

Kursinio darbo užduotis

Kursinio darbo pagrindą turi sudaryti laboratorinių medžiaga papildyta ir praplėstą teorinę kurso medžiagą.

Kursinio darbo struktūra:

1. Įvadas: sprendžiamų uždavinių pristatymas, naudojamų priemonių (tame tarpe ir didžiųjų duomenų sistemų) pristatymas.
2. Pirmojo uždavinio sprendimas:
 - (a) Pateikite MapReduce programavimo modelį kaip iliustraciją naudodami pirmojo laboratorinio darbo medžiagą.
 - (b) Aptarkite Combiner tipo užduoties įtraukimo galimybes į laboratorinio uždavinio sprendimą, pagrįskite Combiner panaudojimo naudą.
 - (c) Aprašykite Hadoop MapReduce sistemos naudojimą pateikdami naudojimosi sistema algoritmą, tame tarpe ir pateikiant naudojamąs komandas.
3. Antrojo uždavinio sprendimas:
 - (a) Parodykite DAG programavimo modelio veikimą kaip iliustraciją naudodami antrojo laboratorinio darbo medžiagą.
 - (b) Aptarkite Spark transformacijas ir veiksmus, kurias naudojote antroje laboratoriniame darbe. Pakomentuokite transformacijų tipus (plačios/siauros), bei duomenų siuntimus dėl transformacijų.
4. Trečiojo uždavinio sprendimas:
 - (a) Python pagalba vizualizuokite duomenų aproksimaciją su tiesinės regresijos modeliu. Pateikite ir pakomentuokite Python bibliotekas ir komandas, panaudotas rezultatui gauti. Pakomentuokite duomenų pertvarkymą tarp skirtingų formatų/duomenų struktūrų.
 - (b) Jeigu kodas būtų vykdomas klasteryje, pakomentuokite kurios kodo vietos yra atsakingos už duomenų siuntimus tarp kompiuterių (serverio mazgai, lokalaus kompiuteris).
5. Priedai: pilnai veikiantys visų uždavinių sprendimo kodai Python kalba.