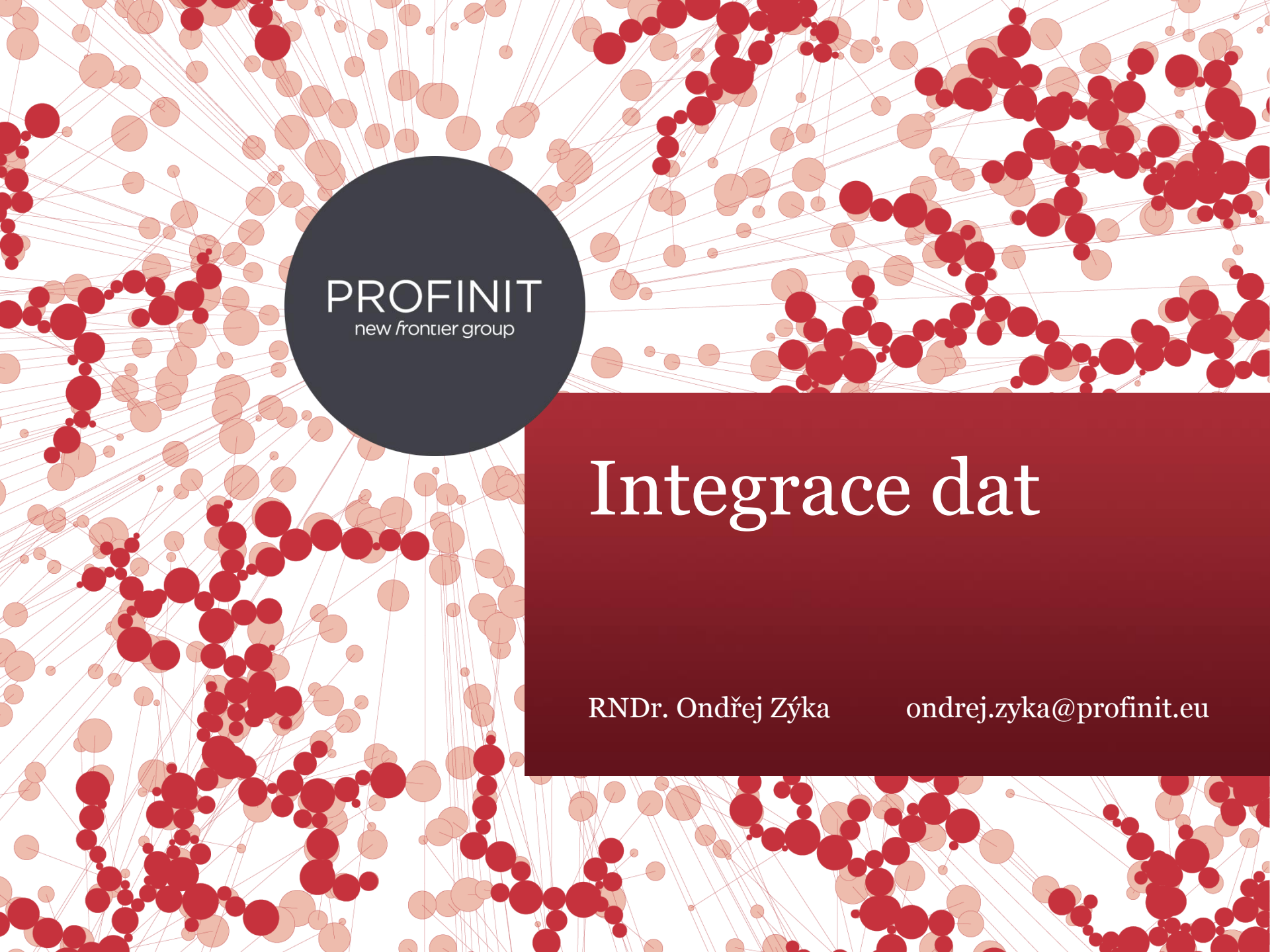



# Integrace dat

RNDr. Ondřej Zýka ondrej.zyka@profinit.eu

# Obsah

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

PROFINIT
new frontier group

## 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

PROFINIT
new frontier group

## 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

PROFINIT
new frontier group

## 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)
- Pouze malý počet dotazů
- Příklady
  - Dohled a provoz

# Datová integrace

PROFINIT
new frontier group

## 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
- Asynchronní integrace

PROFINIT
new frontier group

## Federation

- **Systém umožňuje (vynucuje) aby požadavky vznikaly jeho prostřednictvím a rozprostírá je do jednotlivých systémů.**
- **Příklady**
  - MDM aplikace
  - ESB

## Mediation

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

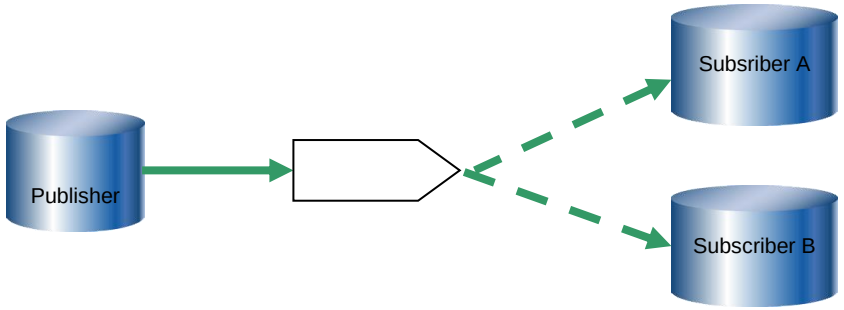
PROFINIT
new frontier group

## 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

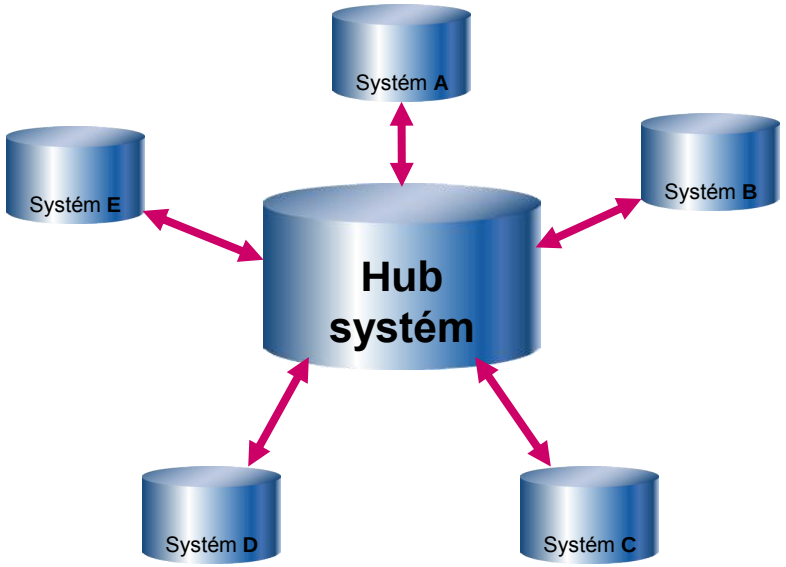


## Point-to-point

- Přímá komunikace mezi systémy navzájem
- Každý systém mnoho partnerů


## Hub and Spoke

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



# Granularita integrace

PROFINIT
new frontier group

## 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

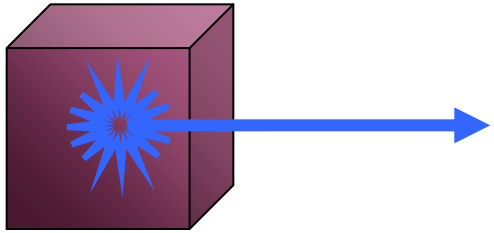
- Snadná identifikace změn
- Jednodušší způsob získávání stavu před a po změně
- Veliké množství malých zpráv - nároky na režii přenosů
- Vysoké nároky na ověření integrity na cílové straně

# 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
  - Timestamp
  - Fronta událostí
    - Technologicky (triggery)
    - Aplikačně
- Indikace rozsahu změn
  - Objekt/záznam
  - Položka/atribut, sloupec
- Data
  - Identifikace změny
  - Nová data
  - Nová i původní data



# Insert

- Nový záznam

- Výsledek
  - Neúplný záznam
  - Nekonzistentní záznam
  - Duplicitní záznam
- Řešení
  - Odmítnutí
  - Dočasný zápis
  - Validační proces

# Update

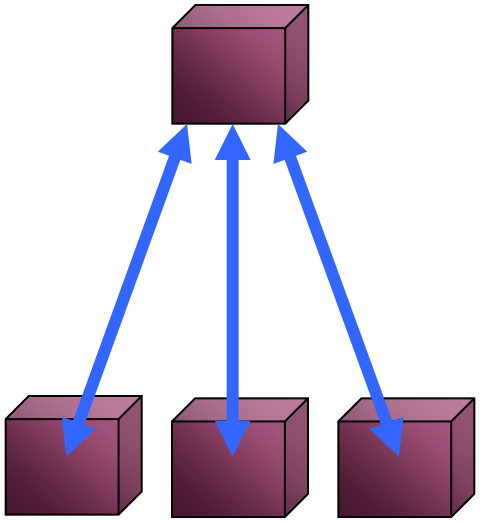
- Update záznamu

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

# Delete

- Delete záznamu
- Mnoho typů zrušení 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

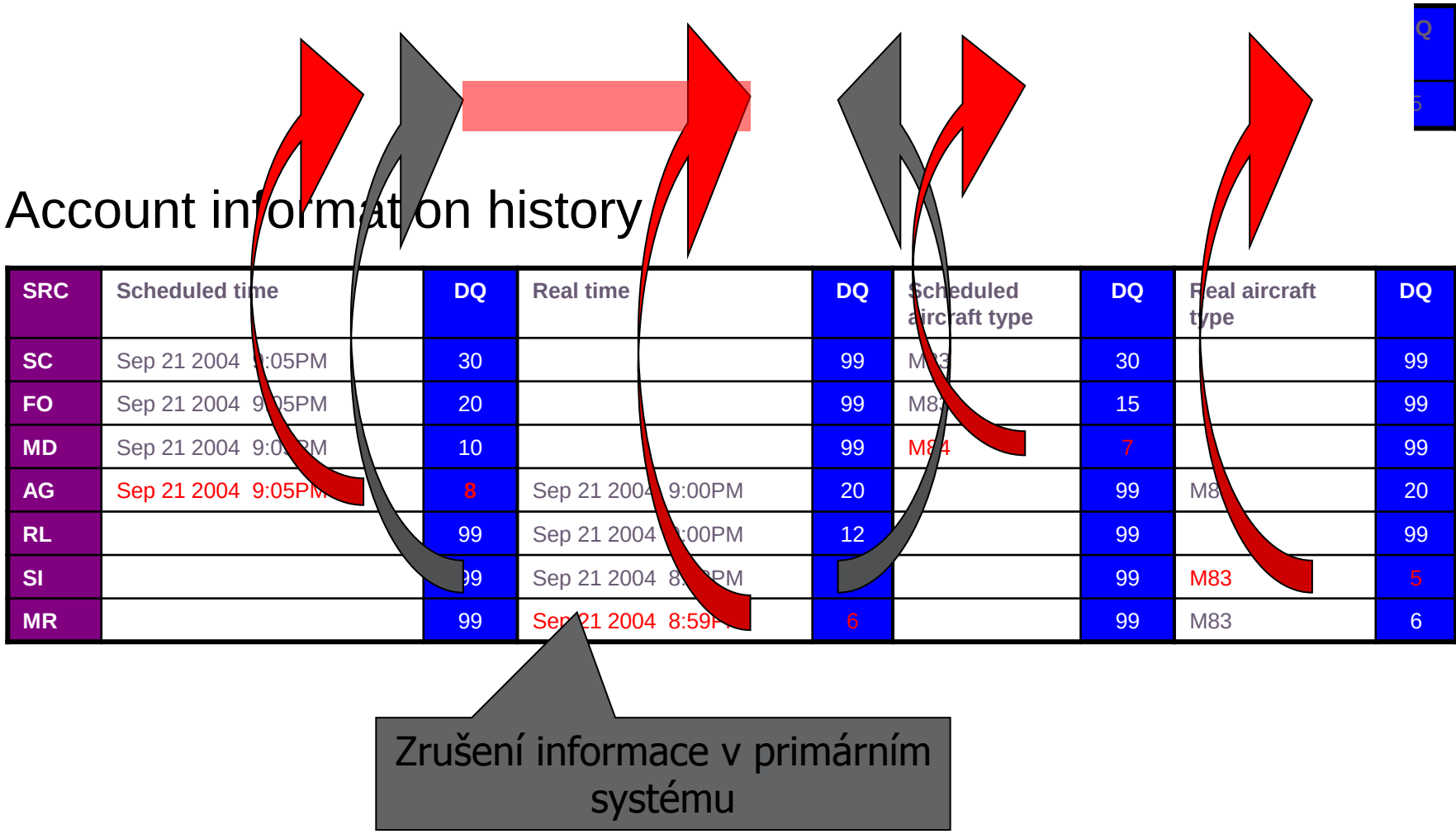
- Který systém má pravdu
- Proč má pravdu
- Jaké jiné hodnoty jsou v některém systému zadány
- Jaké hodnoty byly v kterém systému v minulosti
- Na základě jakých příčin se měnily dat v jednotlivých systémech

# 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

PROFINIT
new frontier group



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

# Řešení nedostupnosti dat

- Definice

| Zdroj | Kvalita dat | Null hodnota |
|---|---|---|
| Datawarehouse | 70 | Ne |
| Systém | 90 | Ne |
| Druhý systém | 80 | Ano |

- Příchozí data

| Zdroj | Jméno | Výsledek |
|---|---|---|
| Datawarehouse | Pavel | ? |
| Systém | Jirka | |
| Druhý systém | Tomáš | |

- Vyšší hodnota kvality dat má přednost

# Řešení nedostupnosti dat

PROFINIT
new frontier group

- Definice

| Zdroj | Kvalita dat | Null hodnota |
|---|---|---|
| Datawarehouse | 70 | Ne |
| Systém | 90 | Ne |
| Druhý systém | 80 | Ano |

- Příchozí data

| Zdroj | Jméno | Výsledek |
|---|---|---|
| Datawarehouse | Pavel | Jirka |
| Systém | Jirka | |
| Druhý systém | Tomáš | |

- Vyšší hodnota kvality dat má přednost

# Řešení nedostupnosti dat

PROFINIT
new frontier group

- Definice

| Zdroj | Kvalita dat | Null hodnota |
|---|---|---|
| Datawarehouse | 70 | Ne |
| Systém | 90 | Ne |
| Druhý systém | 80 | Ano |

- Příchozí data

| Zdroj | Jméno | Výsledek |
|---|---|---|
| Datawarehouse | Pavel | Tomáš |
| Systém | | |
| Druhý systém | Tomáš | |

- Vyšší hodnota kvality dat má přednost

# Použití Null hodnot

PROFINIT
new frontier group

- Definice

| Zdroj | Kvalita dat | Null hodnota |
|---|---|---|
| Datawarehouse | 70 | Ne |
| Systém | 90 | Ne |
| Druhý systém | 80 | Ano |

- Příchozí data

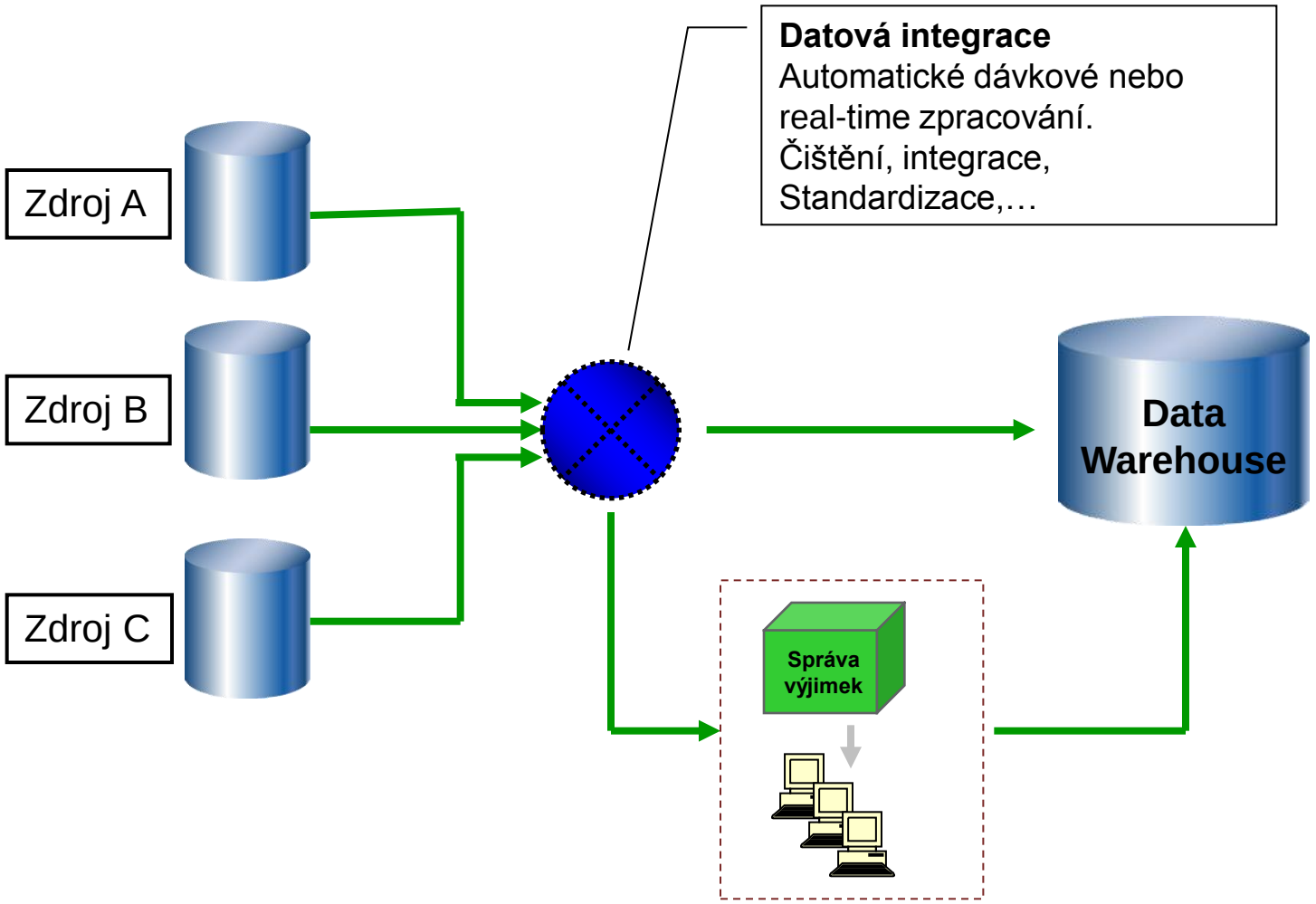
| Zdroj | Jméno | Výsledek |
|---|---|---|
| Datawarehouse | Pavel | |
| Systém | | |
| Druhý systém | | |

# Master Data Management

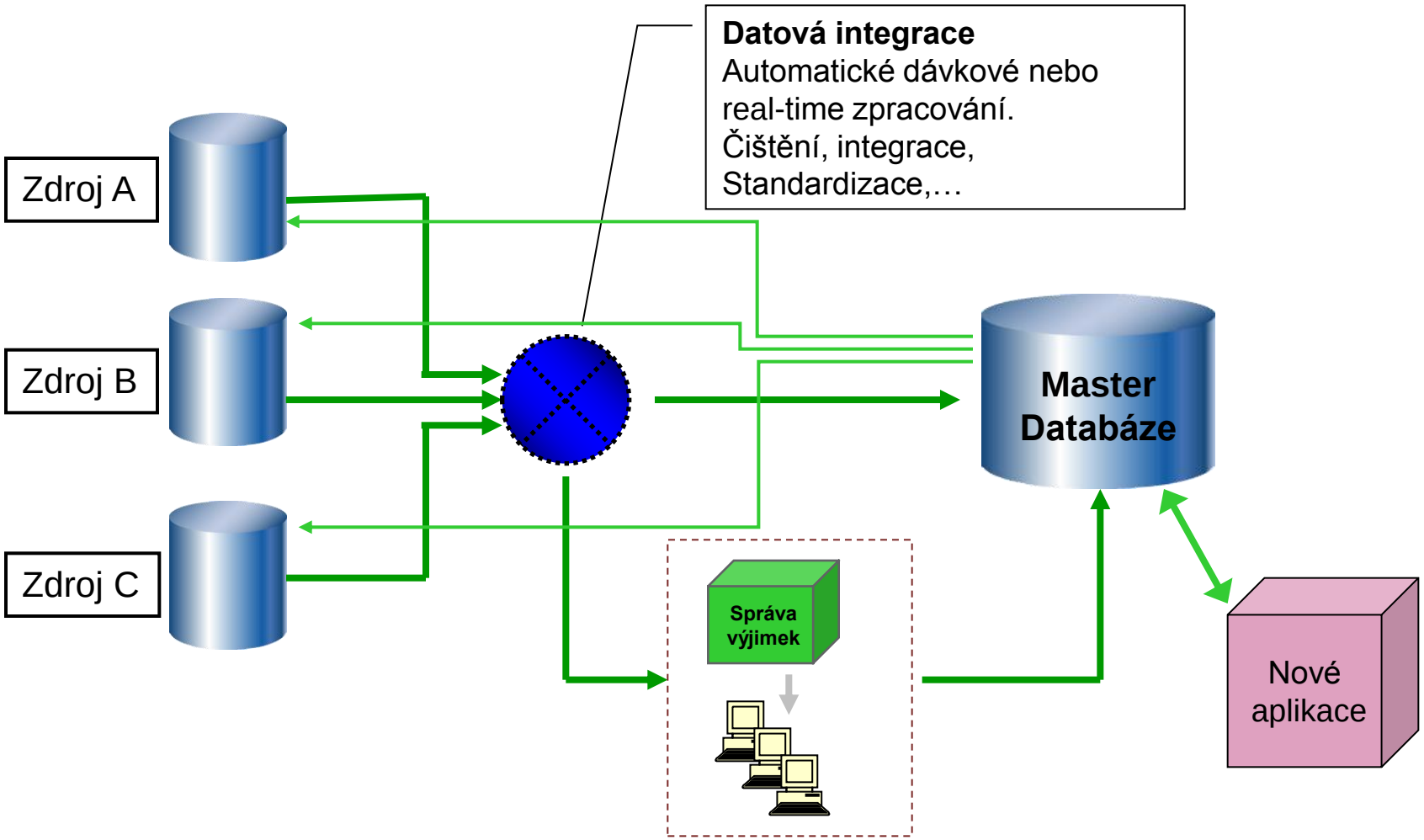
PROFINIT
new frontier group

- 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
- Master Reference Data
- Master Systém of Records
- Master Registry
- Synchronizace

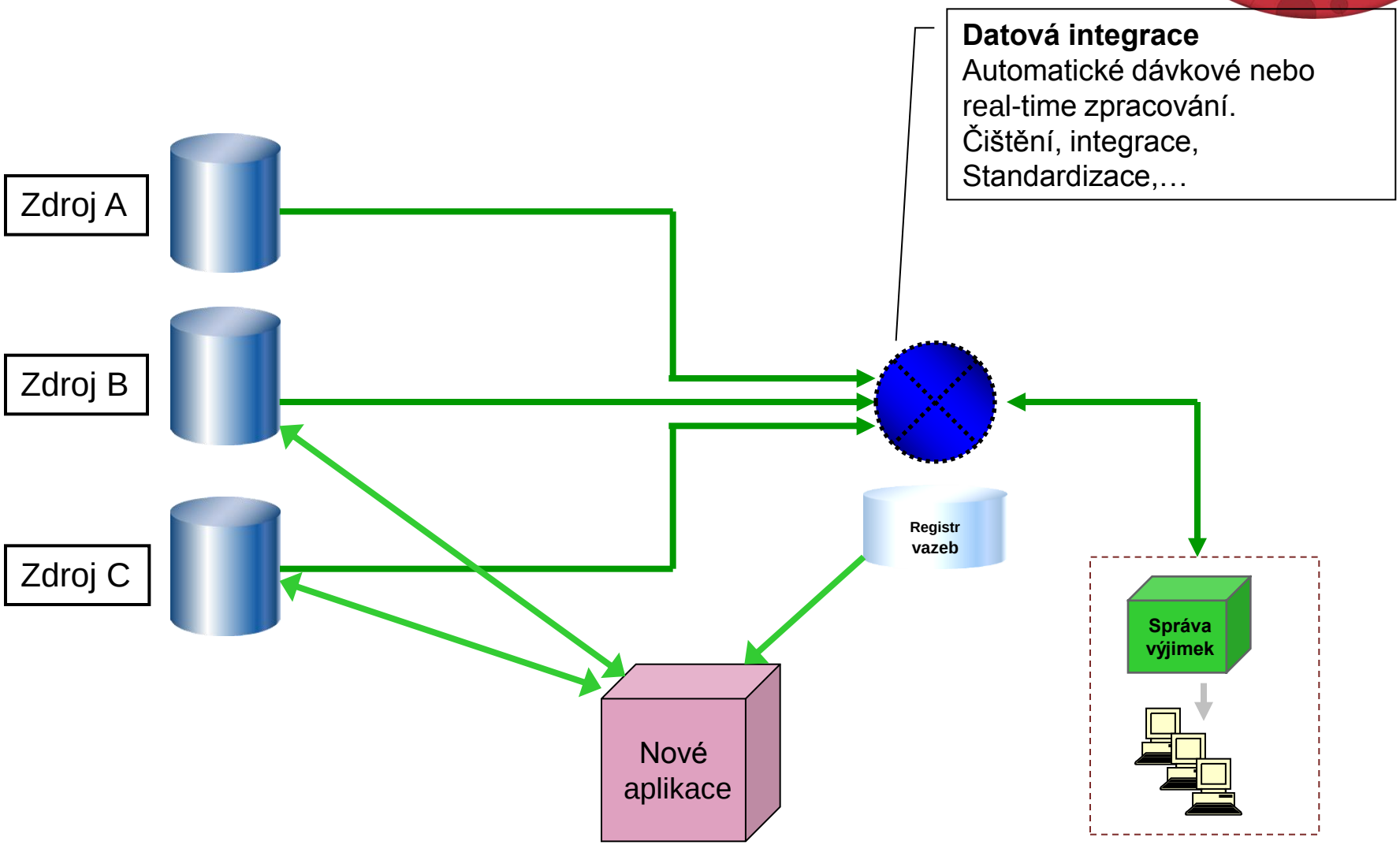
# Master Reference Data



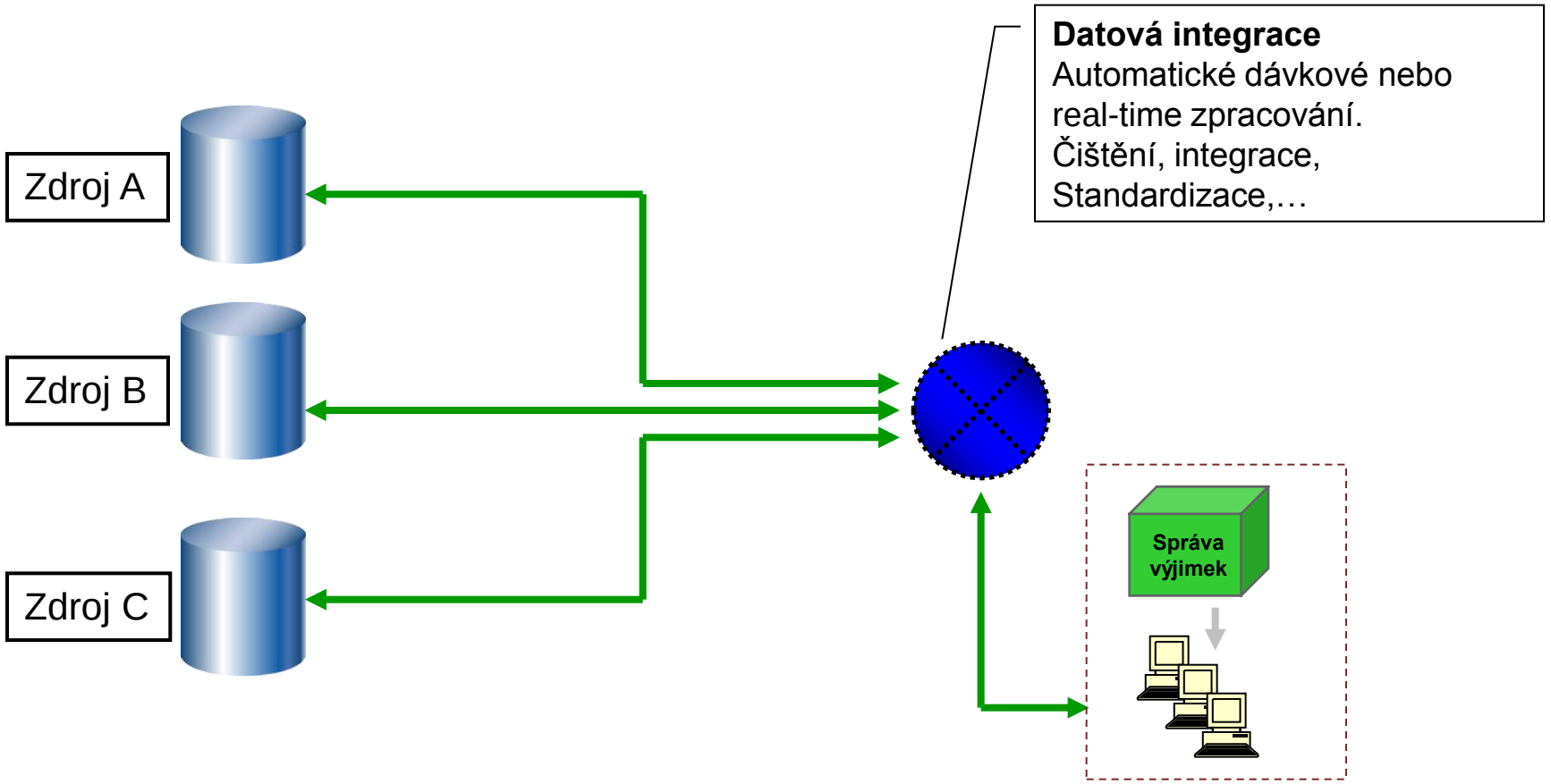
# Master System of Record



# Master Registry



# Synchronization



# Integrace

- Integrací vzniká nová kvalita.
- 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

PROFINIT
new frontier group

- 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

PROFINIT
new frontier group

- 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

PROFINIT
new frontier group

# Diskuse