozjeté“ (na cestě)
- změna přepravní smlouvy
 - aktualizace dat v CNP a DTNP
- dodej nebo výstup zásilky
 - základní údaje o zásilce jsou aktualizovány a rozšířeny, zpracovány v CNP a následně předány do DTNP jako zásilky „provozní“ (neauditované)
- audit zásilky
 - údaje o zásilce jsou aktualizovány dle stavu auditovaného na OPT

Při každém plnění dat do DTNP jsou prováděny kontroly úplnosti a správnosti vůči číselníkům (dimenzím) a všechny chybové zprávy jsou následně vyhodnocovány a zpracovávány návaznými aplikacemi.

Obrázek 1 – Schéma komunikace mezi systémy



V současné době jsou tedy v DTNP uložena data o přepravách „rozjetých“ (na cestě), „neauditovaných“ (provozní data) a auditovaných. Data o zásilkách „rozjetých“ a data provozní poskytují možnost aktuálního pohledu na stav nákladní přepravy s minimálním časovým odstupem od skutečné události. Nyní je tento odstup nastaven na minimálně 24 hodin od okamžiku poslední

události zásilky. Tento časový odstup je nastaven tak, aby údaje o zásilkách vybraných do DTNP byly dostatečně vypovídající a vhodné ke zpracování.

DTNP obsahuje data, která nabízí uživateli odpověď na základní otázky týkající se nákladní přepravy:

- KDY se přeprava uskutečnila (např. datum, období)
- ODKUD KAM se přeprava uskutečnila (např. stanice, místo, stát)
- CO bylo přepravováno (např. zboží, třída nebezpečnosti)
- ČÍM byla přeprava provedena (např. druh a vlastnictví vozu, typ kontejneru)
- OD KOHO a KOMU se přeprava uskutečnila (např. odesílatel, příjemce, zákazník, plátce)
- PODMÍNKY přepravy (např. zákaznický tarif, typ přepravy)

Aktualizace dat v DTNP probíhá jednou denně dle přesně stanoveného harmonogramu se zajištěním kontroly správnosti jednotlivých kroků.

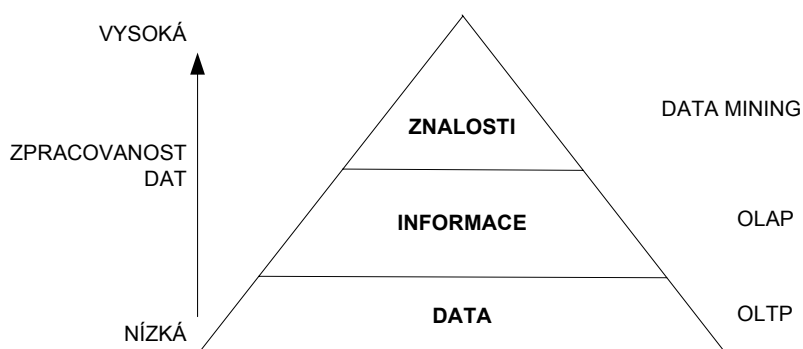
Použitá technologie zpracování dat

Jak již bylo zmíněno dříve zpracovává DTNP velké množství dat popisujících uskutečněné přepravy. Aby toto rozsáhlé množství údajů bylo možno zpracovat, uložit a prezentovat v co nejkratším čase byla zvolena technologie datových skladů (data warehouse). Tato technologie svou filozofií umožňuje zpracování velkého množství vstupních primárních dat a jejich transformaci do multidimenzionálního datového modelu.

Postup zpracování dat:

- 1) „výroba“ dat – primární systémy – technologie OLTP (On Line Transaction Processing)
- 2) skladování dat – technologie datových skladů
- 3) prezentace („prodej“) dat – technologie OLAP (On Line Analytical Processing)

Obrázek 2 – Schématické znázornění struktury business inteligence



Hlavní rozdíly mezi systémy s technologiemi OLTP a OLAP:

- OLTP uchovává a zpracovává velké množství transakcí, provádí se průběžně zápis i čtení dat
- OLAP zpracovává a ukládá data v agregované podobě, zápis dat se provádí většinou jednorázově a čtení i analýza dat se provádí průběžně

Standardní definice datového skladu podle Billa Inmona:

Datový sklad je podnikově strukturovaný depozitář předmětově orientovaných, vzájemně provázaných, časově proměnných, historických dat používaný na získávání informací a podporu rozhodování. V datovém skladu jsou uložena detailní a sumární data.

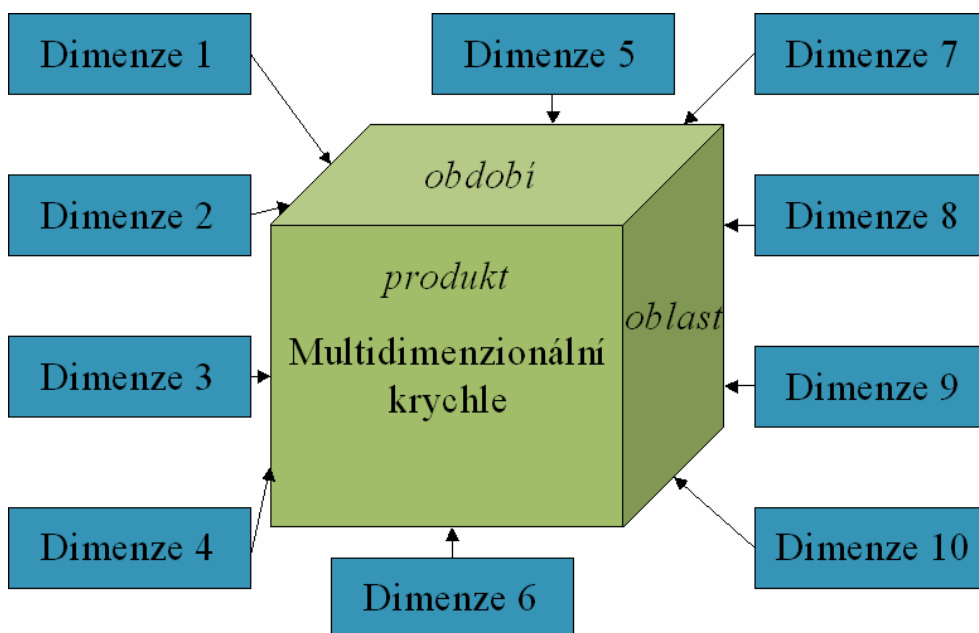
Základní vlastnosti datového skladu:

- předmětová orientace – data se do datového skladu zapisují podle předmětu zájmu, nikoliv podle zdrojové aplikace
- integrovanost – data týkající se konkrétního předmětu se do datového skladu zapisují pouze jednou
- časová variabilita – data zapsaná do datového skladu reprezentují vždy určitý časový úsek
- stálost (relativní) – data v datovém skladu se většinou zásadně nemění ani neodstraňují

Systém obsahující data integrovaná z primárních informačních systémů, data historická, detailní a transformovaná do vhodného tvaru může mít následující architekturu:

- zdroje dat – provozní, technologické, účetní systémy apod.
- pracovní oblast – při přenosu zdrojových dat do této oblasti se provádí čištění a transformace
- datový sklad – slouží pro uložení všech dat odrážející historii i současnost firmy a stává se tak „jediným zdrojem pravdy“
- datová tržiště – vznikají dle potřeb uživatelů a jsou tedy přizpůsobena jejich požadavkům na snadnou dostupnost informací

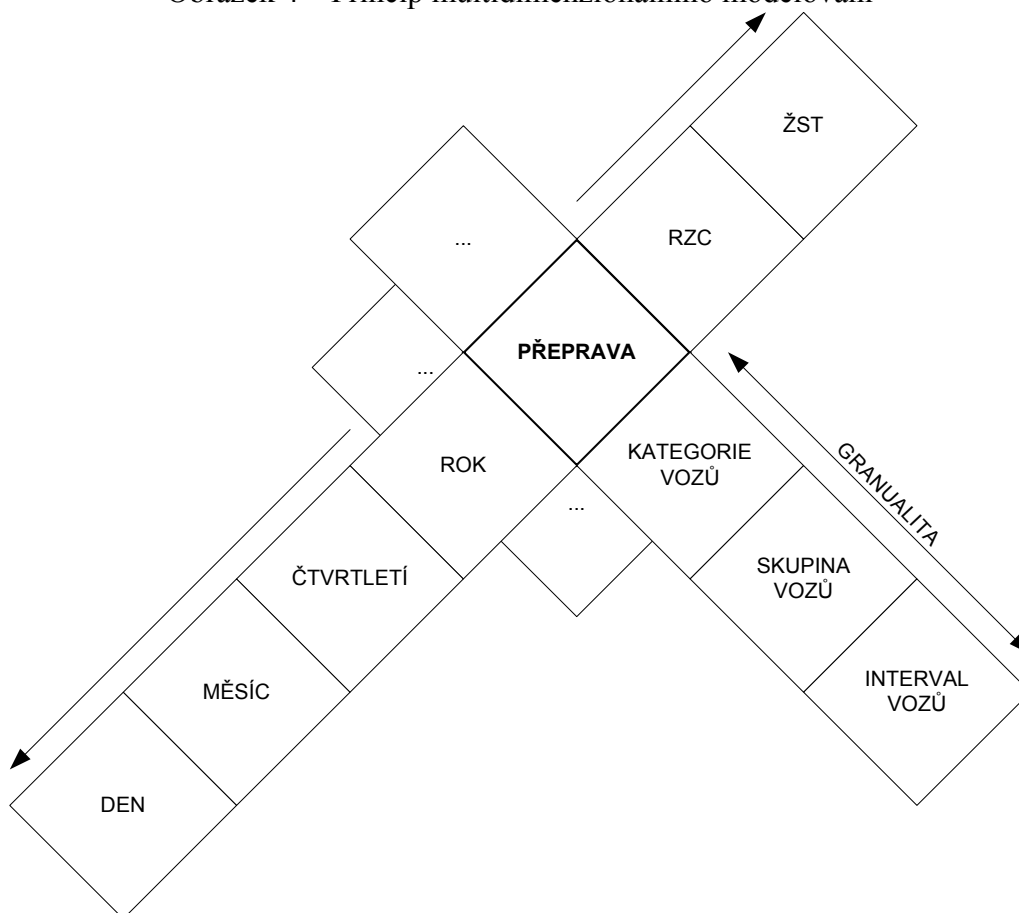
Obrázek 3 – Schéma multidimenzionální krychle



Pro tvorbu datových skladů se většinou využívá standardních relačních databází, ale s využitím multidimenzionálního modelování. Dimenzionální modelování klade důraz na srozumitelnost modelu uživateli. Vzniká tak model složený ze dvou hlavních prvků – tabulek

faktů (obsahují vlastní hodnoty) a tabulek dimenzí (popisující hodnoty obsažené v tabulkách faktů). Princip multidimenzionálního modelu je znázorněn na obrázcích X a X.

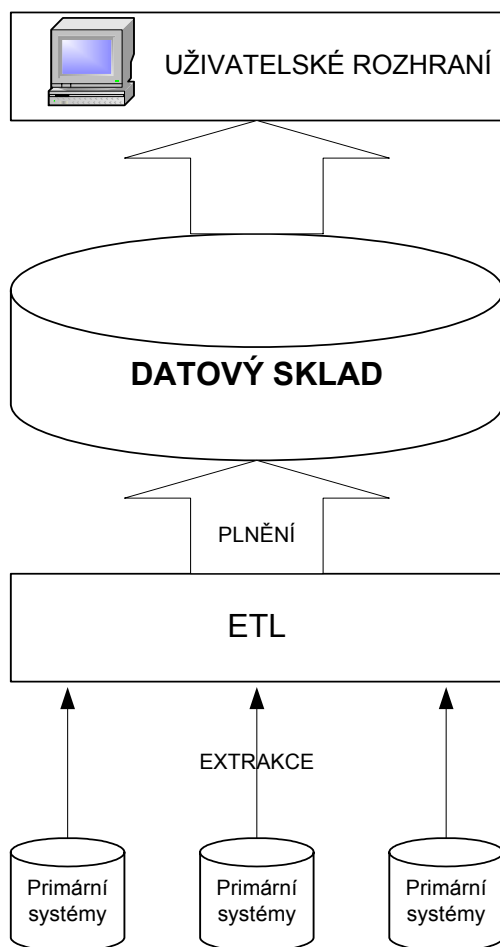
Obrázek 4 – Princip multidimenzionálního modelování



Vlastní naplnění daty ze zdrojů dat do datového skladu se provádí procesem ETL, který se skládá z těchto částí:

1. extrakce – výběr odpovídajících dat z primárních systémů
2. transformace – kontrola správnosti dat vůči dimenzím (číselníkům), integrace a agregace dat
3. plnění (loading) – získaná a zpracovaná data jsou plněna do příslušných tabulek faktů, případně i tabulek dimenzí

Obrázek 5 – Schéma technologie plnění DTNP



Přehled dimenzí

Délkové kategorie PKD

Dimenze slouží ke sledování výkonů v kombinované přepravě. Nejnižší uživatelskou úrovní je kód délky PKD.

Druh přepravy

Dimenze umožňuje sledovat výkony podle druhu přepravy.

Druh vozu

Dimenze umožňuje sledování výkonů z pohledu vozového hospodářství. Nejnižší uživatelskou úrovní je interval vozu.

Hmotnostní stupeň

Dimenze slouží ke sledování výkonů podle hmotnostního stupně.

Zařazení zboží do OS a KS

Dimenze umožňuje sledovat výkony podle definovaných obchodních skupin a komodit.

Místo nakládky/vykládky



Dimenze umožňuje sledovat výkony podle místa nakládky/vykládky. Ve vnitrostátní přepravě a ve vývozu to jsou vlečky, VNVK, kontejnerová překladiště eventuálně soukromé nebo pronajaté železnice. V dovozu a průvozu se místem nakládky/vykládky rozumí země původu.

NHM – zásilka

Dimenze umožňuje sledovat přepravu podle druhu zboží v zásilce. Nejnižší uživatelskou úroveň dimenze je osmimístné číslo NHM.

NHM – PKD

Dimenze umožňuje sledovat kombinovanou přepravu podle druhu zboží prostředku kombinované přepravy. Nejnižší uživatelskou úroveň dimenze je osmimístné číslo NHM.

NHM - obsah PKD

Dimenze umožňuje sledovat kombinovanou přepravu podle obsahu prostředku kombinované přepravy. Nejnižší uživatelskou úroveň dimenze je osmimístné číslo NHM.

NHM – vozidlo

Dimenze umožňuje sledovat přepravu podle druhu zboží na voze. Nejnižší uživatelskou úroveň dimenze je osmimístné číslo NHM.

Odesílatel zásilky

Dimenze umožňuje sledovat výkony podle základních identifikačních údajů odesílatele.

Plátce zásilky

Dimenze umožňuje sledovat výkony podle základních identifikačních údajů plátce.

Příjemce zásilky

Dimenze umožňuje sledovat výkony podle základních identifikačních údajů příjemce.

Kilometrické pásmo

Dimenze slouží ke sledování výkonů podle přepravní vzdálenosti.

Režim vozidla

Dimenze umožňuje sledování výkonů podle vlastníků železničních nákladních vozů.

Nebezpečné věci

Dimenze slouží ke sledování přeprav nebezpečných věcí. Nejnižší uživatelská úroveň jsou kódy nebezpečných věcí.

Zákaznický tarif

Dimenze umožňuje sledovat výkony podle zákaznických tarifů. Nejnižší uživatelskou úroveň je označení zákaznického tarifu.

Nakládka/vykládka

Dimenze umožňuje sledovat výkony podle jednotlivých upřesňujících kritérií stanovených předpisem. Touto dimenzí lze postihnout např. skutečnosti, kdy se přeprava nezapočítává do nakládky a vykládky, ale započítává se do přepravy, zda se jedná o kolejové vozidlo prázdné nebo ložené, zda jde o mimořádnou zásilku

Železniční podnik

Dimenze umožňuje sledování výkonů podle vlastnického železničního podniku železničních nákladních vozů. Nejnižší uživatelskou úroveň je vlastnický železniční podnik.

Statistický tarif

Dimenze umožňuje sledovat výkony podle tarifů (v mezinárodní přepravě pouze v řezu ČD). Nejnižší uživatelskou úroveň je statistické číslo tarifu. Další úrovně jsou tarifní sazebníky a kategorie

Datum přepravy

Dimenze umožňuje sledovat výkony podle skutečného data události, tzn. u vnitrostátní přepravy a dovozu datum dodeje, u vývozu datum podeje a u průvozu datum výstupu z ČD. Nejnižší uživatelskou úroveň dimenze je položka Den, která jednoznačně odpovídá jednomu kalendářnímu dni.

Kalendářní období

Dimenze umožňuje sledovat výkony podle skutečného data události, tzn. u vnitrostátní přepravy a dovozu datum dodeje, u vývozu datum podeje a u průvozu datum výstupu z ČD. Nejvyšší uživatelskou úroveň dimenze je položka Měsíc.

ŽST odesílací/určení

Dimenze umožňuje sledovat výkony podle železničních stanic odesílacích/určení. Nejnižší uživatelskou úroveň dimenze je 6 místné evidenční číslo železniční stanice.

ŽST odesílací/určení v cizině

Dimenze umožňuje sledovat výkony podle železničních stanic odesílacích/určení v cizině. Nejnižší uživatelskou úroveň dimenze je 8 místné evidenční číslo železniční stanice včetně kódu železničního podniku.

Tratě odesílací/určení dle KJŘ

Přiřazení stanic odesílacích/určení na tratě v členění dle KJŘ (jedinečný i duplicitní pohled).

Seznam skupin základních přehledů

Seznam skupin pevně definovaných přehledů poskytne hrubou představu a rozsahu a možnostech reportování z DTNP.

- 501 Přeprava zboží – celkový přehled
- 502 Přeprava zboží a tržby
- 503 Mezinárodní přeprava zboží
- 504 Přeprava zboží podle vozových řad
- 505 Přeprava zboží a tržby – podle jednotlivých RZC
- 507 Tržby za přepravu podle sazebníků
- 509 Kombinovaná přeprava zboží
- 510 Přepravy podle kilometrických pásem a zboží
- 511 Přeprava zboží dle Tabulky č. 3
- 512 Přeprava zboží a tržby dle Tabulky č. 3
- 513 Poměrné složení vozů a zboží
- 514 Přeprava nebezpečných věcí

515 Přeprava zboží a tržby z přepravy podle RZC

521 Přehledy pro UIC

522 Přehledy pro O02 GR ČD

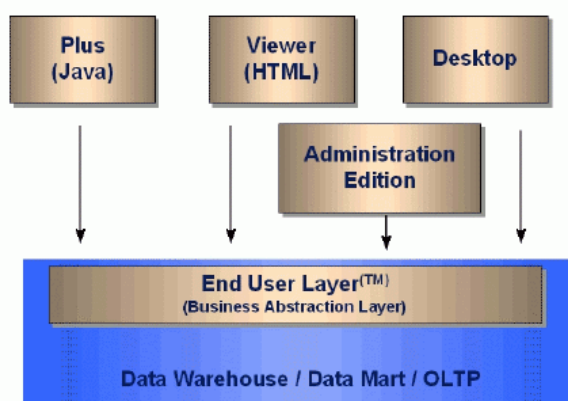
Uživatelské rozhraní

Pro přístup uživatelů k datům uloženým v DTNP se využívá aplikace Oracle Discoverer. Tato aplikace existuje ve více verzích v závislosti na tom, jaký rozsah práce s daty má uživatel povolen a v závislosti na způsobu instalace/komunikace.

Verze aplikace Oracle Discoverer:

- Desktop (aplikace pro koncového uživatele instalovaná na PC)
- Administrator (aplikace pro administrátora)
- Plus (webová aplikace pro koncového uživatele s možností tvorby přehledů)
- Viewer (webová aplikace pro koncového uživatele)

Obrázek 6 – Přehled verzí aplikace Oracle Discoverer



Koncovému uživateli aplikace Discoverer umožňuje libovolným způsobem si prohlížet data zpracovaná procesem ETL do datového skladu a zpřístupněná prostřednictvím vrstvy koncového uživatele (EUL – End User Layer) a v ní vytvořené pracovní oblasti (BA – business area). Pracovní oblast obsahuje tabulky faktů a dimenze upravené již do uživatelsky přívětivé podoby (názvosloví, popis, nápověda) včetně vytvořených vazeb mezi jednotlivými tabulkami faktů a dimenzemi. V pracovní oblasti jsou také vytvořena měřítka tzv. odvozená a vypočítaná na základě měřítek (hodnot) základních. Dimenze upravené pro přístup koncových uživatelů také obsahují nedefinované hierarchie a předvolené filtry.

V současné době DTNP obsahuje 2 tabulky faktů (hodnoty s měsíční a denní podrobností) a 33 dimenzí (pohledů) z nichž některé se ještě dále člení.

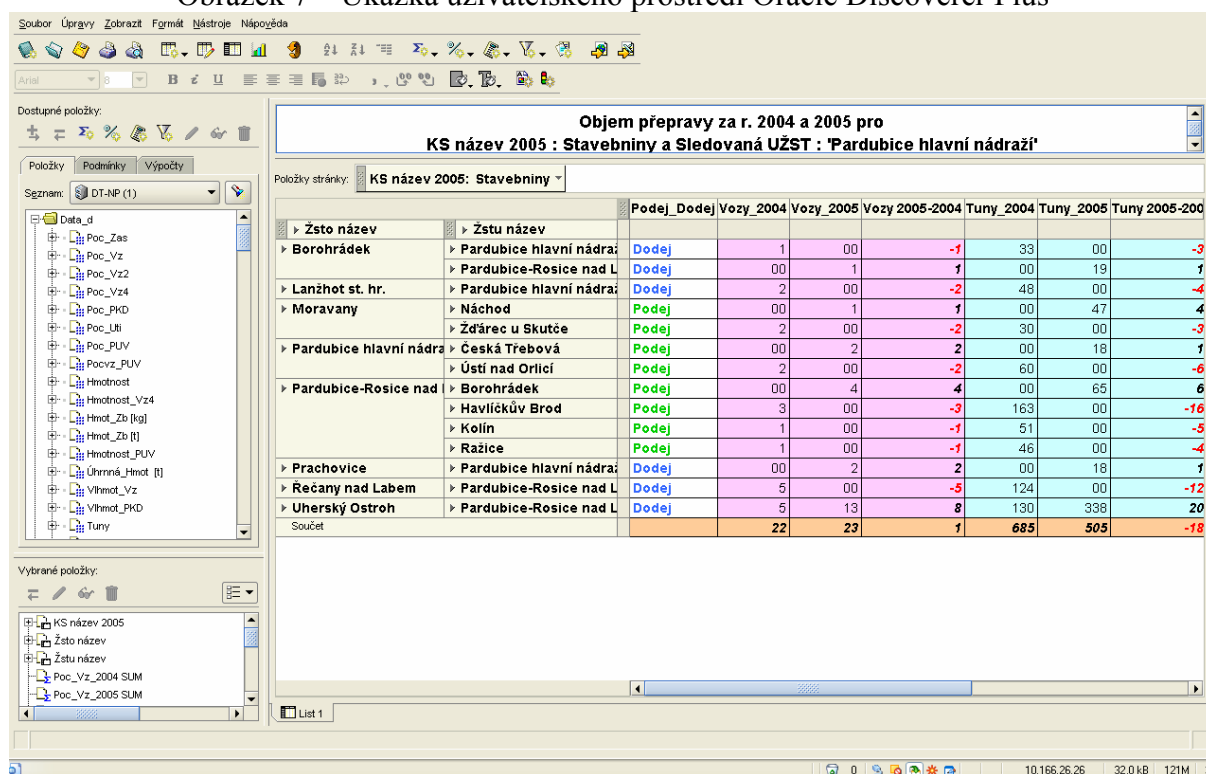
Základním prvkem jehož prostřednictvím se uživateli zobrazují data dle definovaného dotazu je tabulka – jednoduchá či křížová, případně doplněná o položky stránky umožňující rozdělit či spojit větší množství dat obsažených ve výsledku dotazu.

Uživatelský postup vytvoření vlastního dotazu (pohledu):

- 1) výběr měřítek z tabulky faktů

- 2) výběr položek z dimenzí
- 3) uspořádání a orientace položek v návrhu
- 4) volitelné vytvoření filtru či parametru pro omezení vybraných hodnot z položek dimenzí
- 5) volitelné vytvoření grafu (výběr ze standardní nabídky – sloupcový, koláčový, bodový, křivkový apod.)

Obrázek 7 – Ukázka uživatelského prostředí Oracle Discoverer Plus



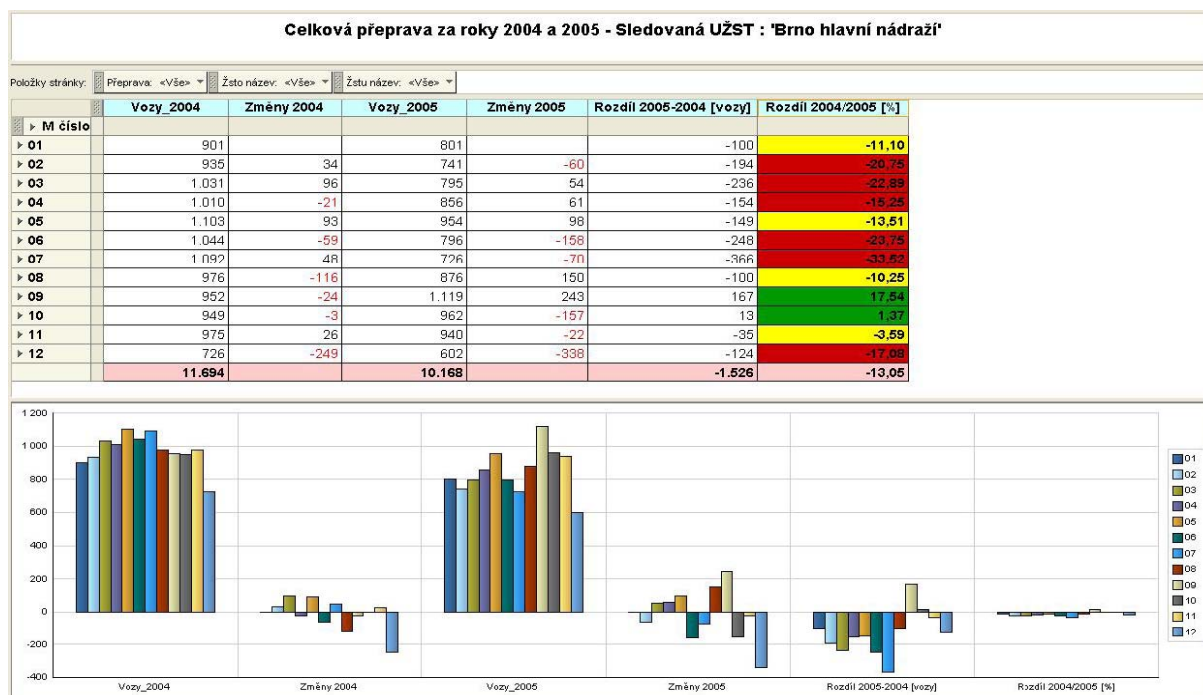
The screenshot shows the Oracle Discoverer Plus interface. The main window displays a table titled "Objem přepravy za r. 2004 a 2005 pro KS název 2005 : Stavebniny a Sledovaná UŽST : 'Pardubice hlavní nádraží'". The table has columns for "Podej_Dodej", "Vozy_2004", "Vozy_2005", "Vozy 2005-2004", "Tuny_2004", "Tuny_2005", and "Tuny 2005-2004". The data is organized hierarchically by station names.

Žsto název	Žstu název	Podej_Dodej	Vozy_2004	Vozy_2005	Vozy 2005-2004	Tuny_2004	Tuny_2005	Tuny 2005-2004
▶ Borohrádek	▶ Pardubice hlavní nádraží	Dodej	1	00	-1	33	00	-3
	▶ Pardubice-Rosice nad L	Dodej	00	1	1	00	19	1
▶ Lanžhot st. hr.	▶ Pardubice hlavní nádraží	Dodej	2	00	-2	48	00	-4
▶ Moravany	▶ Náchod	Podej	00	1	1	00	47	4
	▶ Zďárec u Skutče	Podej	2	00	-2	30	00	-3
▶ Pardubice hlavní nádraží	▶ Česká Třebová	Podej	00	2	2	00	18	1
	▶ Ústí nad Orlicí	Podej	2	00	-2	60	00	-6
▶ Pardubice-Rosice nad L	▶ Borohrádek	Podej	00	4	4	00	65	6
	▶ Havlíčkův Brod	Podej	3	00	-3	163	00	-16
	▶ Kolín	Podej	1	00	-1	51	00	-5
	▶ Ražice	Podej	1	00	-1	46	00	-4
▶ Prachovice	▶ Pardubice hlavní nádraží	Dodej	00	2	2	00	18	1
▶ Rečany nad Labem	▶ Pardubice-Rosice nad L	Dodej	5	00	-5	124	00	-12
▶ Uherský Ostroh	▶ Pardubice-Rosice nad L	Dodej	5	13	8	130	338	20
Součet			22	23	1	685	505	-18

Výsledný přehled je možné libovolně upravovat a zpracovávat použitím následujících funkcí:

- řazení
- součtování
- statistické, matematické, procentuální výpočty
- drillování v položkách u nichž existuje hierarchie
- podmíněné formátování jednotlivých položek
- grafické formátování sestavy (nadpisy, záhlaví, zápatí)
- exportování do formátu XLS, PDF, TXT, odeslání e-mailem
- uložení do databáze
- sdílení jiným uživatelům

Obrázek 8 – Příklad exportovaného výstupu



Geografická prezentace

Součástí DTNP je základní mapová prezentace spočívající v zobrazování pevně definovaných map vytvořených nad daty DTNP. Tyto mapy jsou vyrobeny v prostředí MapInfo a zobrazovány buď prostřednictvím aplikace MapViewer nebo webové aplikace MapExtreme. Možností je také export do formátu PDF nebo některého z běžných grafických formátů.

Možností a omezení obou přístupů:

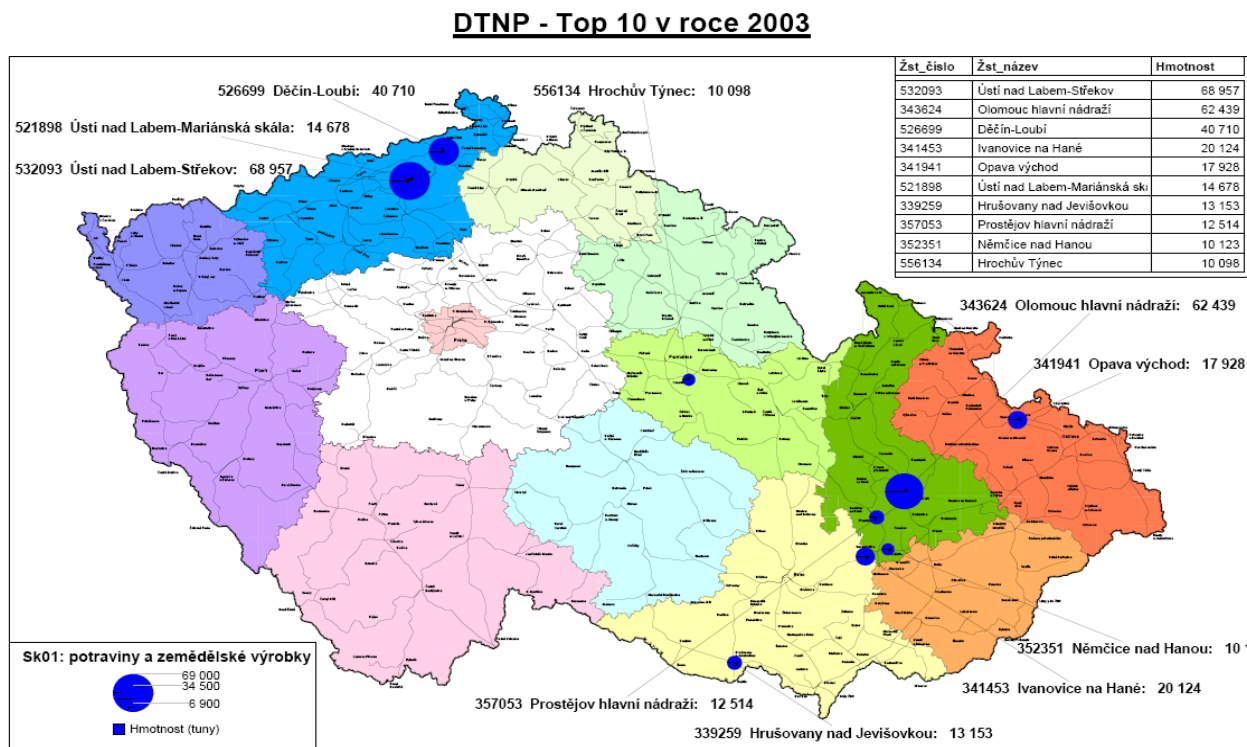
- Mapinfo:
 - možnost úpravy vzhledu mapy
 - uživatelský komfort
 - import dat (možnost práce bez přímého spojení s DTNP)
 - kvalita tiskových výstupů
 - rychlost kreslení mapy
 - práce s rozsahy dat
 - exporty map (PDF, běžné grafické formáty)
- aplikace WWW:
 - přístup omezený pouze přístupem k Internetu
 - snadná dostupnost dalších Web aplikací (IS KDS ČD)
 - aktuální data
 - nekvalitní tiskové výstupy
 - omezené možnosti uživatelského rozhraní (úpravy vzhledu mapy)
 - rychlost (delší časové odezvy při každé akci)

A z nich vzplývajících zaměření obou přístupů:

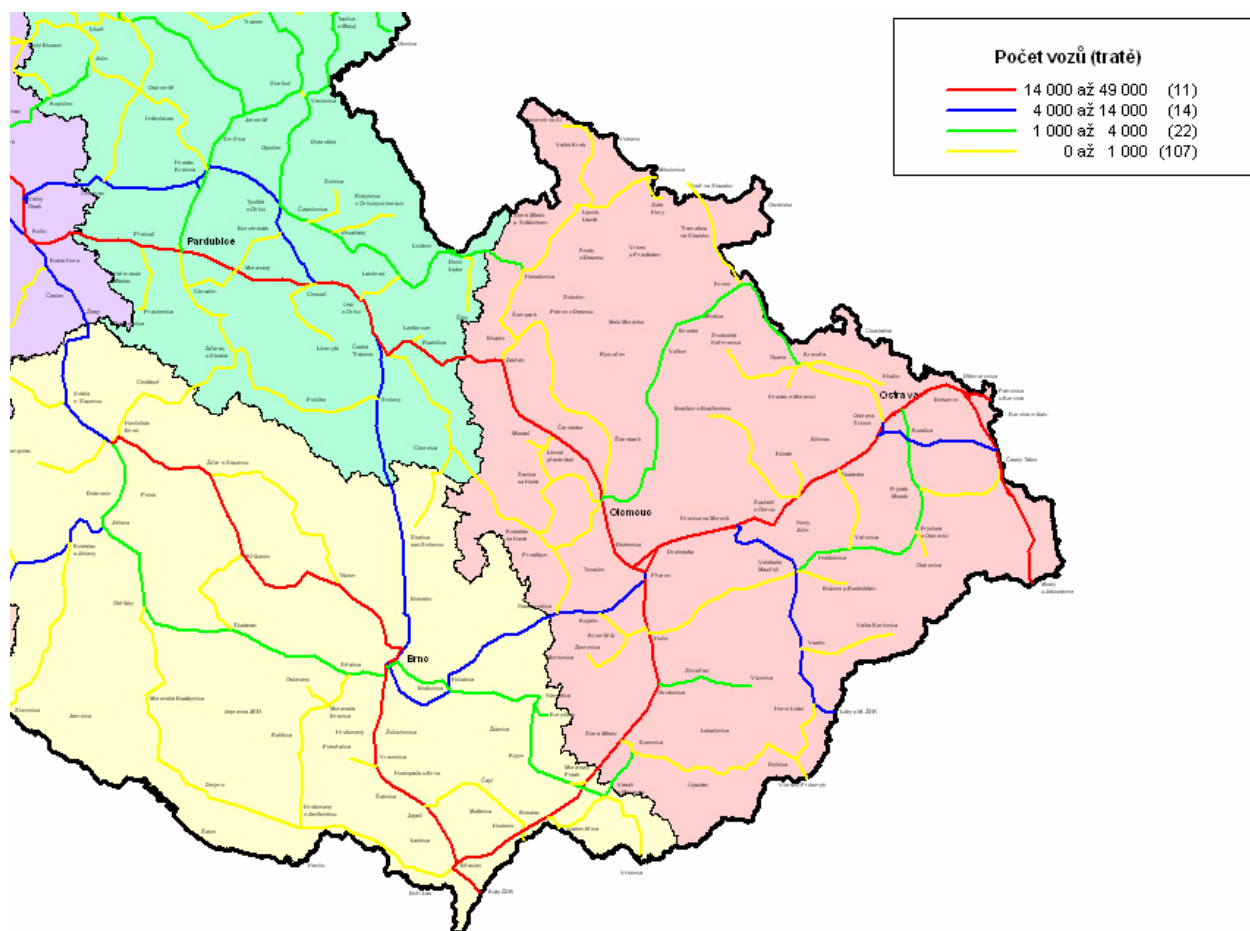
- Mapinfo

- interaktivní manažerský nástroj
 - zdroj papírových výstupů
- aplikace WWW
- prohlížení předdefinovaných dotazů
 - získání základní informace o objektu analýzy

Obrázek 9 – Příklad kombinovaného výstupu (mapa a tabulka) do formátu PDF



Obrázek 10 – Příklad jednoduchého výstupu (mapa) do formátu GIF



Závěr

DTNP je systém, který za téměř 7 let svého rutinního provozu i vývoje nashromáždil značný objemem aktuálních i historických dat, jejichž využití je plně v rukou uživatelů. Záleží především na znalostech a šikovnosti uživatele jaká data ze systému získá k dalšímu využití.

V nejbližším období se vývoj DTNP zaměří na rozšiřování a zkvalitňování datové základny, zlepšování uživatelského rozhraní a integraci s dalšími systémy v oblasti nákladní dopravy a přepravy. Uživatelské rozhraní se bude měnit do podoby portálu datových tržišť, součástí bude interaktivní mapová prezentace a celkově bude systém směřován k business intelligence, tedy převod informací na znalosti a jejich využití k podpoře rozhodování.

Použitá literatura

- [1] Lacko B. Databáze: datové sklady, OLAP a dolování dat, Computer Press, 2003.
- [2] <<http://www.oracle.cz>>
- [3] Dokumentace a uživatelská příručka k DTNP, ČD-Telematika, a. s.
- [3] Volně dostupné zdroje na Internetu.

V Praze, listopad 2006

Lektoroval: Ing. Irena Sluková
GŘ ČD a.s., Odbor strategie a informatiky