

2. Domaća zadaća - ROVKP

Vinko Kolobara
9. travnja 2017.

1 ZADATAK: IZRADA MAPREDUCE PROGRAMA

Koliko različitih vozila se nalazi u ulaznoj datoteci?
U ulaznoj datoteci se nalazi 9834 različitih vozila.

Koliko je trajala najdulja ukupna vožnja jednog taksija? Koja je minimalna, a koja najdulja vožnja tog taksija?
Najdulja ukupna vožnja jednog taksija je trajala 225600s, minimalna 0s, a maksimalna 3360s.

Koje ste promjene morali napraviti na izvornom kodu prilikom uvođenja optimizacijske funkcije Combine?
Potrebno je samo postaviti sa `job.setCombinerClass` razred koji će se koristiti za Combiner, i to isti razred koji se koristi za Reducer.

Koliko vremena su se izvodile inačice programa? Je li to u skladu s vašim očekivanjem? Objasnite zašto.

Inačice programa su se izvodile približno jednako (26915ms bez, 25750ms sa Combiner-om), ali ipak sevarijanta s Combiner-om pokazala nešto bržom. Nije bilo za očekivati veću razliku zbog toga što se sve izvodilo u virtualnom stroju u pseudo-raspodijeljenom načinu s malom datotekom itd. pa hadoop nije mogao pokazati svoju moć.

2 ZADATAK: KORIŠTENJE OBRASCA DIJELJENJA PODATAKA (PARTITIONING PATTERN)

Koliko je vožnji realizirano u pojedinom području, tj. u užem centru i u širem gradskom području, i to tako da je broj putnika bio 1, 2-3 putnika ili 4 i više putnika?

U užem centru je realizirano 500061, a u širem gradskom području 763328 vožnji.

Koje ste promjene morali napraviti na izvornom kodu prilikom uvođenja funkcije Partition? Potrebno je samo poslu postaviti koji će se razred koristiti za Partitioner sa `job.setPartitionerClass`.

Koliko je vožnji navedeno u svakoj podskupini?

- Uži centar, 1 putnik - 244814
- Uži centar, 2-3 putnika - 70458
- Uži centar, 4+ putnika - 86294
- Šire područje, 1 putnik - 525525
- Šire područje, 2-3 putnika - 151509
- Šire područje, 4+ putnika - 184789

3 ZADATAK: IZVRŠAVANJE ULANČANIH MAPREDUCE PROGRAMA

Koliko ste MapReduce poslova izvršili u vašem kôdu?

Jedan posao za razdvajanje podataka na 6 particija, i još po jedan posao za svaku particiju što je ukupno 7 MapReduce poslova.

Koliko je različitih taksija realiziralo vožnje u pojedinom području, tj. u užem centru i u širem gradskom području, i to tako da je broj putnika bio 1, 2-3 putnika ili 4 i više putnika?

- Uži centar, 1 putnik - 6601
- Uži centar, 2-3 putnika - 4138
- Uži centar, 4+ putnika - 3425
- Šire područje, 1 putnik - 8714
- Šire područje, 2-3 putnika - 5487
- Šire područje, 4+ putnika - 3983