

Operativni sistemi 2 – II popravni kolokvijum

Institut za matematiku i informatiku, PMF Kragujevac

4. februar. 2015. godine

Apache Hadoop (6 poena)

Neka su dati fajlovi u kojima se u svakom redu nalazi lista prijatelja jednog korisnika socijalne mreže Facebook. Format jedne linije fajla je dat u nastavku:

User_ID - Friend1_ID, Friend2_ID, ..., FriendN_ID

User_ID predstavlja ID korisnika za koga je data lista njegovih prijatelja u nastavku linije. Svi korisnici za koje je data lista prijatelja su prijatelji između sebe. Potrebno je kreirati **map/reduce** programski model koji pronalazi zajedničke prijatelje za svaki par korisnika iz datih fajlova, a zatim ih ispisuje na standardni izlaz. Ispis u svakoj liniji treba da sadrži par ID-eva korisnika koji imaju zajedničke prijatelje i spisak njihovih zajedničkih prijatelja. Za potrebe rešenja potrebno je kreirati fajl **CommonFriends.java** u kome se nalaze potrebne klase i metode za konfiguraciju i pokretanje posla na Hadoop platformi. **Main** metod pored putanje do ulaznih i izlaznih fajlova prihvata još i broj reduktora i broj mapera koje je potrebno setovati prilikom konfiguracije posla.

Primer:

Ulaz

```
user_1-user_5,user_2,user_4,user_6,user_10,user_3,  
user_2-user_3,user_6,user_11,user_1,  
user_3-user_6,user_10,user_22,user_2,user_1,
```

Izlaz

```
user_1, user_3 -> user_10, user_6, user_2,  
user_1,user_2 -> user_3, user_6,  
user_2,user_3 -> user_1, user_6,
```

Parametarski posao

Za potrebe uštede novca prilikom nabavke robe lanca prodavnica „Od igle do lokomotive“, potrebno je kreirati program koji će se izvršavati na Klaster računaru. Lanac prodavnica nabavlja robu od dobavljača koji su dati u fajlu **dobavljac.txt** formatom:

Id_dobavljacka X_koordinata Y_koordinata domaci

gde **Id_dobavljacka** predstavlja njegov jedinstveni identifikacioni broj, zatim slede 2D coordinate (date u kilometrima) dobavljača koje određuju njegov položaj i parametar **domaci** koji ima vrednost 1 ukoliko se

dobavljač nalazi u Srbiji i 0 ukoliko je dobavljač iz inostranstva. Pored pomenutog fajla dat je i fajl sa cenama proizvoda, **cene.txt**, kao:

Id_proizvoda id_dobavljacka cena id_valute

gde **id_proizvoda** predstavlja njegov jedinstveni identifikacioni broj, **id_dobavljacka** određuje dobavljača za koga važi **cena** i **id_valute** predstavlja jedinstveni identifikacioni broj valute u kojoj je data cena. Vrednost jednog apoeni stranih valuta u dinarima je dat u fajlu **valute.txt** kao:

id_valute dinarska_vrednost.

Neka se u folderu “**narudzbenice**” nalaze **tri fajla** koja za tri različite prodavnice sadrže spisak robe koju treba nabaviti. Fajlovi sa spiskom robe imaju sledeći format:

X_koordinata Y_koordinata

ID_proizvoda1 kolicina

ID_proizvoda2 kolicina

...

gde se u prvom redu nalaze 2D koordinate koje određuju položaj prodavnice, a zatim u ostalim redovima identifikacioni brojevi proizvoda i njihova **količina** koju treba nabaviti.

Potrebno je implementirati programsko rešenje koje će za svaki spisak iz foldera “**narudzbenice**” kreirati poseban fajl u kome će se za svaki traženi proizvod odrediti najpovoljniji dobavljač (njegov ID) i cena nabavke proizvoda, u dinarima. U cenu je potrebno uračunati prevoz i carinu robe od 10% na ukupnu sumu u slučaju da dobavljač nije domaći i cena robe prelazi 10 000 dinara. Prevoz robe se naplaćuje 5 dinara po kilometru. Format izlaznih fajlova je dat u nastavku:

ID_proizvoda kolicina ID_najpovoljnijeg_dobavljacka ukupna_cena

Nakon kreiranja fajlova sa najpovoljnijim cenama za svaki spisak potrebno ih je sortirati prema ukupnoj ceni nabavke u nerastućem poretku u formatu:

ime_fajla ukupnaCena

Programsko rešenje prilagoditi klaster računaru u vidu parametarskog posla koji će se izvršavati paralelno na tri računara i za to je potrebno kreirati:

- Bash skripta za pokretanje parametarskog posla. Rešenje mora sadržati fajl **pokreni.sh** koji pokreće parametarski posao, kreira fajlove sa spiskom najpovoljnijih dobavljača za svaki proizvod, nakon čega kreira fajl **rezultati.txt** u kome su oni sortirani prema ukupnoj ceni nabavke u nerastućem poretku.
- Odgovarajuće PBS komandne fajlove za pokretanje parametarskih poslova.
- Programske kodove za obradu fajlova i kreiranje rešenja.

