

Porovnanie platforiem pre 'Big Data' riešenia



Abstrakt práce: Predmetom skúmania práce je porovnávanie platforiem pre Big Data. Cieľom práce je analyzovanie už existujúcich Hadoop ecosystem platforiem, ktoré slúžia na analýzu dát, a pomocou konkrétnych testovacích úloh zhodnotiť vhodnosť zvolených platforiem. V teoretickej časti práca obsahuje základné poznatky v oblasti mnohorozmerných dát a teoretické poznatky jednotlivých vybraných existujúcich Hadoop ecosystem platforiem. Teoretická časť bakalárskej práce ďalej obsahuje vytvorenie rôznych kritérií pre analýzu zvolených platforiem. Praktická časť práce obsahuje implementáciu jednotlivých platforiem na operačnom systéme Linux 14.04. Bakalárska práca ďalej obsahuje návrh dvoch testovacích úloh, ktoré boli aplikované na všetky zvolené platformy. V praktickej časti sa tiež nachádza podrobný opis aplikácie testovacích dopytov pre jednotlivé platformy. V závere praktickej časti sa nachádza vyhodnotenie vhodnosti zvolených platforiem. Predpokladaným prínosom práce je zhodnotenie zvolených platforiem na základe zvolených kritérií.

Kľúčové slová: hadoop, apache pig, apache hive, apache sqoop, apache spark

Univerzita Mateja Bela

Fakulta prírodných vied

Katedra Informatiky

Autor: Michal Buzák

Vedúci práce: PaedDr. Mgr. Vladimír Siládi, PhD