

PA036: Projekt z DB systémů

Vlastislav Dohnal

Cíle předmětu

- Praktické používání databázového systému
 - Využití rozšířených vlastností
- Vytvoření týmového projektu a jeho prezentace
 - Tým = 4 studenti
- Témata a průběh projektů: „Nové“ možnosti DB systémů
 - Představení problematiky
 - Otestování výkonnosti
 - Porovnání s jinými možnostmi
 - Max. 2 týmy na jedno téma

Průběh předmětu

1. fáze

- Sestavení týmu
- Volba tématu projektu, volba náhradního tématu projektu; určení kompetencí v týmu

2. fáze

- Bližší specifikace projektu, konzultace s vyučujícím, stanovení finálního rozsahu
- Včetně Gantt diagramu prací (co, kdy, kdo) včetně odhadů pracnosti (v člověkohodinách) s týdenní granularitou termínů

3. fáze

- Odevzdání specifikace projektu
- Vypracování posudku na projekt týmu „s náhradním tématem“

4. fáze

- Prezentace projektu ostatním, představení posudku oponenty, diskuze

5. fáze

- Vlastní realizace (doma 😊, konzultace s vyučujícím v rámci hodin semináře, týmové schůzky a týdenní reporty)

6. fáze

- Odevzdání hotového díla, včetně původního Gantt diagramu a upraveného Gantt diagramu podle skutečného průběhu (tj. plán vs. skutečnost)

7. fáze

- Finální prezentace výsledků projektu

Přesné časování a termíny sledujte v interaktivní osnově v ISu.

Osnova

- Databázové systémy a krátké povídání k NoSQL
- Průběh řešení projektu
- Témata projektů

Databázové systémy a přístup k nim

- Relační – PostgreSQL
- NoSQL – dokumentové, klíč-hodnota, sloupcové, grafové
- Přístup k DBMS
 - Aplikační frameworky (Nette, Laravel, ...)
 - REST API (DBCore)
 - JDBC/ODBC

Dokumentové NoSQL databázové systémy

- MongoDB
 - Datový model - JSON dokument
 - Dynamické schéma
 - Primární přístup k dokumentu pomocí ID
 - Sekundární přístup – index nad vybraným atributem dokumentu
- CouchDB
 - Analogický k MongoDB

Klíč-hodnota NoSQL databázové systémy

- Redis
 - Datový model – dvojice klíč-hodnota
 - Primitivní, vysoce výkonné operace nad seznamy, množinami a asoc. poli
 - Často jako in-memory cache – lze nastavit expiraci záznamů
 - Zpracování GEO souřadnic – včetně rozsahových dotazů
- Riak
 - Analogický k Redis, navíc zvládá JSON dokumenty jako hodnoty, umí sekundární indexy
 - Indexy se musí udržovat ručně

Sloupcové / grafové NoSQL DB systémy

- Cassandra
 - Relační datový model
 - Variabilní schéma – skupiny sloupců
 - CQL jazyk podobný SQL
- Neo4j
 - Datový model – orientovaný multigraf

Průběh řešení projektu - První prezentace

- Zvolení tématu a specifikace cílů projektu
 - přihlášení se v rozpisu témat v ISu na preferované a náhradní téma
- Vytvoření projektu v GitLab.fi.muni.cz (podle instrukcí)
- Podrobný plán a cíle projektu, způsob práce
 - definice kompetencí členů týmu, časový harmonogram (Gantt)
 - metodika testování a vyhodnocení (jaká data, co/jak měřit)

Průběh řešení projektu - Během semestru

- Doplnování a aktualizace Gantt diagramu o čas strávený na projektu (sdílený Gitlab repozitář)
- stačí jedna věta a číslo -> identifikace nepracujících a jejich hodnocení „X“
- konzultace s vyučujícím

Průběh řešení projektu - Druhá prezentace

- Výsledky projektu do závěrečné zprávy
 - zopakování (stručně) zadání a cílů projektu
 - přístup k řešení, např. popis technologií, dat
 - výsledky experimentů
 - zhodnocení časového plánu

Poznámka k realizaci projektu

- Experimentování
 - Každý test musí být proveden opakovaně
 - se zachycením požadovaných metrik
 - Prezentovat průměrnou hodnotu metriky, její odchylku
- Data
 - lze použít pgbench, který generuje data odvozená z TPC-B benchmarku
 - jiný TPC benchmark (viz <http://www.tpc.org/information/benchmarks.asp>)
- HW - vlastní, [PC v učebně](#), [Stratus@FI](#)
- Časové nároky
 - dotace 2 kredity -> cca 52 člověkohodin na semestr a studenta
 - započítat povinnou přítomnost na semináři (4x prezentace, plus úvodní hodina)
 - plánovat v Gantt diagramu

Témata projektů pro 2022

- Cizí datové zdroje
- Řízení přístupu k záznamům relace
- Rozšířené statistiky a výkon DB
- Monitorování databázového systému
- Aplikační framework a kešování DB
- Notifikace DB serveru a aplikační framework
- High Availability in PostgreSQL
- Master-master replikace
- Optimalizace úložiště DB
- Horizontální dělení relací
- Výkonnost zpracování JSON dokumentů
- Výkonnost zpracování XML dokumentů
- Zpracování časových řad
- Full-text vyhledávání
- Archivace dat a obnovení
- Analytické dotazy a materializované pohledy
- Prohledávání grafů
- Šifrování dat
- Volba typu primárního klíče