

PTS Naloga 1 (50 točk)

Kontakt: martina.sestak@um.si

1. Iz svojega besedila naloge določite podatke, ki bodo obravnavani v vašem sistemu. Za tiste podatke pripravite ločena podatkovna modela za MongoDB in MySQL podatkovni bazi. **(10 točk)**
 2. V poljubnem programskem jeziku implementirajte REST API, ki podpira **najmanj** izvajanje POST (dodajanje novega dokumenta), PUT (posodabljanje dokumenta), DELETE (brisanje dokumenta) in GET zahtev (pridobivanje vseh dokumentov iz zbirke) za osrednjo zbirko (kolekcijo), v kateri beležite transakcije (dogodke) v MongoDB bazi (**opomba**: pri implementaciji ni dovoljeno uporabljati tretjih knjižnic kot ODM rešitev). **(5 točk)**
 3. Ustvarite novo MySQL bazo z ustreznimi tabelami v MS Azure platformi ali na lokalnem strežniku. Uvozite podatke v tabele in generirajte datoteko CSV (eno ali več njih) z združenimi podatki iz tabel, ki vsebuje(jo) najmanj 100 vrstic (opomba: CSV datoteka za tabele, ki niso povezane z ostalimi tabelami lahko vsebuje manj vrstic, vendar ne manj kot 10 vrstic). **(5 točke)**
 4. Napišite Map Reduce posel, ki bo opravljal analizo nad podatki v HDFS-u z uporabo filtriranja in agregacije podatkov (**opomba**: primer posla ne sme biti popolnoma enak primeru na vajah):
 - 1-5 točke: posel je podoben primeru z vaj in/ali analiza podatke iz le ene datoteke
 - 5-10 točk: posel združuje in analizira podatke iz več datotek
 5. Ustvarite primeren HBase podatkovni model za vse podatke iz vaše domene. Uvezite podatke v HBase glede na oblikovani model. **(5 točk)**
 6. V Apache Drill-u izvedite 2 SQL povpraševanji nad podatki v HBase-u:
 - 1-4 točki: izpis vrstic z uporabo filtriranja in agregacije:
 - 5-10 točk: izpis vrstic z uporabo filtriranja, agregacije in združevanja HBase tabel
-

Navodila za oddajo naloge

Nalogo 1 oddajte na eŠtudiju kot ZIP dokument, ki **nujno** vsebuje:

1. sliko ER modela MySQL baze,
2. posnetek zaslona, iz katerega so razvidni podatki v MySQL bazi (npr. rezultat SQL povpraševanja, v katerem združujete vsaj 2 tabeli),
3. posnetek zaslona iz HBase-a, iz katerega je razvidno kateri podatki so shranjeni v bazo (z uporabo *scan* ukaza),

4. Word dokument, ki vsebuje:

- prikaz podatkovnega modela MongoDB baze (slika ali BSON opis),
- posnetek zaslona z MongoDB Atlas-a, iz katerega je razvidna struktura MongoDB baze oz. baza, zbirke in dokumenti znotraj zbirke (informacija je dostopna pod opcijo *Collections* v MongoDB Atlasu-u za vašo gručo),
- posnetek zaslona iz Postmana, iz katerega je razvidno, da ste klicali API in pridobili rezultat,
- kodo Mapper in Reducer razredov ter posnetek zaslona, v katerem se vidi rezultat izvedbe MR posla (uporabite *cat* ukaz v HDFS-u),
- opis HBase podatkovnega modela v obliki teksta ali slike (stolpčne družine, ključi itd.),
- sintakso SQL povpraševanj v Apache Drill-u in posnetek zaslona z rezultati izvedbe le-teh.

Naloga 1 bo ocenjena na podlagi oddanih dokumentov in sicer glede na število točk za posamezne podnaloge. Pri zagovoru naloge bo tudi preverjeno razumevanje in poznavanje konceptov in tehnologij uporabljenih za reševanje naloge z dodatnimi vprašanji ter se bo kakovost zagovora ovrednotila z maksimalno 5 točkami.