

```
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);:GOT
```

```
4890 PRINT
```

```
4900 NEXT
```

```
4910 PRINT
```

```
4920 FOR I
```

```
4925 PRINT
```

```
4930 IF MD
```

```
Q";:GOTO 4
```

```
4935 PRINT
```

```
4940 NEXT:PRINT
```

```
4950 PRINT"XXXXXXXXXXXX";
```

```
4960 FOR I=2 TO 24 STEP 2
```

```
4965 PRINT" | ";
```

```
4970 IF MD$(I+W-1)="000 000" THEN PRINT"0"
```

```
MS$(I)"0";:GOTO 4980
```

```
4975 PRINT MS$(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.
- Všechny strany vidí najednou stejná data.
 - Technicky nerealizovatelné
- Výkon odpovídá nejslabšímu článku systému
- Aby proběhla transakce, musí být celý systém funkční
- Dvojfázový commit

Asynchronní

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

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

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

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

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

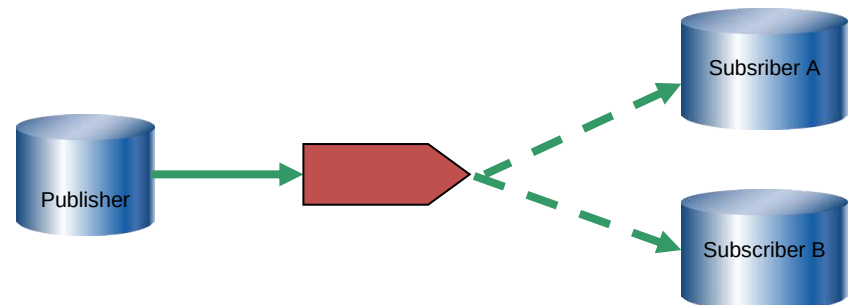
Sender – Receiver

- Zdroj zná své cíle
- Zdroj je schopen reagovat na zprávy od cíle
 - 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 opětnou synchronizaci



Datová integrace

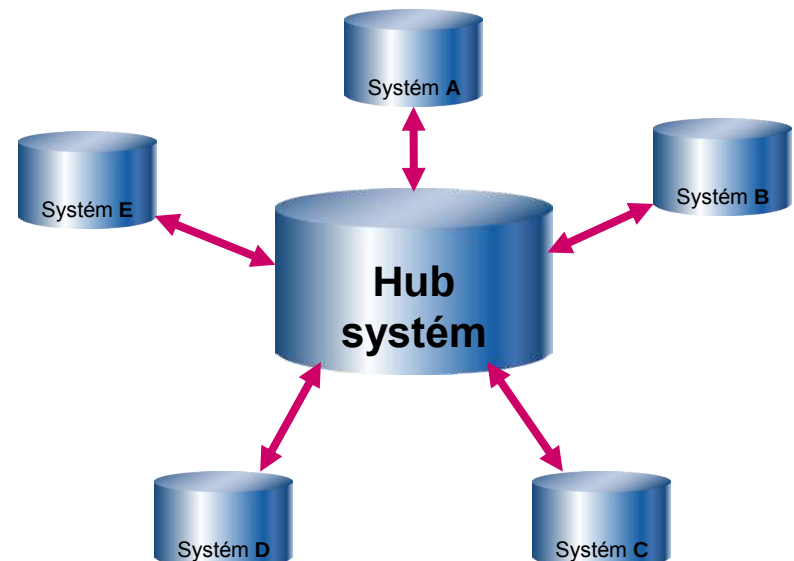
Point-to-point

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



Hub and Spoke

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



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

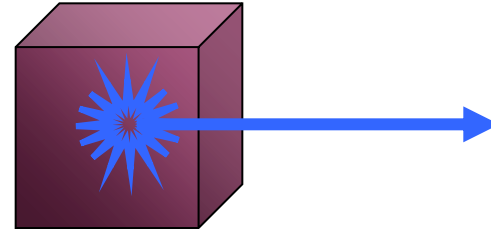
- Snadná identifikace změn
- Jednodušší způsob získávání stavu před a po změně
- Veliké množství malých zpráv - nároky na režii přenosů
- Vysoké nároky na ověření integrity na cílové straně

Kroky integrace

- Identifikace změny
- Insert záznamu
- Update záznamu
- Delete záznamu
- Problematika více systémů
- Integrace na základě času
- Integrace na základě datové kvality
- Řešení nedostupnosti dat

Identifikace změny

- Indikace změn pomocí
 - Timestamp
 - Fronta událostí
 - Technologicky (triggery)
 - Aplikačně
 - Technologicky - replikace
- Indikace rozsahu změn
 - Logická úroveň - objekt/záznam
 - Technologická úroveň - řádek, sloupec
- Data změny
 - Identifikace změny
 - Nová data
 - Původní data



Insert



- Nový záznam
- Výsledek
 - Neúplný záznam
 - Nekonzistentní záznam
 - Duplicitní záznam
- Řešení
 - Odmítnutí
 - Dočasný zápis
 - Validací proces

Update



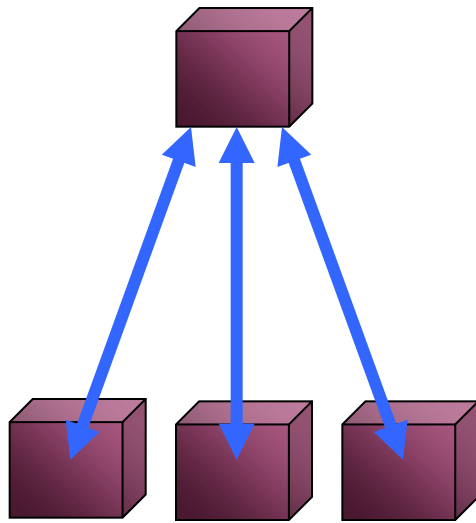
- Update záznamu
- Výsledek
 - Porušení konzistence
 - Vytvoření duplicity
 - Vytvoření neúplného záznamu
 - Nerozpoznání měněného záznamu (ztráta informace o změně)

Delete



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

Problematika více systémů



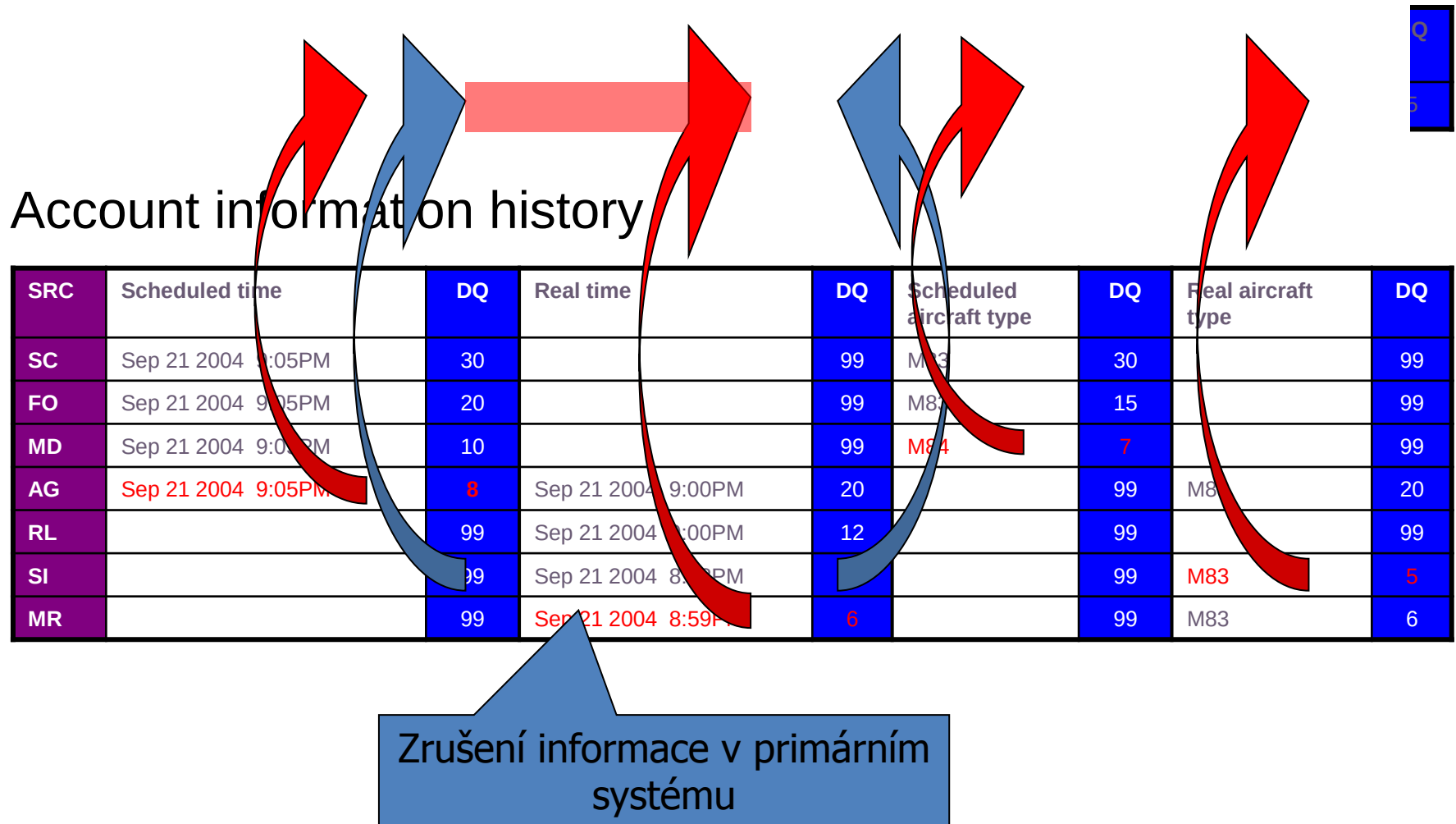
Nové typy otázek

- 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

Integrace na základě času

- Novější údaje jsou přesnější
- Definice času údaje
 - Zadání do primárního systému
 - Doba přenesení do cílového systému
 - Jak řešit paralelní zadávání dat?
- Granularita identifikace času
 - Pro celý záznam
 - Pro jednotlivé datové položky

Příklad použití datové kvality



Řešení nedostupnosti dat

- Definice

Zdroj	Kvalita dat	Null hodnota
Datawarehouse	70	Ne
System	90	Ne
Druhý systém	80	Ano

- Příchozí data

Zdroj	Jméno	Výsledek
Datawarehouse	Pavel	Jirka
System	Jirka	
Druhý systém	Tomáš	

- Vyšší hodnota kvality dat má přednost

Řešení nedostupnosti dat

- Definice

Zdroj	Kvalita dat	Null hodnota
Datawarehouse	70	Ne
System	90	Ne
Druhý systém	80	Ano

- Příchozí data

Zdroj	Jméno	Výsledek
Datawarehouse	Pavel	Tomáš
System		
Druhý systém	Tomáš	

- Vyšší hodnota kvality dat má přednost

Řešení nedostupnosti dat

- Definice

Zdroj	Kvalita dat	Null hodnota
Datawarehouse	70	Ne
System	90	Ne
Druhý systém	80	Ano

- Příchozí data

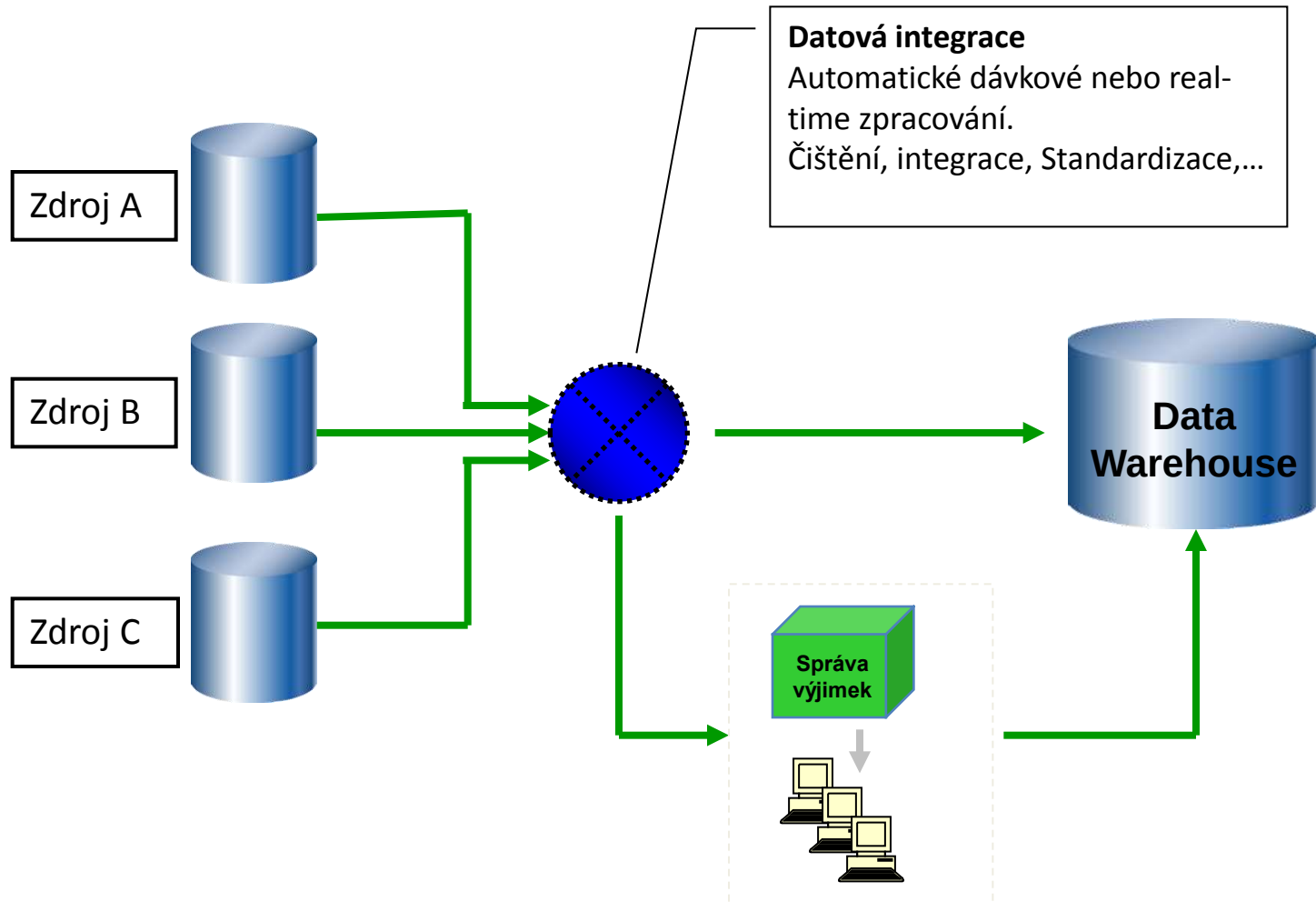
Zdroj	Jméno	Výsledek
Datawarehouse	Pavel	null
System		
Druhý systém		

- Vyšší hodnota kvality dat má přednost

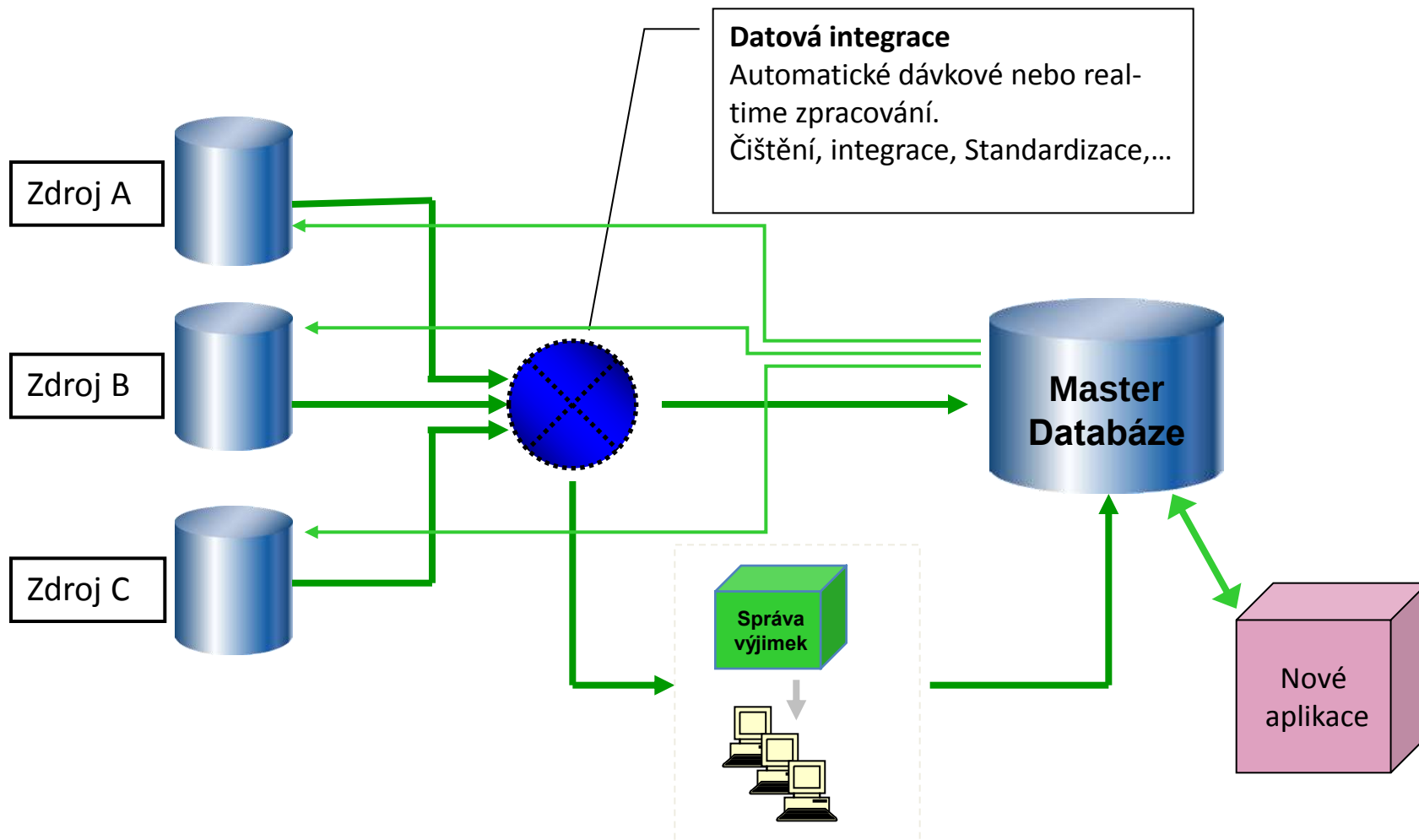
Master Data Management

- Procesy a nástroje k přesné integraci nejdůležitějších entit pro podnik.
- Správa klientů
 - PARTY
 - Role a vazby (Hausholding, ekonomicky spjaté subjekty, externí informace, scoring, ...)
- Správa produktů
 - Dodavatelé, Obchodní proces, Design, Marketing, Nacenění, Partneři, Interní systémy, Náklady, Reporting, Konsolidace produktů
- Správa centrálních číselníků
 - Historizace, plánování, různé verze pravdy, propagace do systémů
- Více architektonických přístupů
 - Master Reference Data
 - Master Systém of Records
 - Master Registry
 - Synchronizace

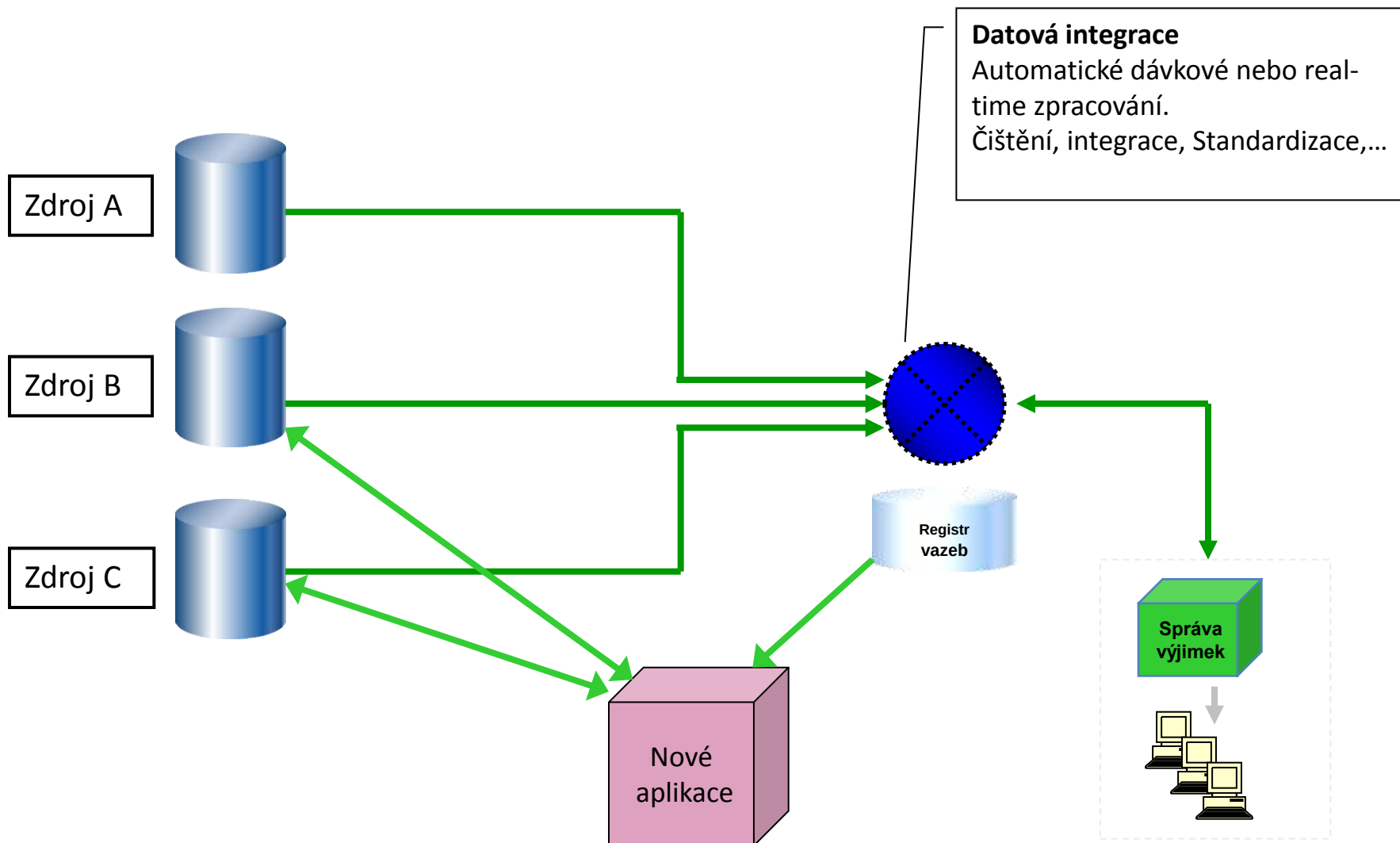
Master Reference Data



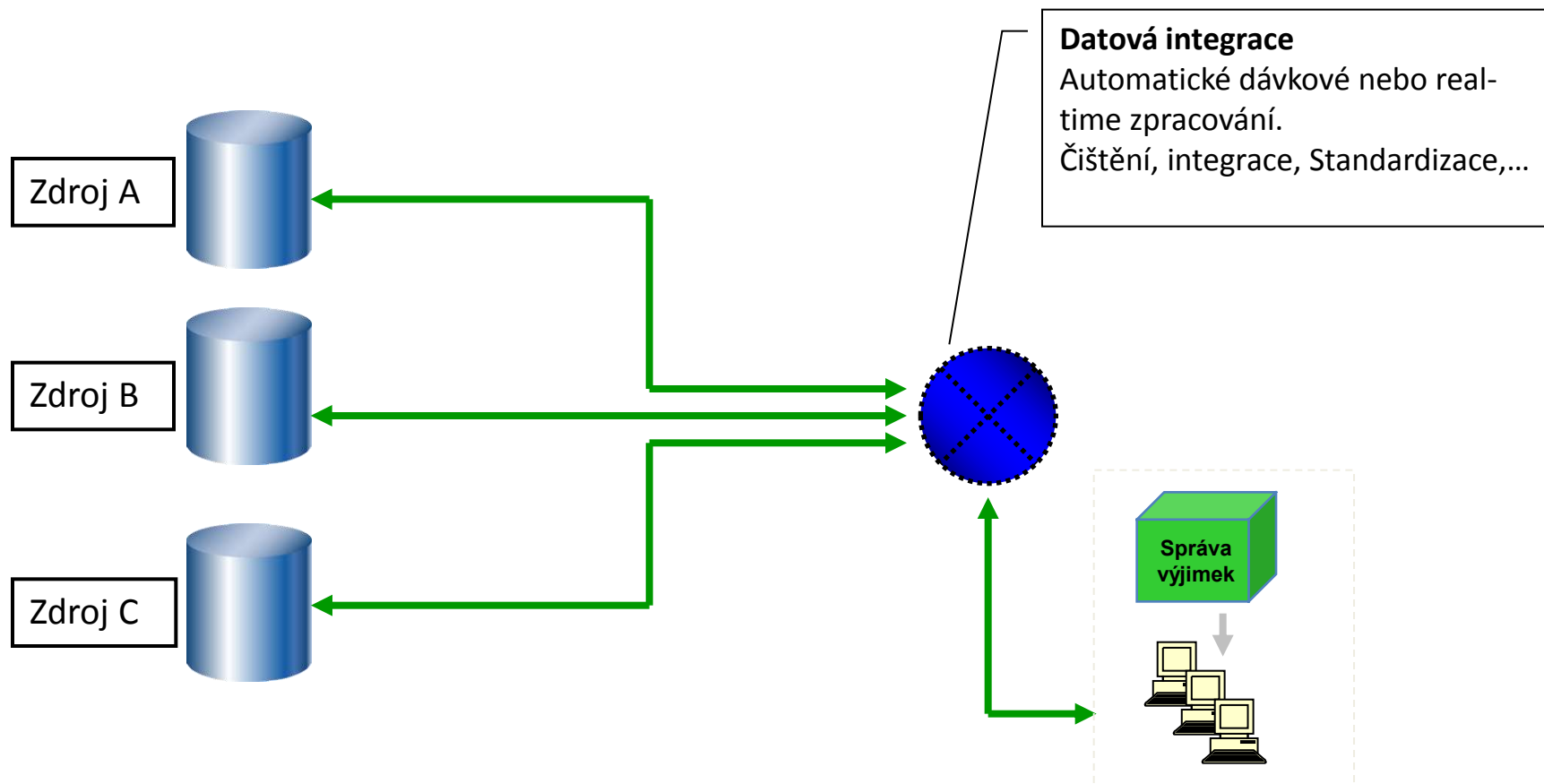
Master System of Record



Master Registry



Synchronization



Integrace

- Integrací vzniká z jednotlivých systémů řešení nové kvality.
- Nutno uvažovat
 - 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.

Integrace – rizika projektů

- 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

Integrace – rizika projektů

- 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
- Techniky indikace dat
- 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 jsou hlavní rizika integračních projektů

```

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

```

Diskuse

- Otázky
- Poznámky
- Komentáře
- Připomínky

