

O grupo de pesquisa tem foco em Sistemas Distribuídos, que consiste em uma união de diversos computadores, por uma rede, cujo objetivo é compartilhar a execução de tarefas. O subgrupo o qual o autor foi incumbido de trabalhar foi o subgrupo MapReduce, cujo objetivo é definir melhoras em aplicações as quais implementem tal modelo.

O modelo de programação MapReduce, tem por objetivo abstrair os detalhes de implementação de aplicativos distribuídos. Em outras palavras, o usuário da plataforma deve preocupar-se apenas com o desenvolvimento de sua solução, que segue o modelo MapReduce, e não com problemas advindos da distribuição do aplicativo em um cluster ou grade. A plataforma MapReduce encarrega-se de controlar a comunicação entre nós, distribuição de tarefas e dados, falhas, dentre outros aspectos comuns em ambientes distribuídos.

Foram realizados trabalhos em uma implementação do modelo MapReduce e um simulador do mesmo. Em ambiente real utilizou-se o Hadoop, uma implementação open source do MapReduce. A segunda, foi em ambiente simulado, utilizando o simulador MRSG (MapReduce over SimGrid), simulador determinístico implementado sobre o SimGrid pelo próprio grupo de pesquisa.

Em ambiente real, foram executados testes com o Hadoop na Grid'5000. Os testes foram agrupados com outros testes que o autor já havia desenvolvido no período passado de vigência da bolsa, para a validação do simulador MRSG. Tal atividade culminou na submissão de um artigo do grupo de pesquisa para o ParCo (Parallel Computing journal). Também foram realizadas a montagem e a configuração do novo cluster do grupo de pesquisa, o qual está atualmente funcionando com o Hadoop. Esse novo cluster é resultado de uma parceria do Instituto de Informática, do grupo de pesquisas GPPD/MapReduce e do projeto JitCloud - RNP, com UFSCar (Universidade Federal de São Carlos).

Em ambiente simulado foi feito um estudo sobre o impacto da rede heterogênea no modelo MapReduce. A motivação desse estudo foi a possibilidade de adaptação do modelo a ambientes de computação voluntária, onde tem-se uma heterogeneidade tanto de máquinas, quanto de rede. Os resultados de tal trabalho foram relatados em um artigo publicado no WSCAD-WIC 2011. Além disso, foram realizados experimentos para executar a análise 2K Fatorial das variações de algoritmos desenvolvidos para o modelo MapReduce em ambiente heterogêneo. O objetivo era de analisar o impacto dos novos algoritmos de divisão de dados, desenvolvidos pelo mestrando Julio Anjos. Os resultados de tais projetos foram incorporados em sua dissertação de mestrado. Tais algoritmos visam a adaptação do modelo para ambientes heterogêneos, tendo em vista que originalmente o MapReduce foi desenvolvido para ser utilizado em ambientes homogêneos, e deixa a desejar quando utilizado em ambientes com uma heterogeneidade de máquinas. Além da variação dos algoritmos, foi analisado o comportamento do modelo com a variação de vários outros parâmetros.