



PROCESSANDO DADOS DISTRIBUÍDOS COM HADOOP

Orientador: FORCHESATTO, André Luiz

Pesquisadora: DELABETHA, Priscila

Curso: Ciência da Computação

Área de Conhecimento: ACET

O homem sempre utilizou informação. Receber conhecimento, usá-lo e compartilhar com o outro é algo feito pela humanidade desde muito antes dos livros, quem dirá a partir das máquinas e dos supercomputadores. Assim, considerando a quantidade de informação que as pessoas têm acesso atualmente, é infinitamente maior do que há poucos anos atrás. Desse modo, é inevitável o surgimento gradativo de novos dados no universo da tecnologia da informação, crescendo também a preocupação com a necessidade de armazenar e manipular um volume de informações cada vez maior. Esta necessidade somente tem sido suprida em virtude da evolução de novas ferramentas e recursos, que surgem com o esforço computacional elevado tanto da comunidade acadêmica quanto da indústria, que vêm se dedicando na criação de novas tecnologias para melhorar a utilização de recursos computacionais no processamento de dados. Para que o processamento de uma grande quantidade de informação seja realizado de forma segura e em tempo viável, faz-se necessária a utilização de conceitos de processamento distribuído e computação paralela. Mas o desenvolvimento de *softwares* para ambientes distribuídos é um tanto complexo, visto que envolve uma série de requisitos a serem considerados pelos programadores como tolerância a falhas, balanceamento de carga, concorrência e distribuição de dados. O MapReduce é um *framework* desenvolvido na busca de solucionar esse problema, pois permite utilizar computação distribuída com um grande número de computadores para processar os dados. O MapReduce é um modelo de programação voltado para a computação em grade, e tem como objetivo processar grandes volumes de dados. O Hadoop é uma das implantações mais populares do MapReduce e vem sendo utilizado por grandes empresas de comunicação e informação pelo mundo. O estudo buscou comprovar o funcionamento da ferramenta Hadoop a partir da realização de testes em computação distribuída, com a codificação própria de projetos que utilizam as funções *map* e *reduce*. Com o desenvolvimento e codificação de processos foi possível conhecer detalhadamente o funcionamento do Hadoop, bem como comprovar o uso da ferramenta como uma solução válida para o desafio de processar e armazenar grande quantidade de informações com confiabilidade e disponibilidade. Os resultados confirmam, demonstram e comprovam a possibilidade de processamento com qualidade e a capacidade que a ferramenta suporta para lidar com o crescimento inevitável da quantidade de informação e a necessidade de manipular e armazenar tais dados com segurança.

Palavras-chave: Hadoop. MapReduce. Processamento distribuído. *Cluster*.

andre.forchesatto@unoesc.edu.br

pri.delabetha@gmail.com

