

Operativni sistemi 2 – II kolokvijum
Institut za matematiku i informatiku, PMF Kragujevac
12.01.2018. god.

Parametarski posao

Banka "JezdaImi" periodično vrši kontrolu izvršenih transakcija. Dati su sledeći podaci:

U fajlu **korisnici.txt** se nalaze informacije o korisnicima banke "JezdaImi" u obliku (idKorisnika stanjeNaRacunu).

U direktorijumu **kvartali** dati su fajlovi **kvartal1.txt**, **kvartal2.txt**, ... koji sadrže podatke u obliku (idKorisnika1 idKorisnika2 iznos). Korisnik *idKorisnik1* je uplati iznos *iznos* na račun korisnika *idKorisnik2*. Podrazumeva se da su podaci unutar fajlova **kvartaliX.txt** sortirani po datumu, npr. transakcija u prvom redu je izvršena pre transakcije u drugom redu itd. Takođe, fajl sa manjim rednim brojem sadrži podatke o transakcijama iz ranijeg kvartala u odnosu na fajl sa većim rednim brojem.

Rezultat kontrole je spisak korisnika (idKorisnika) koji su u bilo kom trenutku imali negativni iznos na računu.

Napomena:

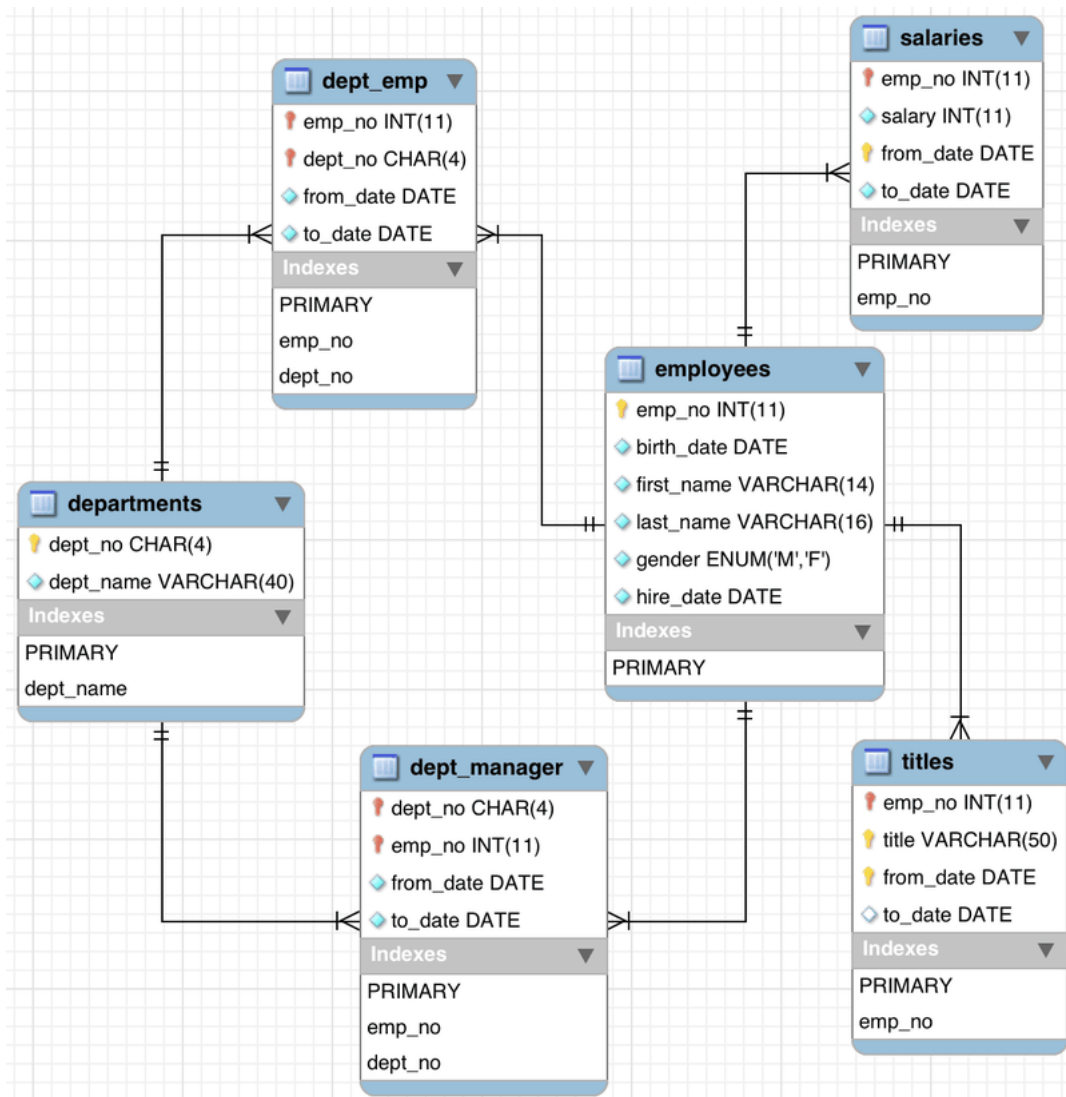
1. idKorisnika je niz brojeva i slova,
2. u svakoj transakciji makar jedan korisnik mora da ima račun u banci "JezdaImi",
3. iznosi su decimalni brojevi.

Rešenje dati u vidu parametarskog posla gde je potrebno kreirati:

1. Bash skriptu za pokretanje parametarskog posla (**pokreni.sh**), kao i integrisanje rezultata u fajl **resenje.txt**.
2. Odgovarajuće PBS komandne fajlove.
3. Programske kodove za obradu podataka i kreiranje rešenja.

Apache Spark

Dati su csv fajlovi koji sadrže podatke koji odgovaraju šemi sa slike.



Napisati pySpark skriptu koja za svaku titulu računa prosečnu dnevnu zaradu.

Napomene:

1. Ukoliko je vrednost datuma, u fajlu, veća od trenutnog datuma, taj datum zameniti sa trenutnim. (npr. 9999-01-01 postaje 2018-01-12)
2. Vrednost polja salary u tabeli salaries je plata na godisnjem nivou, broj dana u godini je 365.

Vreme za rad: 120 min.