



***PULSAROPENSTACK-AD – INTRODUÇÃO AO PROCESSAMENTO DISTRIBUÍDO DE DADOS UTILIZANDO O HADOOP E OPENSTACK (SAHARA): UMA VISÃO DO USUÁRIO***

**Marianne Linhares Monteiro<sup>1</sup>, Andrey Elísio Monteiro Brito<sup>2</sup>**

**RESUMO**

Diante da rápida evolução tecnológica, constantemente surgem novos dispositivos com capacidade de processamento e comunicação e, conseqüentemente, novas fontes de dados. Desta forma uma questão cada vez mais em voga é a necessidade de processamento de grandes massas de dados de maneira eficiente, acessível e simples. Nesse sentido a ferramenta de Big Data Hadoop se encontra em ascensão. Além da utilização direta do Hadoop, projetos como o Sahara do Openstack se mostram como uma excelente alternativa tanto para usuários mais experientes quanto para novos utilizadores, oferecendo todo o poder de processamento do Hadoop, associado às vantagens do armazenamento em nuvem, uma interface amigável, fácil escalonamento e configuração automática das máquinas virtuais. Neste trabalho, foi feito um estudo a respeito do Hadoop, Openstack e OpenStack Sahara buscando introduzir essas tecnologias de maneira sucinta e eficaz, resultando tanto em uma referência para a introdução de um usuário à essa tecnologia, como também em um modelo de aplicação para recebimento e armazenamento contínuo de dados para um futuro processamento.

**Palavras-chave:** big data, computação em nuvem, map reduce.

***DISTRIBUTED PROCESSING OF DATA USING HADOOP AND OPENSTACK(SAHARA): AN USER'S POINT OF VIEW***

**ABSTRACT**

Due to the increasingly fast technological evolution, constantly new smart devices appear and consequently become new sources of data. Therefore, an important application is the need for processing big amounts of data in a simple, accessible and efficient mode. In this direction, the use of the Big Data tool Hadoop is increasing common and the new project from Openstack called Sahara is an excellent choice for both the experienced and the beginner user, offering all the processing power of the best known big data tool, associated with the advantages of cloud computing, a friendly interface, easy configuration and scalability. A study was made on Hadoop, Openstack and OpenStack Sahara targeting to introduce these technologies in a summarized and efficient way, resulting in the production of an user's introduction reference and in an application model for continuously receiving and storing large amounts of data for later processing.

<sup>1</sup>Aluna do Curso de Ciência da Computação, Departamento de Sistemas e Computação, UFCEG, Campina Grande, PB, e-mail: marianne.monteiro@ccc.ufcg.edu.br

<sup>2</sup>Ciência da Computação, Professor Doutor, Departamento de Sistemas e Computação UFCEG, Campina Grande, PB, e-mail: andrey@dsc.ufcg.edu.br