

MapReduce para Aplicações Intensivas de Dados

Francisco Brasileiro

*Universidade Federal de Campina Grande
Departamento de Sistemas e Computação
Laboratório de Sistemas Distribuídos
Av. Aprígio Veloso, s/n, Bodocongó
58.429-900 - Campina Grande - PB
fubica@dsc.ufcg.edu.br*

Resumo

O modelo de programação Map/Reduce tem atraído bastante atenção da indústria e da academia, oferecendo uma forma simples, eficiente e escalável para o desenvolvimento de aplicações paralelas que processam grandes massas de dados. A popularização desse modelo de programação se deve em grande parte ao sucesso da implementação aberta Hadoop, desenvolvida pela Apache. O arcabouço Hadoop possui mecanismos para prover alta disponibilidade dos dados armazenados e mecanismos para prover tolerância a falhas dos nós de processamento. De maneira complementar, implementações simplistas do modelo Map/Reduce, como o BashReduce baseada em *scripts shell*, também ajudaram a disseminar a utilização desse paradigma, até mesmo para executar aplicações que processam pequenas quantidades de dados, haja vista que a utilização dessas implementações não requer a implantação de um aparato especial para dar suporte à execução de aplicações MapReduce. Este minicurso tem como objetivos apresentar o modelo Map/Reduce e alguns exemplos de aplicações que usam esse paradigma. Adicionalmente, serão discutidas as implementações Hadoop e BashReduce.