

Studiul expunerii unui sistem MapReduce sub formă de servicii REST

Flavius Condrea

Rezumat

Seturile de date mari devin tot mai răspândite. Chiar și în calitate de cercetători, în ziua de astăzi, trebuie lucrat în mod curent cu seturi de date care sunt de până la câțiva petabytes în dimensiune. Aplicațiile interesante din lumea reală produc volume imense de date nestructurate. După cum datele devin din ce în ce mai abundente tot așa apar și instrumente scalabile și ușor de utilizat pentru procesarea distribuită. Printre acestea, MapReduce este larg acceptat atât în mediul academic, cât și în industrie.

O utilizare largă a acestui model de programare o găsim în cadrul serviciilor Web de regăsire a informațiilor (căutare, sortare, construire index inversat), cercetare (descompunerea valorilor singulare, sortare distribuită) și aplicații ale inteligenței artificiale (traducerea de text, document clustering, machine learning).

Scopul proiectului este studiul MapReduce și al posibilității de a expune această funcționalitate sub formă de servicii REST.

Pentru a realiza această expunere s-a dezvoltat o aplicație având 2 module: un modul având o interfață web cu ajutorul căruia se pot încărca aplicații de tip MapReduce, iar un alt modul care permite lansarea în execuție a aplicațiilor încărcate și rularea acestora în cadrul mai multor instanțe. Prin intermediul interfeței web se pot seta diverși parametri de configurare (precum numărul de instanțe a fiecărui program, parametri suplimentari trimiși prin intermediul liniei de comandă). Aceste funcționalități sunt necesare pentru a simula comportamentul unui sistem de tip MapReduce. Comunicarea dintre cele 2 module ale sistemului se realizează prin intermediul unor servicii de tip REST.

Avantajele soluției propuse ar putea fi expunerea mult mai largă a sistemului MapReduce (deoarece serviciile REST sunt independente de platformă), scăderea costurilor hardware (deoarece mașinile pe care rulează programele MapReduce pot fi folosite de toți utilizatorii aplicației) și configurarea simplă care se realizează prin intermediul unei interfețe web.