

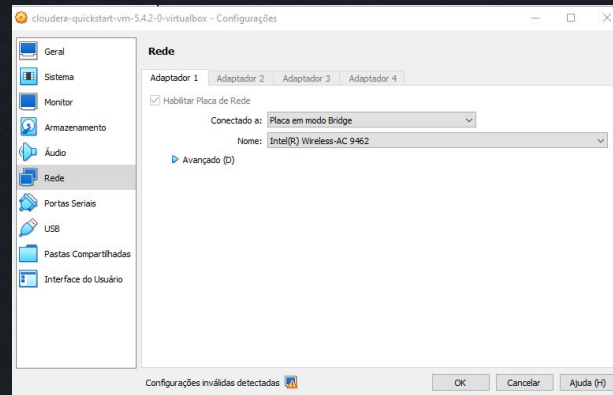
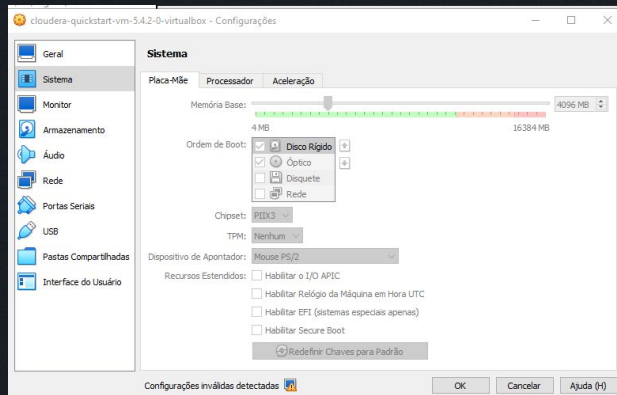


Aula 1 - Hadoop - Prática

Lucio Monteiro

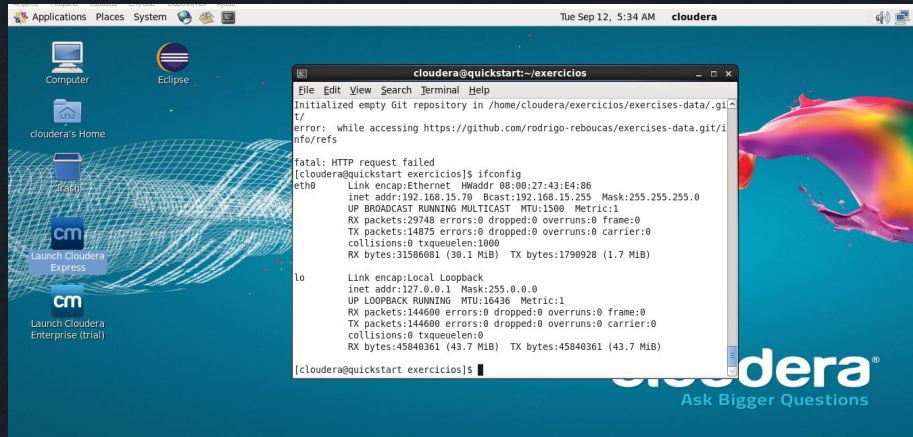
Ambiente Hadoop

- 1) Baixar o Oracle VM VirtualBox - [Oracle VM VirtualBox - Downloads | Oracle Technology Network | Oracle Brasil](#)
- 2) Baixar a Imagem do Linux com ambiente Hadoop
https://downloads.cloudera.com/demo_vm/virtualbox/cloudera-quickstart-vm-5.12.0-0-virtualbox.zip
- 3) Baixar o arquivo cloudera-quickstart-vm-5.4.2-0-virtualbox.zip e descompactar em uma pasta
- 4) No Oracle VM VirtualBox
 - a) Clicar em Arquivo
 - b) Importar Appliance
 - c) Selecionar o arquivo que baixou e descompactou
- 5) Fazer 2 configurações no Oracle VM Virtualbox para evitar problemas ao acessar:
 - a) Botão direito Em configurações:

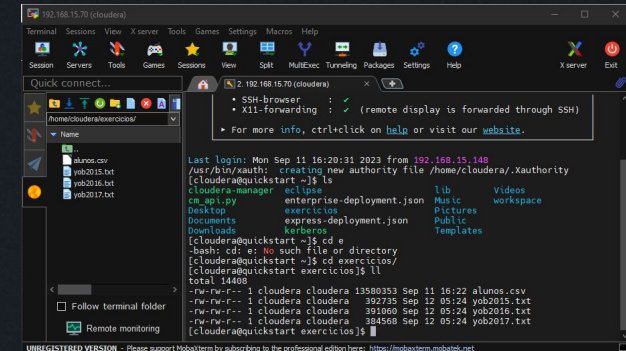


Ambiente Hadoop

- 1) Ao término - clicar no nome da máquina com o botão direito e clicar em Iniciar
- 2) Abrir um terminal do Linux



- 3) Baixar e instalar o programa MobaXterm
<https://mobaxterm.mobatek.net/download-home-edition.html>
Vai facilitar o acesso e upload de arquivos para o ambiente Cloudera



Ambiente Hadoop

- 1) Abrir um terminal do Linux
- 2) mkdir exercicios
- 3) cd exercicios
- 4) Baixar e salvar na pasta acima:

<https://dados.ifpb.edu.br/he/dataset/alunos/resource/29c2b593-ed14-4b73-b30c-d6135f072cf7>

5) Comandos HDFS:

- a) hdfs dfs -help
- b) hdfs dfs -ls -R /
- c) hdfs dfs -ls -R /user
- d) hdfs dfs -mkdir /user/aluno
- e) hdfs dfs -put alunos.csv /user/aluno
- f) hdfs dfs -ls -R /user
- g) hdfs dfs -mkdir /user/aluno/teste
- h) hdfs dfs -cp /user/aluno/alunos.csv /user/aluno/teste
- i) hdfs dfs -ls -R /user/aluno
- j) hdfs dfs -mv /user/aluno/teste/alunos.csv /user/aluno/teste/alunos1.csv
- k) hdfs dfs -ls -R /user/aluno
- l) hdfs dfs -rm /user/aluno/teste/alunos1.csv
- m) hdfs dfs -rmdir /user/aluno/teste/
- n) hdfs dfs -du -h /user/aluno/
- o) hdfs dfs -df -h /
- p) hdfs dfs -cat /user/aluno/alunos.csv | head -n 2
- q) hdfs dfs -tail /user/aluno/alunos.csv
- r) hdfs dfs -setrep 2 /user/aluno/alunos.csv
- s) hdfs dfs -checksum /user/aluno/alunos.csv

Ambiente Hadoop

- 1) Comandos YARN:
 - a) `yarn application -list`

Obrig.ada