

OPERATIVNI SISTEMI 2 - II KOLOKVIJUM

INSTITUT ZA MATEMATIKU I INFORMATIKU, PMF KRAGUJEVAC

08. februar 2013. god.

Glavni problem (cluster,grid)

Neka je dato 8 tekstualnih datoteka (`data1.txt`, `data2.txt`, ..., `data7.txt`, `data8.txt`) gde svaka od njih sadrži po 100000 linija u formatu:

```
idKlijentaKojiUplacuje:idKlijentaKomeSeUplacuje:iznos:datumTransfera
```

Ukoliko u nekoj od datoteka imamo liniju:

```
113455:87345:12670.58:20120921
```

nju tumačimo kao novčani transfer koji se odigrao 21.09.2012. u iznosu od 12670.58, koji je klijent sa id-jem 113455 uplatio klijentu sa id-jem 87345.

Datoteke predstavljaju primitivnu particionisanu tabelu baze podataka o novčanim transferima iz različitih izvora. Potrebno je napisati programska rešenja (u C programskom jeziku) kojim će se odrediti podaci koji bi bili formulisani kao:

- Koji su različiti id-jevi klijenata koji su izvršili plaćanja počev od 1.07.2007 do danas sa iznosom većim od 10000.00?
- Kolika je ukupna uplaćena suma klijentu sa id-jem 18806 u transferima počev od 01.07.2007?

Rešenja koja se predaju u datotekama `upita.c` i `upitb.c`, uvezati sa odgovarajućim **bash** skriptama i skriptama za pokretanje `job`-ova na `cluster`-u i `grid`-u. `Job`-ovi moraju da imaju sledeće karakteristike:

- Svaki `job` pokreće 4 paralelna procesa obrade, gde svaki od njih obrađuje po 2 datoteke sa podacima;
- U slučaju `grid` rešenja, datoteke sa podacima o transferima moraju da budu postavljene na **storage element**, i da se odatle koriste;

Pored pokretanja pomoćnih akcija u obradi podataka prilikom odvijanja `job`-ova, obezbediti i skripte koje će integrisati rezultate oba `upita` iz 4 procesa u datoteke `izlaza.txt` i `izlazb.txt`. U rešenju pored svih pomenutih datoteka predati i detaljno uputstvo za pokretanje svih akcija uključujući i postavljanje ulaznih datoteka na **storage element**.

Dopunski (bonus) problem - PowerShell

Napisati PowerShell skriptu `analiza.ps1` koja:

- Sa komandne linije preuzima 2 parametra - putanju direktorijuma i ekstenziju (sa tačkom, recimo `.pdf`);
- U direktorijumu pronalazi sve datoteke sa ekstenzijom (rekurzivno, računajući poddirektorijume) i datoteku sa najvećom veličinom;
- U odgovarajućoj poruci ispisuje prosečnu veličinu pronađenih datoteka i datum zadnje promene najveće od njih.

Napomena : veličine ispisivati u MB.