

# Instalação e Caso de Uso da Ferramenta *Hadoop*

Ariel F. Azevedo<sup>1</sup>, Rodrigo B. Acosta<sup>2</sup>, Rodrigo M. Duarte<sup>1</sup>

<sup>1</sup>Laboratório de Sistemas Paralelos e Ubíquos - Universidade Federal de Pelotas  
– Pelotas – RS – Brasil

<sup>2</sup>Centro de Desenvolvimento e Tecnologia - Universidade Federal de Pelotas  
– Pelotas – RS – Brasil

{afazevedo, rbacosta, rmduarte}@inf.ufpel.edu.br

**Abstract.** *This article describes how to install the Hadoop tool, showing step by step of the all things we've made*

**Resumo.** *Este artigo descreve como instalar a ferramenta Hadoop , mostrando passo a passo de todas as coisas que fizemos*

## 1. Introdução

O projeto do *Apache Hadoop* desenvolve um *software* para computação confiável, escalonável e distribuída. A biblioteca dessa ferramenta é um *framework* que permite processamento distribuído de grandes quantidades de dados através de *clusters* de computação usando modelos simples de programação. Ele é projetado para escalar a partir de um único servidor para milhares de máquinas , cada um oferecendo computação e armazenamento local.

Esse artigo tem como objetivo mostrar como foi feito a instalação passo a passo dessa ferramenta em uma máquina, bem como seus requisitos, e foi projetado da seguinte forma: na Sessão 1 foi feita uma introdução ao trabalho, na Sessão 2 foi mostrado a Metodologia usada no trabalho, na Sessão 3 foi mostrado o passo a passo da instalação da ferramenta.

## 2. Metodologia

Esta sessão tem como objetivo mostrar a metodologia, com intuito de informar o leitor a máquina onde a instalação foi feita, bem como os requisitos necessários e quais as dificuldades da instalação.

### 2.1. Máquina

A estação onde foi feita a instalação possui a seguinte configuração:

- Processador Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz, com cache L1 de 8Mb
- Memória RAM DDR3 de 4Gb
- Sistema Operacional Ubuntu 14.04 AMD64 bits
- Disco Rígido de 1Tb

### 2.2. Requisitos

Os seguintes requisitos são necessários para fazer a instalação:

- SSH instalado e configurado
- Java da Oracle<sup>1</sup>

---

<sup>1</sup>Em distribuições Linux não pode ser a opção OpenJDK

## 2.3. Dificuldades

Para conseguir efetuar a instalação com maior facilidade em distribuições Linux, é preciso um nível razoável de conhecimento em *shell*, pois como tudo é feito por linha de comando, caso haja algum problema é necessário conhecimento para facilitar na instalação.

## 3. Instalação

Essa sessão tratará do passo a passo para a instalação, começando mostrando como instalar e configurar os requisitos mostrados na Subsessão 2.2.

### 3.1. Instalação de Requisitos

#### 3.1.1. SSH

- *sudo apt-get install openssh-server*
- *ssh-keygen -t rsa -P ""*
- *\$HOME/.ssh/id\_rsa.pub >> \$HOME/.ssh/authorized\_keys*

#### 3.1.2. Oracle Java

- *sudo apt-get purge openjdk\**
- *sudo add-apt-repository ppa:webupd8team/java*
- *sudo apt-get update*
- *sudo apt-get install oracle-java8-installer*

### 3.2. Instalação do Hadoop

#### 3.2.1. Download, extração e movimentação do pacote

- *wget http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.6.0/hadoop-2.6.0.tar.gz*<sup>1</sup>
- *tar xvzf hadoop-2.6.0.tar.gz*<sup>2</sup>
- *sudo mv \* /usr/local/hadoop*<sup>3</sup>
- *sudo chown -R user:group /usr/local/hadoop*<sup>4</sup>

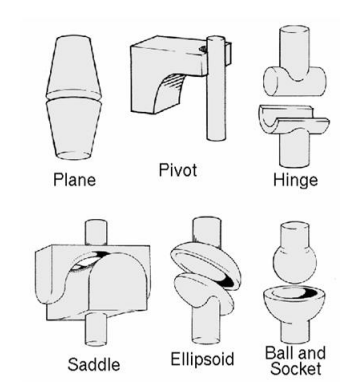
### 3.3. Arquivos de Configuração

Os Arquivos a seguir serão modificados para completar a configuração do *Hadoop*:

1. *~/.bashrc*
2. */usr/local/hadoop/etc/hadoop/hadoop-env.sh*
3. */usr/local/hadoop/etc/hadoop/core-site.xml*
4. */usr/local/hadoop/etc/hadoop/mapred-site.xml.template*
5. */usr/local/hadoop/etc/hadoop/hdfs-site.xml*



**Figure 1. A typical figure**



**Figure 2. This figure is an example of a figure caption taking more than one line and justified considering margins mentioned in Section 4.**

## 4. Figures and Captions

Figure and table captions should be centered if less than one line (Figure 1), otherwise justified and indented by 0.8cm on both margins, as shown in Figure 2. The caption font must be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.

In tables, try to avoid the use of colored or shaded backgrounds, and avoid thick, doubled, or unnecessary framing lines. When reporting empirical data, do not use more decimal digits than warranted by their precision and reproducibility. Table caption must be placed before the table (see Table 1) and the font used must also be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.

<sup>1</sup>Download do Pacote

<sup>2</sup>Extração do pacote

<sup>3</sup>Substitua \* pelo nome da pasta do Hadoop

<sup>4</sup>substitua user e group pelos respectivos usuario e grupo

**Table 1. Variables to be considered on the evaluation of interaction techniques**

|  | Chessboard<br>top view | Chessboard<br>perspective view |
|--|------------------------|--------------------------------|
| Selection with side<br>movements         | 6.02 $\pm$ 5.22        | 7.01 $\pm$ 6.84                |
| Selection with in-<br>depth movements    | 6.29 $\pm$ 4.99        | 12.22 $\pm$ 11.33              |
| Manipulation with<br>side movements      | 4.66 $\pm$ 4.94        | 3.47 $\pm$ 2.20                |
| Manipulation with in-<br>depth movements | 5.71 $\pm$ 4.55        | 5.37 $\pm$ 3.28                |

## 5. Images

All images and illustrations should be in black-and-white, or gray tones, excepting for the papers that will be electronically available (on CD-ROMs, internet, etc.). The image resolution on paper should be about 600 dpi for black-and-white images, and 150-300 dpi for grayscale images. Do not include images with excessive resolution, as they may take hours to print, without any visible difference in the result.

## 6. References

Bibliographic references must be unambiguous and uniform. We recommend giving the author names references in brackets, e.g. [Knuth 1984], [Boulic and Renault 1991], and [Smith and Jones 1999].

The references must be listed using 12 point font size, with 6 points of space before each reference. The first line of each reference should not be indented, while the subsequent should be indented by 0.5 cm.

### References

- Boulic, R. and Renault, O. (1991). 3d hierarchies for animation. In Magnenat-Thalmann, N. and Thalmann, D., editors, *New Trends in Animation and Visualization*. John Wiley & Sons Ltd.
- Knuth, D. E. (1984). *The T<sub>E</sub>X Book*. Addison-Wesley, 15th edition.
- Smith, A. and Jones, B. (1999). On the complexity of computing. In Smith-Jones, A. B., editor, *Advances in Computer Science*, pages 555–566. Publishing Press.