

MapReduce: Simplified Data Processing on Large Clusters

Maksim Norkin, ISIfm-13

Vilniaus Gedimino technikos universitetas
Fundamentinių mokslų fakultetas
Informacinių Sistemų katedra
`maksim.norkin@ieee.org`

2013 m. lapkričio 20 d.

Problema

- Bendras protokolas didelių duomenų apdorojimui tarp atskirų sistemos mazgų

Darbo tikslas

- Automatinis skaičiavimų lygiagretumas ir vykdymas tarp didelio kiekio skaičiavimo mazgų

Uždaviniai

- Aprašyti programavimo modelį
- Aprašyti MapReduce įgyvendinimą
- Aptarti galimus patobulinius
- Pateikti įvykdymo greitaveikos matavimus
- Parodyti įgyvendinimo pavyzdį Google sistemose
- Aptarti panašius ir būsimus darbus

lšvados

- Sėkmingas modelio panaudojimas Google sistemose
- Lengvas modelio panaudojimas
- Platus problemų sąrašas yra lengvai aprašomas per MapReduce skaičiavimus
- MapReduce algoritmo įgyvendinimas, kuris lengvai paskirstomas tarp tūkstančių skaičiavimo taškų
- Griežtas programavimo modelis leidžia lengviau skirstyti skaičiavimus tarp mašinų
- Tinklo pralaidumas yra esminė problema

Vertinimas I

- Turinys ir pavadinimas
 - Straipsnio pavadinimas tiesiogiai atitinka straipsnio turinį
- Aktualumas
 - Analizuojama problema yra labai aktuali
- Argumentavimas
 - Autoriai pateikia grafinius skaičiavimų įrodymus, pateikiamas konkretus veiksmų planas, kiekvienas žingsnis yra detaliai aprašomas ir apžvelgiamas
- Metodika
 - Modeliavimas, sisteminė analizė

Vertinimas II

- Nuoseklumas
 - Pradedama nuo metodo aprašymo, tuomet keliama į bendrą algoritmo įgyveninimą, aptariami galimi nesėkmių atvejai. Pateikiami realių problemų sprendimų pavyzdžiai.
- Problema, tikslas, uždaviniai, išvados
 - Iškelti uždaviniai išspręsti,
- Bloomo taksonomija
 - Pasiekiamas Veiksmų plano lygmuo
- Stilius
 - Darbo stilius yra nuoseklus ir suprantamas kiekvieno, kas tik yra susidūręs su IT pobūdžio straipsniais

Vertinimas III

- Literatūra
 - Literatūros sąrašas stiprus