

Tehnologija upravljanja podatkov

Matjaž Kukar
Luka Šajn

2025/26





Splošne informacije...

* Predavanja

- Izr. prof. dr. Matjaž Kukar, sreda ob 10:15, P04
matjaz.kukar@fri.uni-lj.si
- Govorilne ure: po dogovoru
 - * kabinet 2.04 (2. nadstropje, desno od dvigala)

* Vaje:

- Doc. dr. Luka Šajn, ponedeljek 14h PR17, četrtek 14h PR08

* Spletna stran:

- Tehnologija upravljanja podatkov
<https://ucilnica.fri.uni-lj.si>



Upravljanje s podatki (data management)

- * Definicija (Data Management Association – DAMA)
Upravljanje s podatki sestavljajo razvoj in izvajanje arhitektur, usmeritev in praktičnih postopkov za podporo celotnemu življenjskemu ciklu podatkovnih potreb sodobnega podjetja.

Obravnavali bomo torej napredne teme s širšega področja podatkovnih baz



Pričakovano predznanje

- * Nadgradnja predmetov Osnove podatkovnih baz/Podatkovne baze
 - Opisovanje in shranjevanje podatkov v PB
 - * Zgradba SUPB, upravljanje z diskom in pomnilnikom
 - * Organizacija in indeksiranje datotek
 - Poizvedovanje v PB
 - * Relacijski podatkovni model, algebra in račun; SQL
 - Načrtovanje PB
 - * Pristopi k načrtovanju PB
 - * Konceptualno, logično in fizično modeliranje ???
- * Nadgradnja programerskih predmetov
 - Osnovno znanje programiranja in uporabe orodij (predvsem Python)
- * Razvoj informacijskih sistemov
 - Presek nekaterih tem (predvsem načrtovanje); poudarek na načrtovanju PB

Predpriprava na predavanja in vaje

- * Podatkovne baze / Osnove podatkovnih baz
- * Osvežite vsebine predmetov!
 - SQL
 - Relacijska algebra
 - Programski dostop do PB?
 - * ODBC, JDBC, DB API, ...
 - Upravljanje z različnimi SUPB:
 - * MySQL/MariaDB, SQLite/DuckDB, PostgreSQL, Oracle, MS SQL Server, ...
 - Sodobne nerelacijske podatkovne baze - NoSQL

Predpriprava na predavanja in vaje

- * Primeri na predavanjih-vajah v Pythonu. Ga obvladate?
 - Uradna Python dokumentacija, <http://docs.python.org>
 - Learn Python the hard way, <http://learnpythonthehardway.org/book/>
 - Learn Python (interaktivno), <http://www.learnpython.org/>
- * Aktualna verzija (oktober 2024):
 - Python 3.12; 3.9 ali novejši je OK (priporočena distribucija Anaconda)
- * Zakaj Python?
 - manjši obseg pisanja (večina primerov na enem zaslonu)
 - hitrejše programiranje
 - uporaba v podatkovni znanosti



Vsebina predavanj

- * Eksterni vidiki obvladovanja podatkov:
 - Sodobne nerelacijske podatkovna baze (NoSQL)
 - * Dokumentne, grafne, stolpčne, vektorske, ... vaše izkušnje?
 - Načrtovanje podatkovnih baz
 - * Konceptualno, logično in fizično načrtovanje
 - * Normalizacija relacij
 - * Analiza uporabniških zahtev
 - Podatkovne baze in podatkovna skladišča
 - * Namen in načrtovanje podatkovnih skladišč
 - * Zagotavljanje kvalitete shranjenih podatkov
 - * Analiza shranjenih podatkov (OLAP, Data Mining)



Vsebina predavanj

- * Interni vidiki obvladovanja podatkov:
 - Dostop do podatkov
 - Upravljanje sočasnosti dostopa do podatkovne baze (transakcije)
 - Varovanje in obnavljanje podatkovne baze
 - Upravljanje delno strukturiranih in nestrukturiranih podatkov
 - * JSON, XML, prostorski podatki, tekst v naravnem jeziku, ...
 - Formalni modeli za zagotavljanje dostopnosti in konsistentnosti podatkov
 - * ACID in BASE modela konsistentnosti podatkov
 - * Teorem CAP

Izhodiščna literatura

1.Relacijski SUPB

- a) Thomas M. Connolly, Carolyn E. Begg (2015). Database Systems, A Practical Approach to Design, Implementation and Management, 6th Edition, Pearson
- b) Ramez Elmasri, Shamkant B. Navathe (2016). Fundamentals of Database Systems, 7th Edition, Pearson
- c) S. Sumathi, S. Esakkirajan: Fundamentals of Relational Database Management Systems, Springer, 2007
- d) Carlos Coronel, Steven Morris (2018). Database Systems: Design, Implementation and Management, 13th Edition, Cengage Learning
- e) Raghu Ramakrishnan, Johannes Gehrke (2003). Database Management Systems, Third Edition, McGraw-Hill
- f) Paul Wilton and John W. Colby (2005): Beginning SQL, Wrox

Izhodiščna literatura

2. Nerelacijski (NoSQL) SUPB

- a) Ian Robinson, Jim Webber and Emil Eifrem: Graph Databases, O'Riley (free download), <https://neo4j.com/graph-databases-book>
- b) Dan McCreary and Ann Kelly: Making Sense of NoSQL: A guide for managers and the rest of us, Manning Publications, 2013
- c) Pramod J. Sadalage and Martin Fowler: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence , Addison-Wesley, 2012
- d) Alex Giamas: Mastering MongoDB 3.x: An expert's guide to building fault-tolerant MongoDB applications, Apress, 2017
- e) Open-source vector databases, Microsoft, 2024
<https://learn.microsoft.com/en-us/azure/cosmos-db/mongodb/vcore/vector-search-ai>



Vsebina vaj

I. Praktična obravnava nekaterih tem s predavanj

- Spoznavanje s programsko opremo
- Nerelacijski SUPB
- Načrtovanje podatkovne baze in podatkovnega skladišča
- Programski dostop do PB
- Normalizacija, priprave na normalizacijo, denormalizacija
- Transakcije
- Spremljanje izvajanja in optimizacija poizvedb

II. Domače naloge (vsaj 50%)

III. Seminarska naloga v obliki projekta (vsaj 50%)

- Na koncu semestra (december, januar)
- Nekoliko obsežnejša kot domače naloge
- Obvezen predstavitveni seminar (vaje ali predavanja)
- Pravočasnost izdelave, oddaje in zagovora!



Orodja

- * MongoDB, Neo4j, Schema/Milvus
- * PowerDesigner 12.5 (licenca FRI)
- * MariaDB ≥ 11 (priporočeno 11.5.2) ali
MySQL ≥ 8 (priporočeno 8.0.39)
- * PostgreSQL ≥ 15 (priporočeno 16.4)
- * MySQL Workbench, HeidiSQL
- * Python ≥ 3.9 + dodatki (priporočeno 3.12)
- * Gonilniki za dostop do baze (npr. ODBC)

- * Lastni računalniki?
- * Docker?



Izpitni red

1. Obveznosti vaj: iz domačih nalog in seminarske naloge morate doseči posamično najmanj 50% možnih točk (pogoj za pristop k pisnemu izpitu).
2. Pisni izpit morate za pozitivno oceno pisati najmanj 50%
3. Vaje sestavljajo pol ocene in veljajo eno šolsko leto!
4. Sodelovanje na predavanjih/vajah se nagrajuje (subjektivno, do 10%)

Torej: domače naloge $\geq 50\%$, seminarska naloga $\geq 50\%$,
izpit $\geq 50\%$

Kako izgleda vaše predznanje v praksi?

- * Programiranje: Python, Java, C/C++/C#, ...
- * Programski dostop do podatkovne baze
- * Nerelacijski SUPB
- * Načrtovanje podatkovnih baz (ER): OPB/PB, IS, RIS
- * Normalizacija PB: OPB/PB, RIS
- * SUPB: MySQL/MariaDB, PostgreSQL, SQLite/DuckDB, Oracle, Microsoft SQL Server, NoSQL ?????
- * Orodja ORM (SQLAlchemy, Hibernate, ...)

Teme seminarских nalog

- * Specifična tematika, večja količina podatkov
- * Praktična izvedba
- * Dva študenta v skupini
- * Teme (objavljene bodo na učilnici):
 - Praktična izvedba obdelave (analiza, vizualizacija, ...) nad izbranimi podatki (več nalog), na platformi SUPB + orodje
 - Specifične pregledne teme z vzorčno implementacijo (več nalog)
 - Stare teme seminarских nalog za vzorec

Domača naloga

- * Ponovite snov predmetov PB/OPB
- * SQL !!!
- * Python !!!
 - Vzpostavite si delovno okolje
 - Naredite čim več primerov
 - Ogromno literature na spletu
- * Razmislite o svojih kompetencah s področja predmeta in o tem kakšen tip seminarske naloge ali seminarja bi vam najbolj ustrezal
- * Oblikujte skupine po natanko dva študenta

Terminologija

- * Sistem za upravljanje s podatkovnimi bazami (SUPB)
- * Podatkovna baza (data base)
- * Podatkovno skladišče (data warehouse)
- * Sodobni (ne)relacijski SUPB:
 - NoSQL = Not only SQL
 - NewSQL

Sistemi za upravljanje s PB (SUPB)

- * Angleško: Database Management System (DBMS)
- * Sistem za upravljanje s podatkovno bazo – SUPB je programska oprema za obvladovanje velikih količin podatkov, shranjenih v vnapij točno določeni obliki (logičnem podatkovnem modelu)
- * Alternativa – shranjevanje v aplikaciji lastni obliki; problemi: neprenosljivost, nefleksibilnost ...
- * Obstaja veliko vrst SUPB. Omejili se bomo predvsem na relacijske in nekatere sodobne nerelacijske:
 - Oracle, Microsoft SQL Server, Postgres, MySQL/MariaDB,
 - MongoDB, Neo4j, MonadDB, ...

Podatkovna baza in podatkovno skladišče

- * Podobno, vendar ne enako!
- * Podatkovna baza (PB oz. DB):
 - OLTP sistem (on-line transaction al processing)
 - opisuje trenutno stanje
- * Podatkovno skladišče (PS oz. DW):
 - OLAP sistem (on-line analytical processing)
 - opisuje zgodovino vsebin OLTP sistema (pogosto več OLTP sistemov)
- * Oba pristopa tečeta na SUPB, vendar z različnimi prioritetami izvajanja (OLTP - hitro izvajanje transakcij, OLAP - hitra analiza)
- * Razlikujejo se tudi postopki načrtovanja

Načrtovanje PB, PS, NoSQL

- * Načrtovanje podatkovnih baz
 - konceptualno načrtovanje:
 - * obsežne PB, veliko število tabel
 - * korak pri načrtovanju inf. sistemov
 - logično načrtovanje (normalizacija relacij oz. tabel):
 - * direktno za manjše PB (nekaj deset tabel)
 - * preverjanje rezultatov konceptualnega načrtovanja
 - fizično načrtovanje
- * Načrtovanje podatkovnih skladišč
 - načrtovanje zvezdnih shem
- * Načrtovanje nerelacijskih PB
 - Relativno novo, nezrelo in neusklajeno področje
 - Neobstoj "naravnih" formalnih pristopov
 - Poudarek na horizontalni skalabilnosti!

Zakaj razlike?

- PB: hitro (sprotno) izvajanje vnaprej definiranih transakcij
- PS: hitro izvajanje spontanih analiz podatkov
- NoSQL: nerelacijski podatkovni modeli, fleksibilne sheme, skalabilnost!