

# HDFS – Hands-On

**Objetivo:** O objetivo deste hands-on é utilizar os principais comandos do HDFS e executar um job usando dados armazenados no HDFS.

## Exercício

1. Faça o download de 3 livros do Projeto Gutenberg (<http://www.gutenberg.org/>). Os livros devem ser baixados no formato Plain Text UTF-8.

- The Outline of Science, Vol. 1 (of 4) by J. Arthur Thomson - <http://www.gutenberg.org/etext/20417>
- The Notebooks of Leonardo Da Vinci - <http://www.gutenberg.org/etext/5000>
- Ulysses by James Joyce - <http://www.gutenberg.org/etext/4300>

2. Execute as operações abaixo:

- Criar uma pasta no diretório HDFS do usuário local (/user/\$USER) chamado gutenber
- Copiar os arquivos para esta pasta no HDFS
- Listar os arquivos existentes na pasta /user/\$USER/gutenber do HDFS

3. Apresente a estrutura de blocos gerados na máquina local (DataNode) localizados no diretório /home/vagrant/hadoop-vagrant/dfs/data/current.

4. Apresente os arquivos com metadados do HDFS (fs\_image e edits) na máquina local (NameNode) localizados no diretório /home/vagrant/hadoop-vagrant/dfs/name.

5. Execute a aplicação wordcount. A aplicação fará a leitura dos arquivos existentes em /user/\$USER/gutenber e armazenará os resultados em /user/\$USER/gutenber-output.

```
$ $HADOOP_HOME/bin/hadoop jar $HADOOP_HOME/share/Hadoop/mapreduce/hadoop-mapreduce-examples-2.9.1.jar wordcount /user/$USER/gutenber /user/$USER/gutenber-output
```

6. Liste o conteúdo do diretório de saída para verificar se os arquivos foram gerados corretamente.

7. Apresente o conteúdo do arquivo de saída sem copiar o arquivo.

8. Copie o resultado do HDFS para o diretório local.

9. Remova os dados utilizados como entrada e saída do HDFS.

10. Faça um relatório apresentando todo o processo realizado para executar as etapas acima.