Specifikacija predmetnog projekta iz predmeta Big data

* Problem koji predstavlja danas prilikom razvoja aplikacija je prevelika količina podataka koja je potreban za rad, složeni upiti , predugo čekanje za povratak informacija iz baze podataka, fizički manjak memorije jer je inicijalna baza može biti prepunjena. Kao moguća solucija je premeštanje arhitekture na cloud, gde je fleksibilnije zauzimanje memorije. U cloud-u postoji mogućnost da baza podeli na više manjih celina to jest u blokove koje će sadržati više baza i tako ubrzati proces pretraživanja. Pošto podaci mogu biti u različitim formatima postoji mogućnost da se koriste različite baze za skladištenje ali to neće biti razmatrano u ovom projektu već će se koristit samo jedan vid podataka i modela.
* Skup podataka koji će biti korišćen se može pronaći na [kaggle](https://www.kaggle.com/datasets). Podaci opisuju studente koji su bili podeljeni u grupe, imali su 3 testa (matematika, citanje, gramatika).
* Projekat bi trebalo da simulira rad baze podataka na cloud-u. Na lokalu će se pretvoriti set podataka u modele, po standardnom principu bi trebalo da se sve sačuva u jednu bazu ali radi simuliranja rada će biti izdeljeni na više baza po nekim logičkim načinom kako bi bilo najlakše da se podele. Web aplikacija je potrebno da poseduje neke forme za pretragu baze podataka i vizuelni prikaz pretraženih podataka.
* Baza podataka - postgres ili postgrREST,
* Web app - flask,
* Vizuelizacija podataka - Chart.js
* Maper podataka u baze - python app,
* Podaci su u csv.
* Cilj projekata je razumevanje/simulacija raspodele podataka po bazama na cloud-u, realizacija i funkcionisanje na lokalu i sam prikaz odabranih podataka.

Petar Canić 22.11.2022