

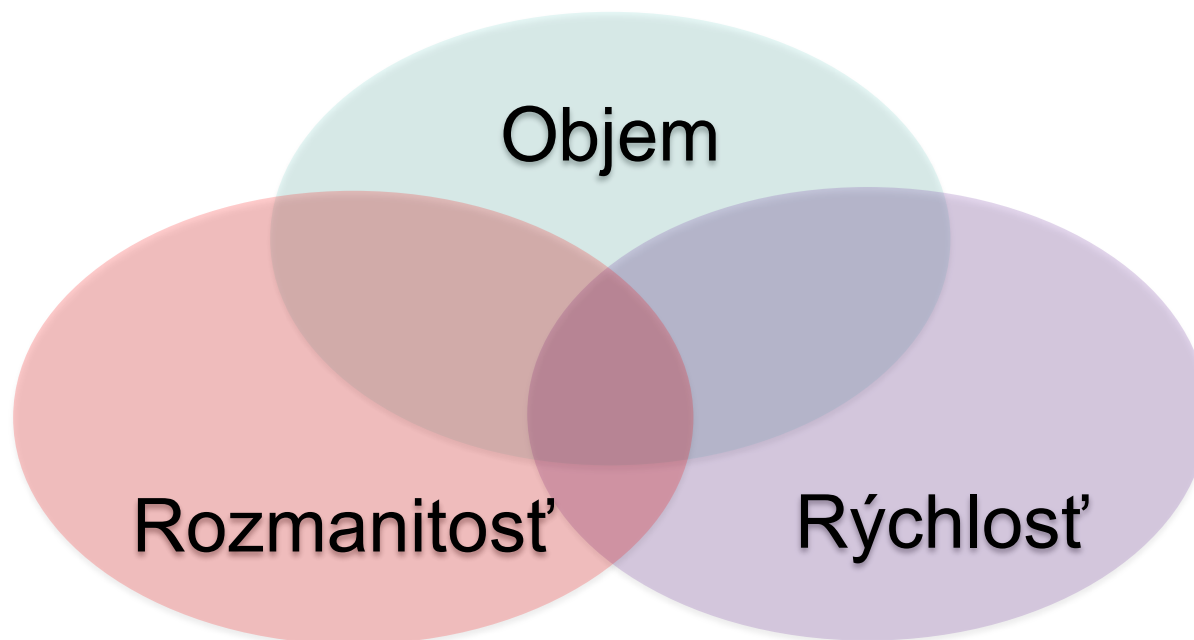
Problém Big Data a ako ho riešiť pomocou NoSQL

Ján Zázrivec
Softec

Dáta dnešného sveta

- Oblasti kde sa spracováva veľké množstvo dát:
 - Internet – Web vyhľadávače, Sociálne siete
 - Veda – Large Hadron Collider, Astronómia, Ľudský génom, Výskum klimatických zmien
 - Obchod, Finančný sektor
- Čo majú spoločné?

Big data – Veľké dáta



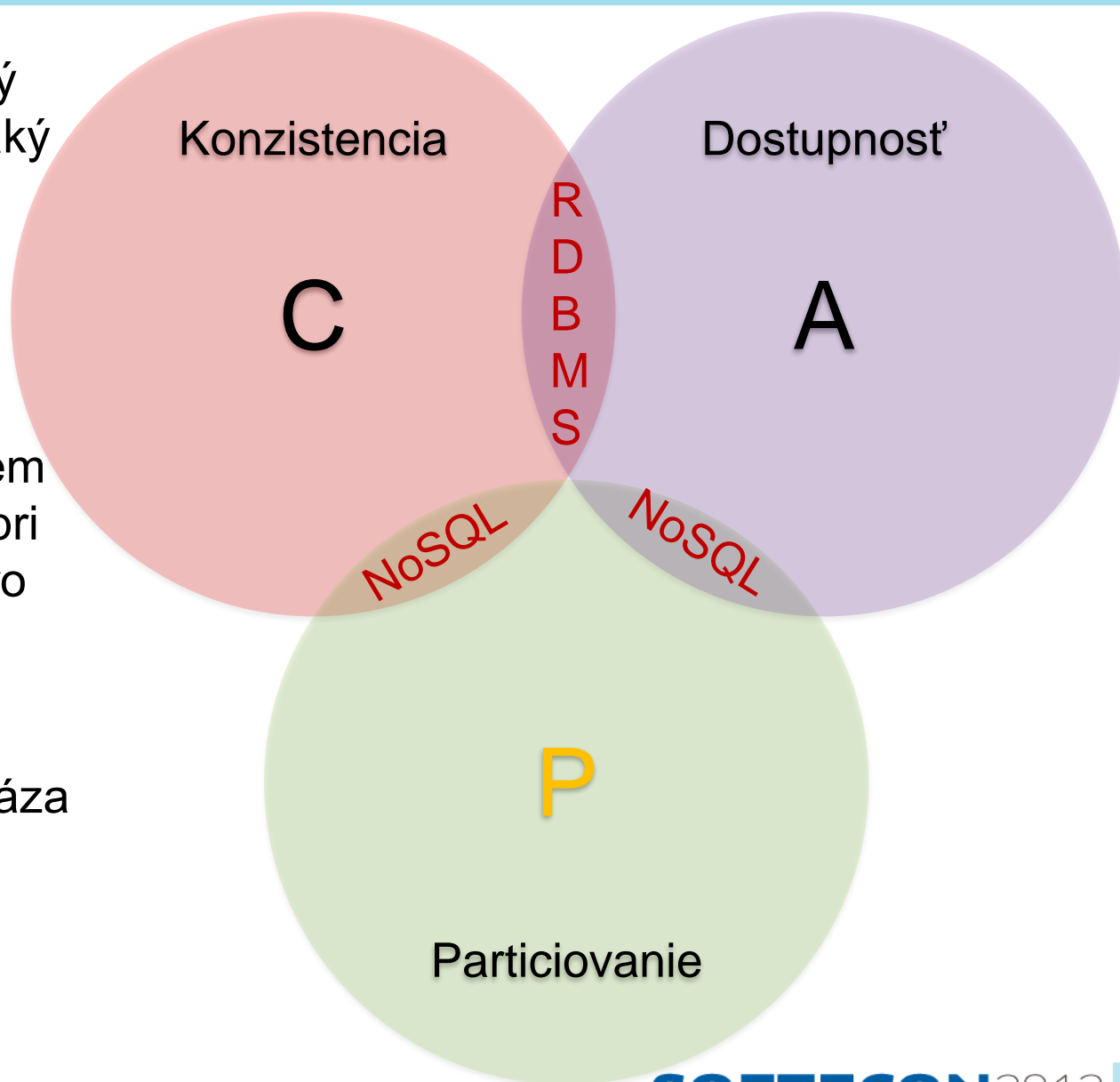
- Dáta: ukladať, spracovávať a prezentovať
- Prečo je použitie relačných databáz neefektívne?

Problémy relačných databáz

- Objem/Rýchlosť
 - Problematické horizontálne škálovanie
 - Dynamické tabuľky pri dotazoch
- Rozmanitosť
 - Join v dotazoch
 - Zmena databázovej schémy vs. dostupnosť
- Náklady na škálovateľnosť sú exponenciálne

Brewerová CAP teoréma distribuovanej DB

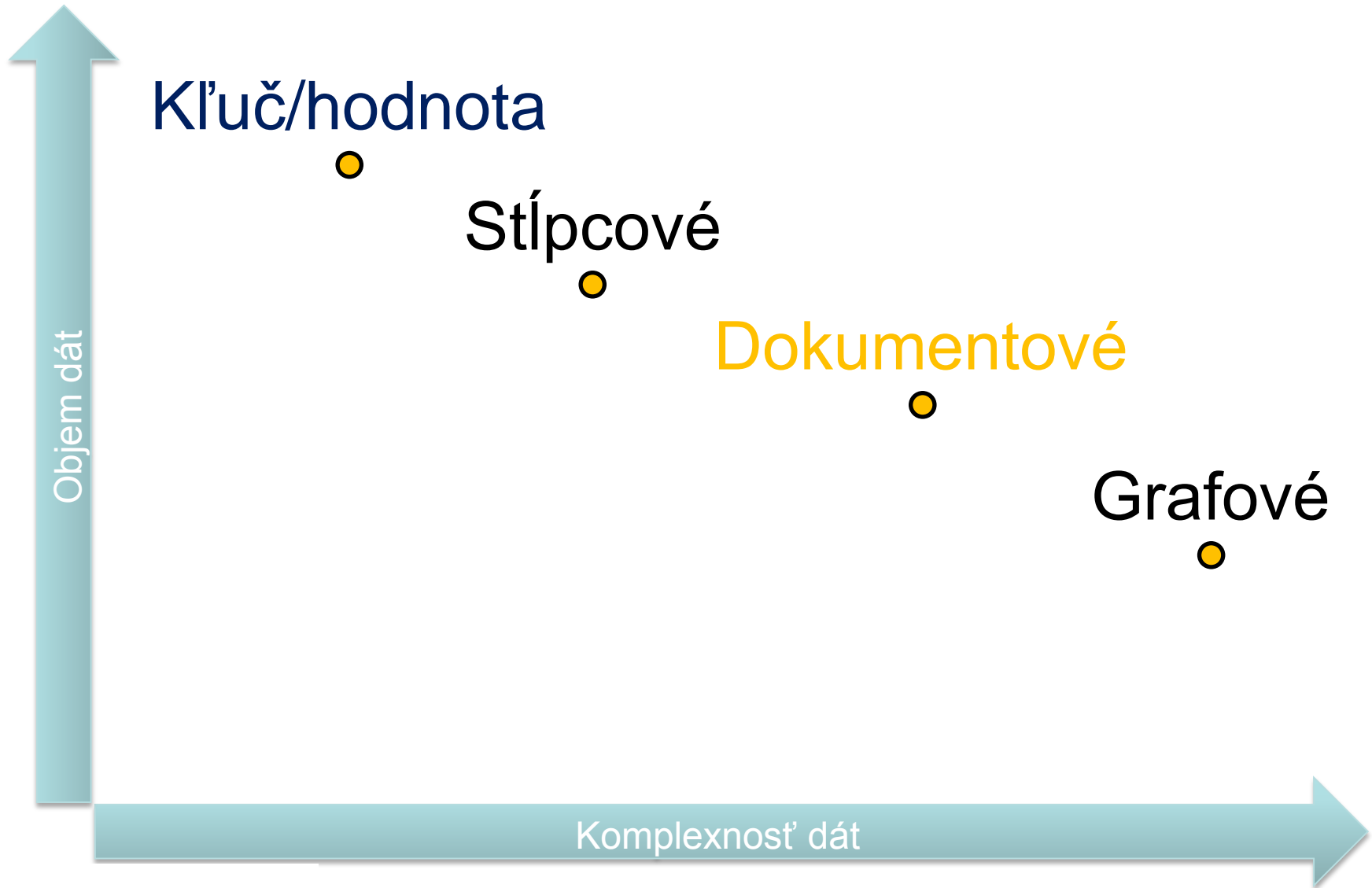
- Konzistencia – každý klient ma vždy rovnaký pohľad na dáta
- Dostupnosť – každý klient dokáže vždy čítať aj zapisovať
- Particiovanie – systém dokáže fungovať aj pri rozpadnutí na viacero častí
- Distribuovaná databáza nedokáže splniť dokonale všetky 3 požiadavky



NoSQL

- Horizontálne škálovateľné
- Špecializácia na konkrétny problém
- Nepoužívajú relačný model
- Neexistuje spoločný dotazovací jazyk *
- Bez schémy (Schemaless)

Typy NoSQL databáz



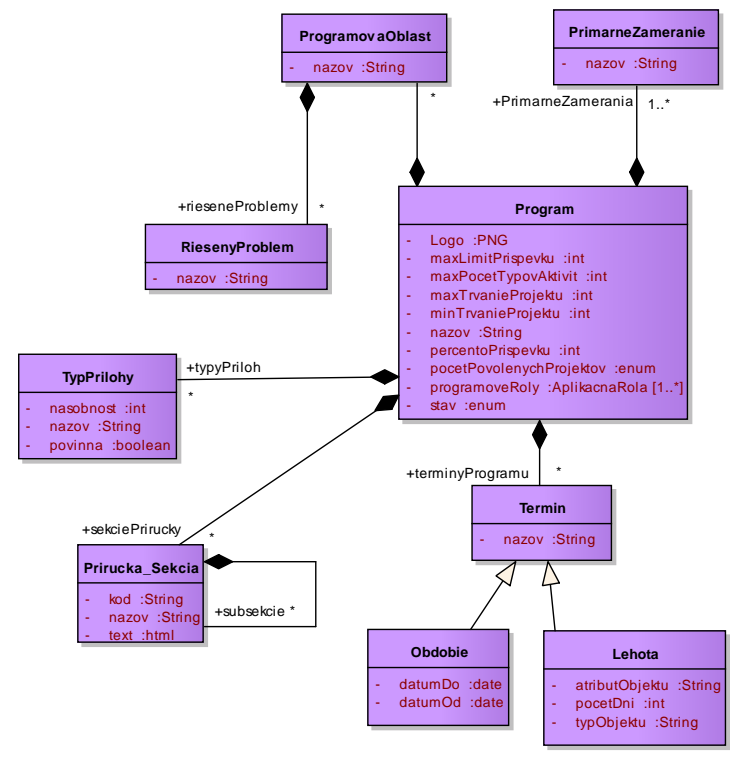
Dokumentové NoSQL

```
{ name:"Hodina deťom 14. ročník",
  primaryFocus:["bežné deti", "zdravotne znevýhodnené"],
  programArea:{
    name:"Dieťa a jeho radosti",
    solvedProblem:["vplyv médií, na dieťaťa", "chýbanie komunikácia"]
  },
  guideSections:{
    name:"Informácie o organizácii",
    abstract:"Ak ste vyplnili ...",
    text:"Ak ste vyplnili pri registrácii ...",
    subSections:{
      name:"Predkladateľ",
      abstract:"Predkladateľom projektu...",
      text:"Predkladateľom projektu...",
      subSections:{
        abstract:"test1,"
      }
    }
  }
}
```

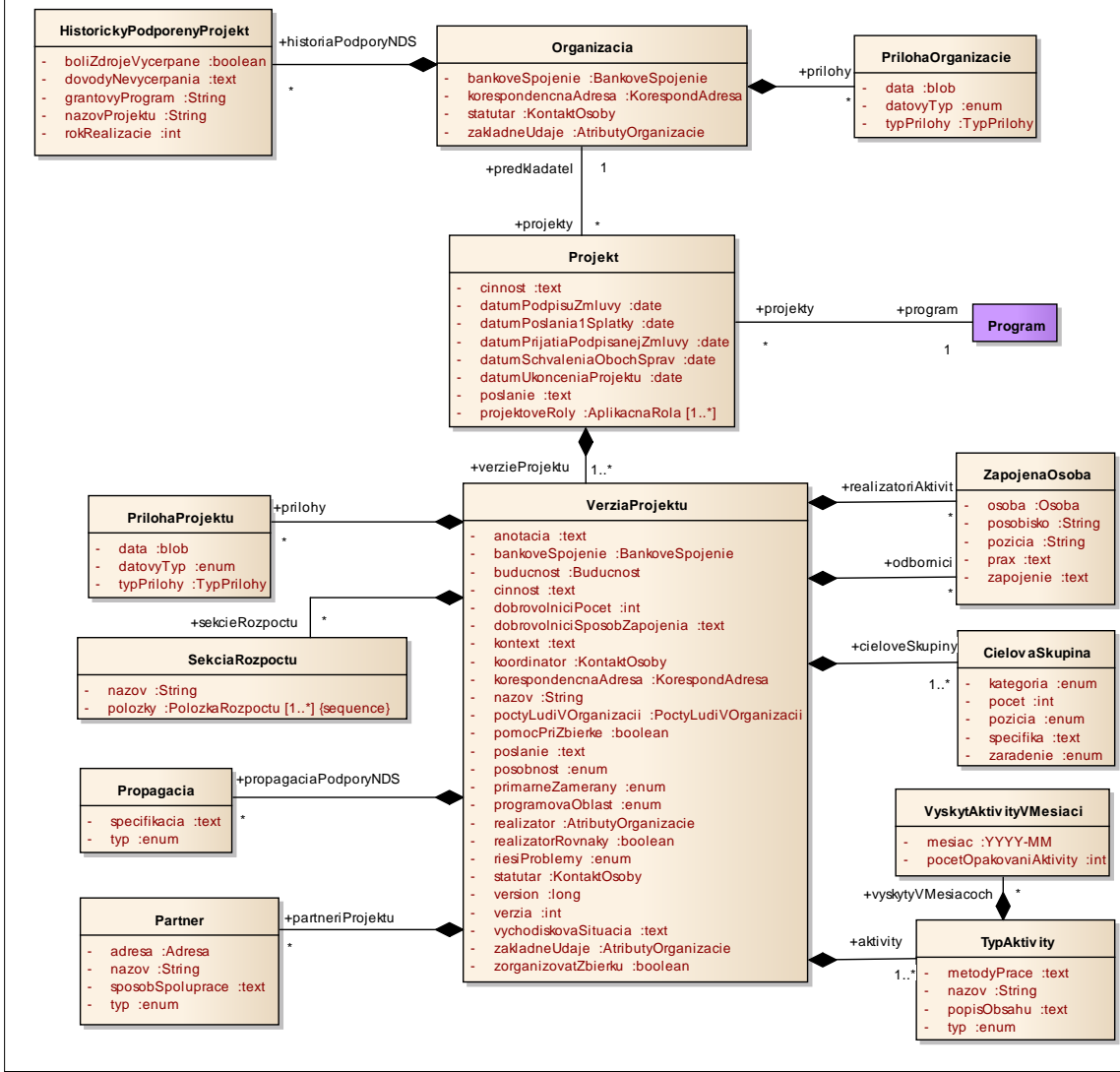
- **Vhodné použitie:**
 - Štruktúrované a polo štruktúrované dokumenty
 - Úložisko metadát, napr. produktový katalóg

Ukážka meta-modelu a modelu

class NDS_Program+datove typy

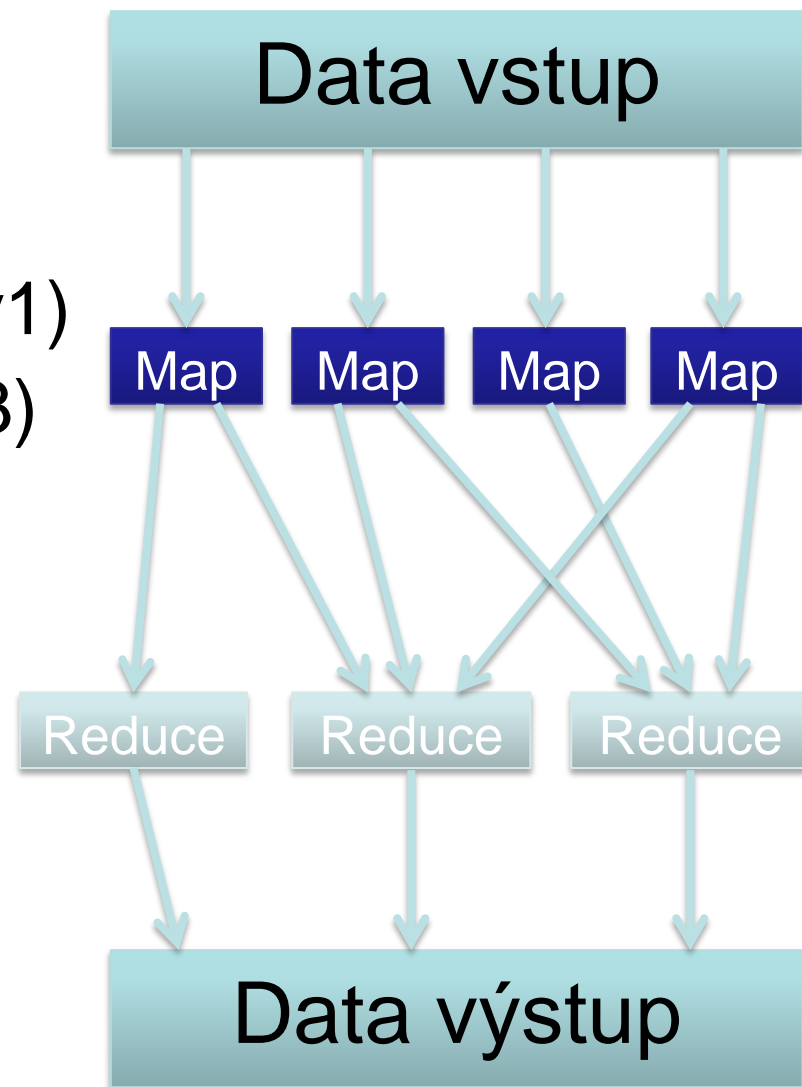


class NDS_LogickyModel



Dotazy a agregácie

- Ad Hoc dotazy (index)
- Map – reduce algoritmus
 - $\text{map}(\text{key}, \text{value}) \rightarrow \text{list}(\text{k1}, \text{v1})$
 - $\text{reduce}(\text{k1}, \text{list}(\text{v1})) \rightarrow \text{list}(\text{v3})$
- Materializovaný pohľad



Architektúra IS s NoSQL

- Zvážiť ktorú NoSQL implementáciu použiť a či vôbec
- Použitie **len** NoSQL je takmer vždy nevhodné
 - Riešením je polyglot perzistencia
- Zmena v existujúcich systémoch na NoSQL
 - Veľmi náročná a nebezpečná
- Integrácia s Business Analytics (napr. Pentaho)

Ďakujem za pozornosť