

# Integrace na datové vrstvě

d-PROG s.r.o.

[Petr.Pomykacek@d-prog.cz](mailto:Petr.Pomykacek@d-prog.cz)

[Pavel.Krutina@d-prog.cz](mailto:Pavel.Krutina@d-prog.cz)

1.4.2014

# d-PROG s.r.o

- Je firma, které se již desátým rokem zabývá podporou relačního databázového systému IBM Informix.
- Její odborníci s Informixem pracují 25 let
- Oblasti působení
  - Podpora systémů 24x pořád....
  - Ladění výkoností
  - Migrace systémů na nový HW nové verze databázového serveru
  - Návrh systémů
  - Tvorba specializovaného SW, (ESQL/C).
  - Účast na EVP (Early Validation Program ) testování nových verzí
  - Aktivní účasti na seminářích
    - CIDUG – sdružení uživatelů informixu v Čechách.
    - IIUG International Informix Users Group

# Integrace na datové vrstvě

- Mýty a omyly při budování systémů
  - Universální aplikace pracující z jakoukoliv databází
  - Pouze aplikace kontroluje a řídí jak se data zpracovávají
  - Aplikace řeší to, co umí databázová vrstva lépe
  - Špatná a chybná data lze integrovat pomocí všemocných nástrojů typu ETL (Extract, Transform, Load)
  - Pouze centralizované systémy dokážou vybudovat odolné a robustnější systémy s konzistentními daty

# Universální aplikace

- Nelze psát pouze obecnou vrstvu pro práci s daty, ale vždy je jí nutno podřídit konkrétním vlastnostem databázového serveru (Oracle, DB2, Informix...)
  - Testování musí probíhat na reálnými objemy dat a v prostředí konkrétního serveru. Co platí pro jeden databázový server, nemusí fungovat na druhém!

# Aplikace dělá vše ...

- Univerzálnost aplikací vede k tomu, že aplikace řeší
  - Přístupová práva
  - Integritní omezení (Check, Default, Datové typy)
  - Referenční integritu (primární a cizí klíče)
- Naopak neřeší
  - Optimalizací SQL příkazů
  - Transakční logiku
  - Úrovně izolace přístupu k datům (zamykání záznamů při update, delete atd.)

# Aplikace dělá, co jí nepřísluší

- Where klausule se řeší na aplikační úrovni
  - Vyberou se všechny záznamy a pomocí IF a regulárních výrazů zúží výběr, místo MATCHES, LIKE nebo jen *col > hodnota*.
- Řada mezivrstev aplikačních serverů zcela zatemní, jak SQL zpracování vlastně probíhá
  - Pro každé pole formuláře se provádí SELECT místo jednoho výběru mnoha řádků najednou.

# Návrh a tvorba systému

- Jen dobře navržený systém umožní pozdější integraci s okolím.
- K dobrému návrhu databáze také patří
  - Diskové rozložení dat a indexů v databázovém systému (Databáze není jen hromada dat na kterou se ukládá nebo se z ní vybírá)
  - Fragmentace dat a indexu
  - Pohled do budoucnosti. Kolik bude dat za rok, za dva, ale i za pět let.
  - Je objem dat
    - uchopitelný,
    - přenositelný v budoucnu,
    - Zálohovatelný a hlavně obnovitelný v reálném čase (V případě katastrofy bude obnova tak dlouhá, že data ztratí faktický význam)

# IBM Informix

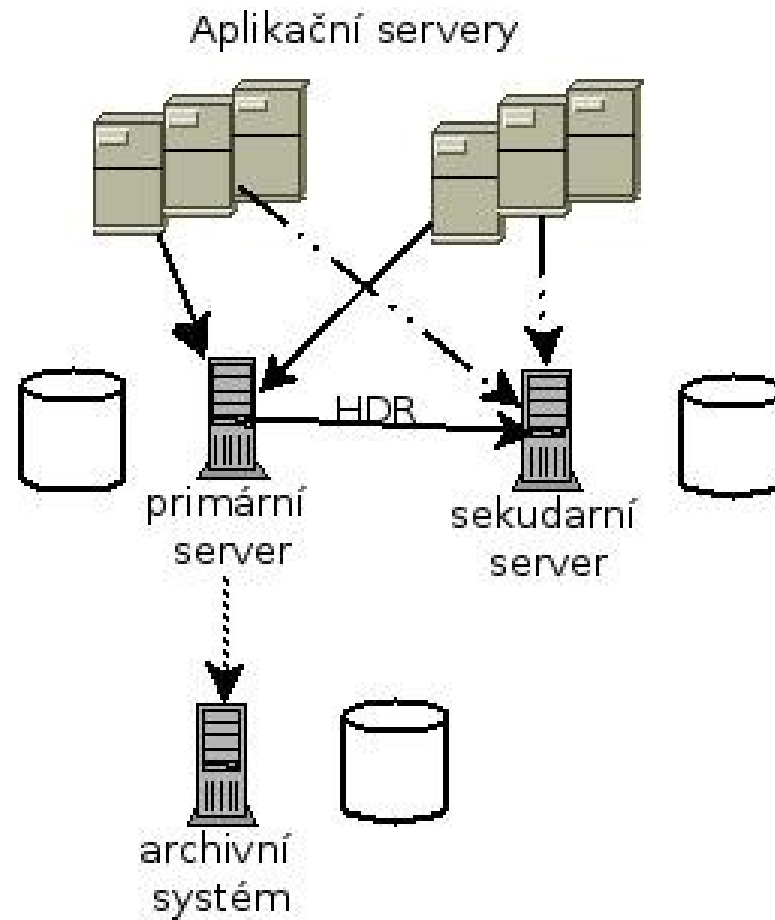
- Je relační databázový systém vyvinutý v 80 letech minulého století
- V České a Slovenské republice byl nasazován zejména 90 letech minulého století a to jak ve státní správě a samosprávně, tak i komerční sféře.
- Jeho rozmach paradoxně přibrzdila jeho akvizice firmou IBM v roce 2001
  - IBM jasně nedefinovalo kam bude Informix směřovat a tak řada zákazníků migrovala ne k DB2 jak IBM doufala, ale odešla k ORACLE či jiným systémům.
  - Síla nasazení a zejména jednoduchost a technologická dokonalost v řadě případů udržela řadu uživatelů až do současnosti a ani další “akce” IBM nedokázala zlomit jejich setrvání u této databázové platformy



# Příklady řešení INFORMIXových systémů

- Globální aplikace pro sledování přepravních zásilek
- Řešení registru živnostenského podnikání
- Řešení IS v přední pojišťovně

# Aplikace sledování zásilek

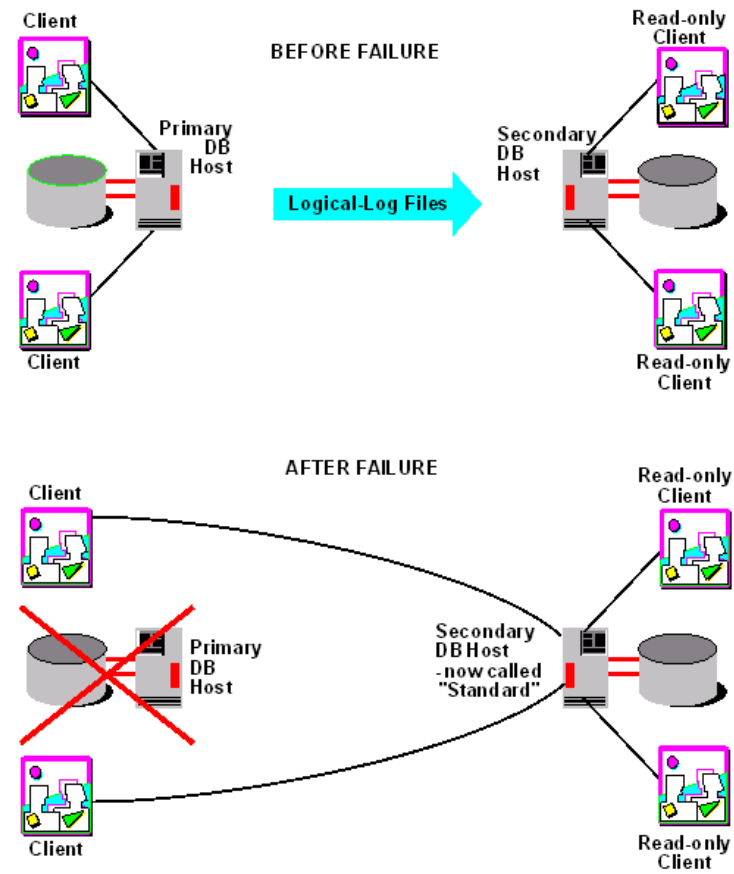


# Aplikace sledování zásilek

- Velký objem malých transakcí směřujících z desítek aplikačních serverů.
  - Využití technologie High-Availability Data Replication (HDR)
    - Data logického logu se přenáší na další server a tím dochází k tvorbě identických kopií dat
  - Samostatné servery
    - Jeden primární
    - Jeden sekundární
    - Jeden a více RSS serverů – zde není využito
  - Connection Manager pro směrování dat k serverům
    - Logické rozdělení RW a RO operací dělí zátěž mezi více serverů

# HDR

INFORMIX: Overview of High-Availability Data Replication (HDR)



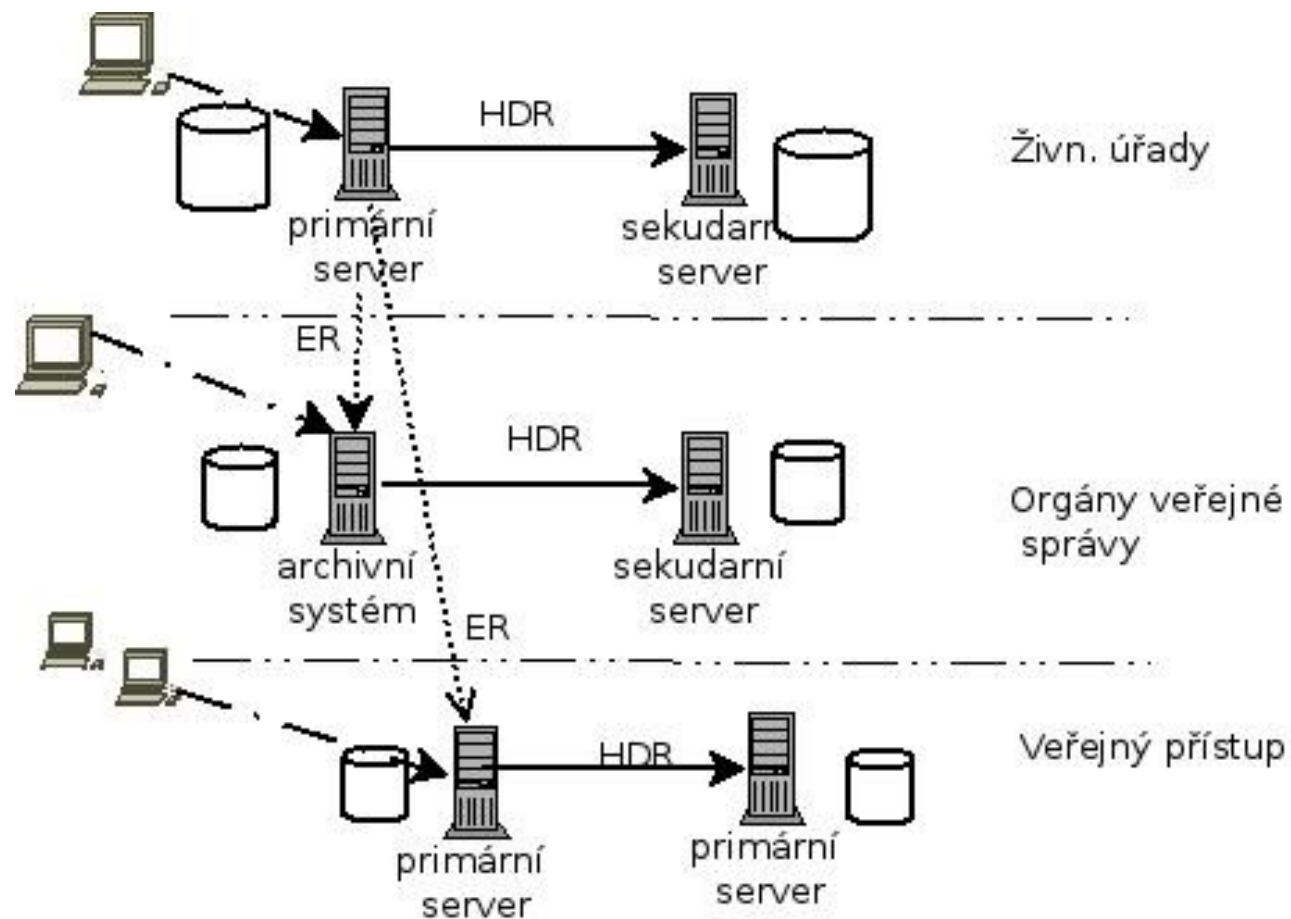
# Struktura databáze

- Provozní databáze 600GB
- Archivní databáze 1,3 TB po kompresi 300GB
- Databáze má 90 tabulek
- 7 hlavních tabulek obsahuje 95% dat
- Maximální počet řádků v tabulce je 1,3 miliardy
  - Neuchopitelné pro správu

# Registr živnostenského podnikání

- Systém využívá technologie HDR spolu s ER (Enterprise Replikace)
- ER slouží pro omezení dat pro jednotlivé úrovně přístupu
  - Živnostenský úřad
  - Státní správa
  - Veřejnost

# Registr živnostenského podnikání



# RŽP objemy dat

- Živnostenské úřady 300GB
  - Už tento objem bude v budoucnu nutné rozdělit na provozní a archivní data
  - Archivace dokumentů
- Státní správa 150GB
- Veřejnost 100GB



# Struktura databáze

- Celkem 450 tabulek
- Maximální počet řádků v tabulce je 36 miliónů ostatní jsou menší. Není problém při manipulaci s daty
- Zóny mají stejnou strukturu
  - Liší se počtem řádků
  - Prázdné tabulky

# Pojišťovna

- Jedna databáze roste od roku 1995, dnes má více než 3TB
- Nárůst dat se zrychluje, přírůstek je 32GB za měsíc
- Systém neřeší archivaci dat
- Pro statistické a analytické účely se vytváří identické kopie dat na úrovni diskového pole
  - Každá kopie má zase 3TB
  - Začalo se s jednou, nyní jich je 6

# Pojišťovna

- Propojení z jinými systémy je problém!!
- SAP
  - Od 17:00 do 24:00 výběr dat z primárního systému do SAPu
  - Od 24:00 do 07:00 příjem dat ze SAPu
- Provozní problémy
  - Přesouvací mezidatabáze má již 1TB
  - Časový problém operace se nestíhají
- Řešení se odaluje, snaha řešit cestou výkonného HW (hrubou silou)

# Pojišťovna

- Co s tím do budoucna
  - Problém se musí řešit!!!
- Možnosti řešení
  - Oddělením provozních a archívních dat
  - Využití ER pro přesun dat do analytických a statistických systémů
  - Budou nutné úpravy aplikací , což je důvod s otálením řešení

# Struktura databáze

- Obsahuje cca 400 tabulek
- Maximální počet řádků v tabulce se blíží k 2 miliardám
  - Opět složité administrativní operace např index, alter table apod.

Děkujeme za pozornost

?

Dotazy