

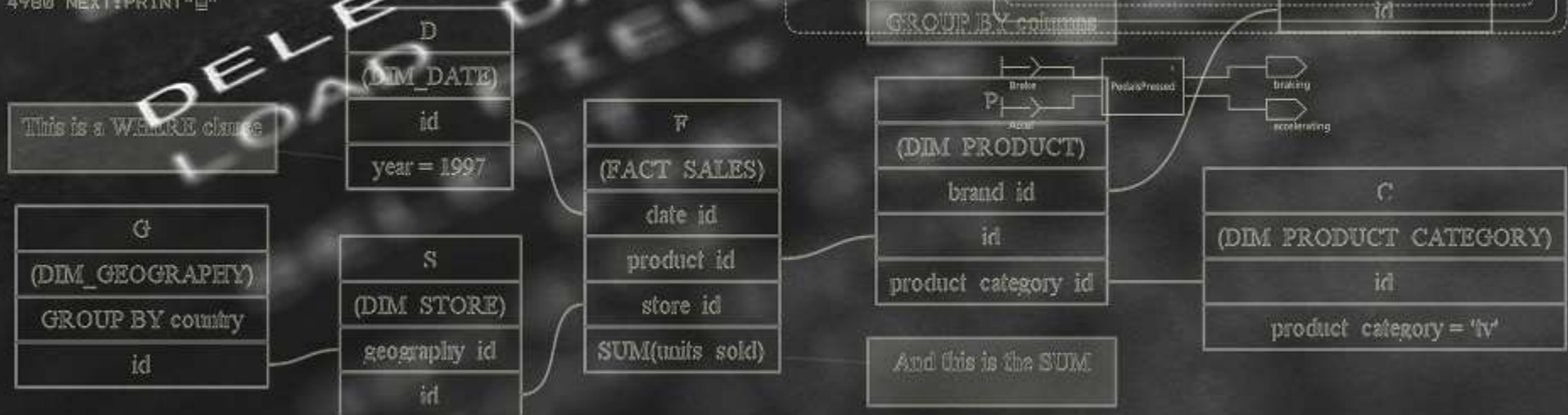
```

4780 GOTO 5000
4790 :
4800 REM
4801 REM
4802 REM
4803 REM
4810 :
4820 PRINT
4825 W=V+1
4830 FOR X
4835 FOR I
4840 PRINT
4850 NEXT:
4860 PRINT
4870 FOR I
4880 IF MD
(I+1);:GOTO
4890 PRINT
4900 NEXT
4910 PRINT
4920 FOR I
4925 PRINT
4930 IF MD
Q";:GOTO 4
4935 PRINT
4940 NEXT:PRINT "
4950 PRINT"#####";
4960 FOR I=2 TO 24 STEP 2
4965 PRINT"|";
4970 IF MD*(I+W-1)="##### THEN PRINT"
M$(I)"|";:GOTO 4980
4975 PRINT M$(I);
4980 NEXT:PRINT"

```

Integrace dat

RNDr. Ondřej Zýka



Obsah

- Kategorizace integračních přístupů
- Kroky integrace a řešení problematických stavů
- Master Data Management

Datová integrace

Synchronní

- Akceptovaný požadavek na primárním systému je akceptován na všech systémech.
- Primární systém nahlásí chybu, pokud nahlásí chybu kterýkoliv z integrovaných systémů.
- Všechny strany vidí najednou stejná data.
 - Technicky nerealizovatelné
- Výkon odpovídá nejslabšímu článku systému
- Aby proběhla transakce, musí být celý systém funkční
- Realizováno nejčastěji jako Dvojfázový commit

Asynchronní

- Akceptovaný požadavek se přenese na všechny systémy, tam není zaručena jeho akceptace. Může tam vyvolat chybu.
- Všechny strany dostanou všechny požadavky.
- Průchodnost jak infrastruktura dovolí.
- Výpadek cílového systému neovlivní schopnost zadat požadavky.
- Různé typy poštovních (messaging) systémů

Datová integrace

ETL, ELT

- Extract-Transform-Load
- Extract-Load-Transform
- Dávkové zpracování
- Podpora složitých transformací
- Full load, přírůstkový load
- Primárně pro Datový sklad

Replikace

- Replikace datových prostorů
- Replikace na úrovni transakcí
- Malé možnosti transformací
- Real-time integrace
- Vyžaduje vyspělejší databáze s transakčním logem
- Asynchronní integrace

Datová integrace

Materializované úložiště

- Vzniká nové úložiště integrovaných dat
- Umožňuje výpočetně náročné algoritmy integrace
- Dotazy na integrovaná data jsou rychlé, zvládají velké množství dotazů
- Příklady
 - DWH
 - ODS

Virtuální pohledy

- Pouze metadata o modelech, vazbách a transformacích
- Data se získávají a transformace se provádějí až při dotazu
- Není třeba udržovat integrovaná data (velikost, výpočtová náročnost, aktuálnost)
- Požadovaný výkon je daný počtem dotazů
- Pouze malý počet dotazů
- Příklady
 - Dohled a provoz

Datová integrace

Federation

- Systém umožňuje (vynucuje) aby požadavky vznikaly jeho prostřednictvím a sám se stará o veškerou komunikaci s jednotlivými systémy. Výsledky předává uživateli.
- Příklady
 - MDM aplikace
 - ESB

Mediation

- Reaguje se na změny v jednotlivých systémech a ty se předávají ostatním systémům
- Příklady
 - Messaging
 - Replikace

Datová integrace

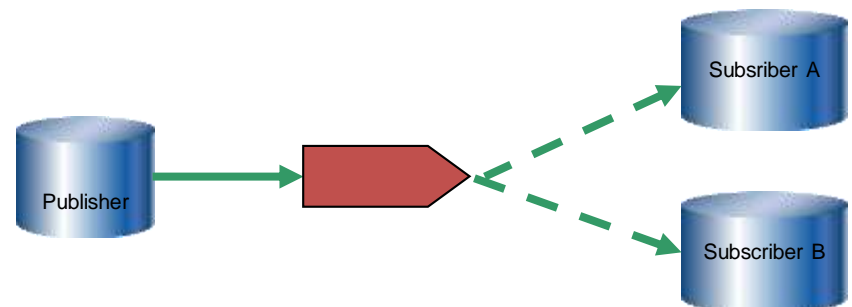
Sender – Receiver

- Zdroj zná své cíle
- Zdroj je schopen reagovat na zprávy od cíle
- Cíl je schopen informovat zdroj
 - Chybná zpráva
 - Žádost o opakování
 - Žádost o synchronizaci (všechna data)



Publisher – Subscriber

- Zdroj se nezajímá o cíle, množství a typy cílů zdroj nijak neovlivňují
- Cíl může odebírat data bez znalosti zdroje
- Cíl nemá zaručeno, že má všechny data
- Cíl nemůže požádat o opětovnou synchronizaci



Datová integrace

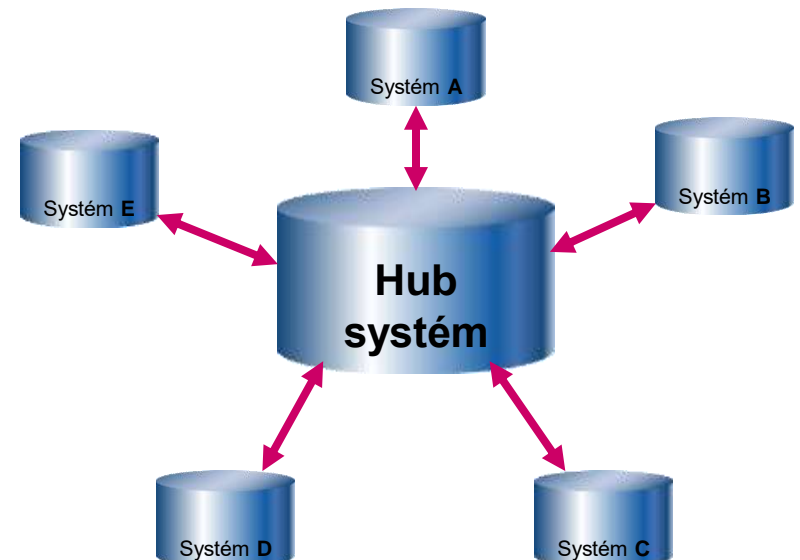
Point-to-point

- Přímá komunikace mezi systémy navzájem
- Každý systém má mnoho partnerů



Hub and Spoke

- Každý systém komunikuje pouze s centrální systémem (Hub)
- Různé technologické úrovně řešení HUBu, materializované i virtuální data
- Příklady: ESB, MDM, ODS, Informatica Data Hub



Datová integrace

Short-live transaction

- Rychlost transakcí závisí pouze na výkonu infrastruktury.
- Provedení maximálně v řádu sekund
- Výpadek infrastruktury transakci ukončí.
- Používá se rollback
- Například databázová transakce

Long-live transaction

- V rámci transakce je možná interakce uživatele
- Může trvat i jednotky dnů
- Transakce přežije výpadek infrastruktury
- Používá se opravný kód
- Například transakce v BPM systémech

Granularita integrace

Full (business) object

- Informace vždy o celém objektu
- Snadná inicializace
- Snadné řešení relačních vazeb a konzistencí
- Nutnost zpracovat celý objekt ve zdroji a cíli
- Vysoké nároky na přenosovou kapacitu

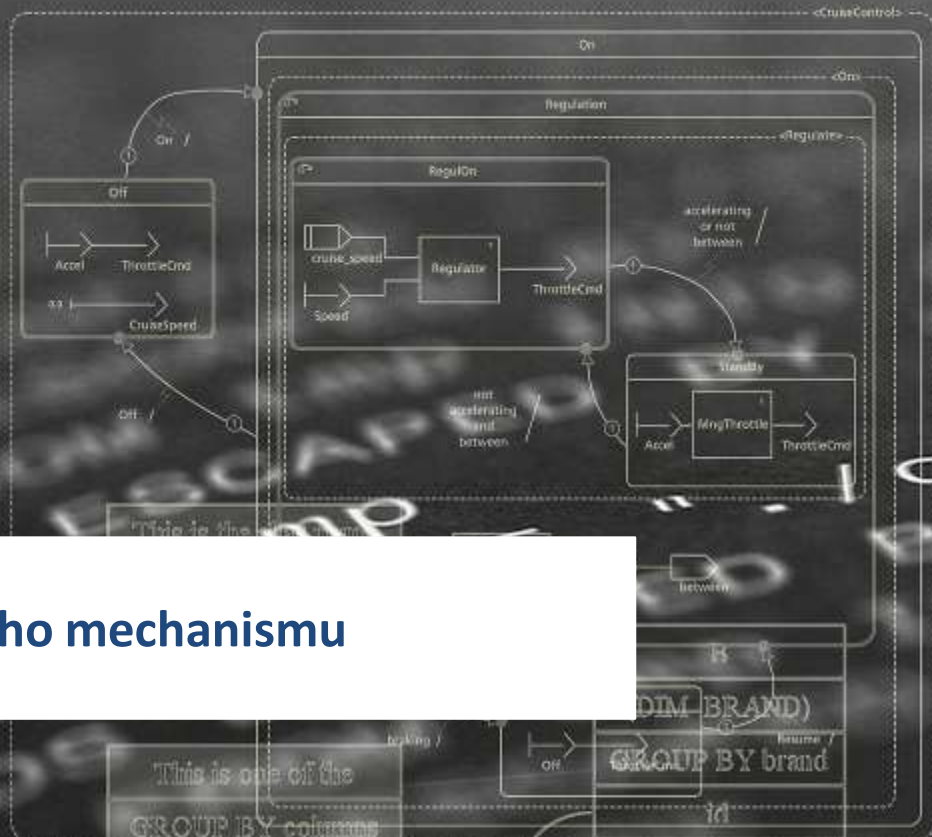
Data record

- Snadná identifikace změn
- Jednodušší způsob získávání stavu před a po změně
- Veliké množství malých zpráv - nároky na režii přenosů
- Vysoké nároky na ověření integrity na cílové straně

```

4780 GOTO 5000
4790 :
4800 REM -----
4801 REM --- DARSTELLUNG ---
4802 REM --- DES MANUALS ---
4803 REM -----
4810 :
4820 PRINT" ";
4825 W=V+1:IF W<8 THEN W=W+14
4830 FOR X=1 TO 2:PRINT"XXXXXXXXXXXX";
4835 FOR I=0 TO 23
4840 PRINT MD$(I+W);
4850 NEXT:PRINT:NEXT
4860 PRINT"XXXXXXXXXXXX";
4870 FOR I=0 TO 23
4880 IF MD$(I+W)=CHR$(32) THEN PRINT M$(
(I+1));:GOTO 4900
4890 PRINT MD$(I+W);
4900 NEXT
4910 PRINT:PRINT"XXXXXXXXXXXX";
4920 FOR I=2 TO 24 STEP 2
4925 PRINT"|";
4930 IF MD$(I+W-1)=
Q";:GOTO 4940
4935 PRINT" ";
4940 NEXT:PRINT"|"
4950 PRINT"XXXXXXXXXXXX";
4960 FOR I=2 TO 24
4965 PRINT"|";
4970 IF MD$(I+W-1)=
M$(I)"|";:GOTO 4980
4975 PRINT M$(I);
4980 NEXT:PRINT"|"

```



Kroky integračního mechanismu

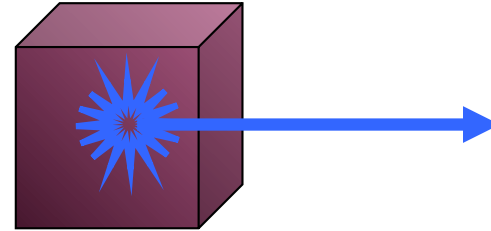


Kroky integrace

- Identifikace změny
- Přenos nového záznamu
- Přenos změny záznamu
- Přenos zrušení záznamu
- Problematika více systémů
- Integrace na základě času
- Integrace na základě datové kvality
- Řešení nedostupnosti dat

Identifikace změny

- Indikace změn pomocí
 - Času
 - Změna Timestamp od posledního přenosu
 - Není možná identifikace fyzického zrušení záznamu
 - Fronty událostí
 - Řešeno programátorsky (triggery zapisující změny)
 - Zvyšuje zatížení databáze
 - Aplikačně – aplikace vytváří frontu změn
 - Nemožnost zaznamenat změny vzniklé mimo aplikaci
 - Technologicky – replikace na základě transakčního logu
- Indikace rozsahu změn
 - Logická úroveň - objekt/záznam
 - Technologická úroveň - řádek, sloupec
- Shromažďovaná data o změně
 - Identifikace změny
 - Iniciátor změny
 - Nové hodnoty dat
 - Původní hodnoty dat



Přenos nového záznamu



- Nový záznam ve zdrojovém systému
- Výsledek v cílovém systému
 - Akceptace
 - Vznik neúplného záznamu
 - Vznik nekonzistentního záznamu
 - Vznik duplicitního záznamu
- Řešení
 - Lepší analýza
 - Odmítnutí záznamu
 - Dočasný zápis
 - Validáční proces

Přenos změny záznamu



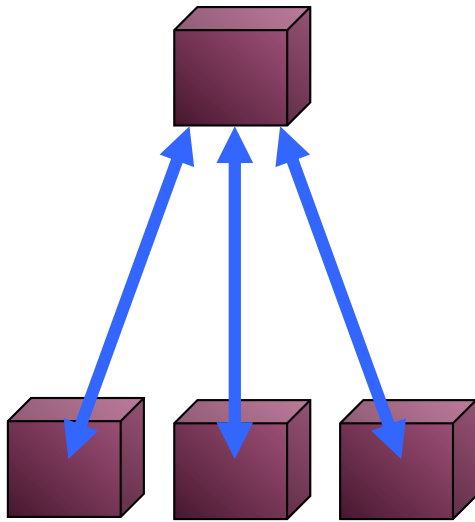
- Změna záznamu ve zdrojovém systému
- Nutno rozlišovat, zda systém data nedodal, nebo zda null je správná hodnota.
- Výsledek v cílovém systému
 - Porušení konzistence záznamu
 - Vytvoření duplicity
 - Vytvoření neúplného záznamu
 - Nerozpoznání měněného záznamu (ztráta informace o změně)
- Řešení
 - Lepší analýza
 - Odmítnutí záznamu
 - Dočasný zápis
 - Validační proces

Přenos zrušení záznamu



- Zrušení záznamu
- Mnoho typů zrušení záznamu
 - neaktivní
 - dokončený
 - zrušený
 - fyzický delete
- Výsledek
 - Vznik nekonzistencí
- Řešení
 - Lepší analýza
 - Logické zrušení (více typů – mapování na stavy zdrojových systémů)
 - Fyzické zrušení

Problematika více systémů



Při integraci více systémů, vznikají pro uživatele nové typy otázek:

- Který systém má pravdu?
- Proč má pravdu?
- Jaké jiné hodnoty jsou v některém systému zadány?
- Jaké hodnoty byly v kterém systému v minulosti?
- Na základě jakých příčin se měnily dat v jednotlivých systémech?
- Kde a kdy vznikly data, která používám?

Integrace na základě času

- Princip: Novější údaje jsou přesnější
- Co je čas údaje
 - Zadání do primárního systému
 - Doba přenesení do cílového systému
 - Jak řešit současnou změnu dat ve více systémech najednou?
 - Nutnost přesné synchronizace času mezi systémy
- Na co se vztahuje identifikace času
 - Na celý záznam
 - Na jednotlivé datové položky

Integrace na základě datové kvality

- Princip: Pro každou hodnotu je definována datová kvalita (jak důvěryhodná je)
- Definice kvality na úrovni systému
 - Data z interního systému jsou důvěryhodnější než data z webu.
- Definice kvality na úrovni atributů
 - To platí, ale pro hodnotu emailové adresy jsou data z webu důvěryhodnější.
- Definice kvality na úrovni typů dat
 - Pro korporátní uživatele jsou data z interního systému důvěryhodnější, pro ostatní uživatele je tomu naopak.
- Definice kvality na základě času
 - Data ze systémů vymáhání jsou nejkvalitnější, ale pouze v době vymáhání.

Příklad použití datové kvality

Nižší hodnota – kvalitnější data

Account information history

SRC	Scheduled time	DQ	Real time	DQ	Scheduled aircraft type	DQ	Real aircraft type	DQ
SC	Sep 21 2004 9:05PM	30		99	M83	30		99
FO	Sep 21 2004 9:05PM	20		99	M83	15		99
MD	Sep 21 2004 9:05PM	10		99	M84	7		99
AG	Sep 21 2004 9:05PM	8	Sep 21 2004 9:00PM	20		99	M83	20
RL		99	Sep 21 2004 9:00PM	12		99		99
SI		99	Sep 21 2004 8:59PM			99	M83	5
MR		99	Sep 21 2004 8:59PM	6		99	M83	6

Zrušení informace v primárním systému

Master Data Management

- Procesy a nástroje k přesné integraci nejdůležitějších entit pro podnik.
- Zákazníci
 - Správa všech partnerů
 - Role a vazby (Hausholding, rodinné vazby, ekonomicky spjaté subjekty, obohacení o externí informace, scoring, ...)
- Správa produktů
 - Jednotné řešení pro všechny oddělení a uživatele
 - Dodavatelé, Marketing, Partneři, Interní systémy, Konsolidace produktů, Prodej, Finance
 - Jednotné řešení pro všechny procesy
 - Design, Nacenění, Náklady, Reporting, Obchod
- Správa centrálních číselníků
 - Jednotné řízení všech číselníků podle kterých se reportuje ve všech odděleních (číselníky pro controlling)
 - Umožňuje: Historizace, plánování, různé verze pravdy, propagace do systémů

Dopady integrace systémů

- Integrací vzniká z jednotlivých systémů řešení nové kvality.
- Nutno zohlednit
 - požadavky na dozor
 - nutnost komunikace se správci jednotlivých systémů
 - údržba jednotlivých systému
 - vytvoření adekvátní organizační struktury
 - řízení změn je nutné na úrovni všech integrovaných systémů
- !! !! Zásah do libovolného systému se může projevit jako závažný problém v ostatních systémech.

Rizika projektů integrace

- Testování
 - Testování je složité a časově náročné
 - Často nutnost míchání různá testovací a produkční prostředí
 - Nutnost zapojení testerů (automatů) do všech systémů
- Nasazení
 - Nemožnost paralelního běhu
- Provoz - nutnost přípravy na výskyt neočekávaných stavů
 - nepředpokládané interakce
 - smyčky v přenosu
 - vzájemné ovlivňování systémů
 - změna chování uživatelů
- Bezpečnost
 - ztráta informací
 - neautorizované modifikace
 - právní odpovědnost
 - pravdivost informací
 - původ informací
 - krádež služeb
 - ztráta důvěry zákazníků
 - příležitost pro fraud

Co si zapamatovat

- Kategorizace integračních přístupů
- Rozdíl mezi synchronní a asynchronní integrací
- Jaké techniky se používají při indikaci dat, které je nutno přenášet v rámci integrace
- Jaké jsou hlavní problémy při zrušení záznamu v integračním systému
- Jak se používá datová kvalita při integraci dat z více systémů
- Co to je Master Data Management (MDM)
- Jaká rizika integračních projektů znáte


```

4780 GOTO 5000
4790 :
4800 REM -----
4801 REM --- DARSTELLUNG ---
4802 REM --- DES MANUALS ---
4803 REM -----
4810 :
4820 PRINT" ";
4825 W=V+1:IF W<8 THEN W=W+14
4830 FOR X=1 TO 2:PRINT"XXXXXXXXXXXX";
4835 FOR I=0 TO 23
4840 PRINT MD$(I+W);
4850 NEXT:PRINT:NEXT
4860 PRINT"XXXXXXXXXXXX";
4870 FOR I=0 TO 23
4880 IF MD$(I+W)=CHR$(32) THEN PRINT M$(
(I+1));:GOTO 4900
4890 PRINT MD$(I+W);
4900 NEXT
4910 PRINT:PRINT"XXXXXXXXXXXX";
4920 FOR I=2 TO 24 STEP 2
4925 PRINT"|";
4930 IF MD$(I+W-1)="  " THEN PRINT"
";:GOTO 4940
4935 PRINT " ";
4940 NEXT:PRINT" "
4950 PRINT"XXXXXXXXXXXX";
4960 FOR I=2 TO 24 STEP 2
4965 PRINT"|";
4970 IF MD$(I+W-1)="  " THEN PRINT"
"
M$(I)" ";:GOTO 4980
4975 PRINT M$(I);
4980 NEXT:PRINT" "

```



Diskuse

- Otázky
- Poznámky
- Komentáře
- Připomínky

