

Exercício 1

1. Baixe o trace dos acessos feito ao departamento de computação de calgary:

http://ita.ee.lbl.gov/traces/calgary_access_log.gz

2. Descompacte :)

3. O formato do arquivo é como se segue:

ORIGEM - - DATA “AÇÃO” STATUS TAMANHO

Onde **ORIGEM** pode ser local ou remote, a data é no formato DD/Mon/YYYY:HH:MM:SS tz a ação é um comando HTTP, o Status é um código de resultado e o tamanho é o tamanho da requisição (ou - quando não existir).

Exemplo:

local - - [24/Oct/1994:13:47:19 -0600] "GET 1.gif HTTP/1.0" 200 1210

4. Sua primeira atividade é descartar as linhas que tem nenhum ou apenas um traço depois de remote ou de local. Exemplo, as linhas abaixo devem ser DESCARTADAS:

local - [01/Feb/1995:14:46:36 -0700] "GET /images/gtr.gif 1350.0" 200 36403
local index.html

5. Sumarize os dados desse arquivo indicando:

- a) quantas requisições locais foram feitas
- b) quantas requisições remotas foram feitas
- c) em média, qual a hora (apenas o campo hora e só o valor inteiro) em que requisições locais são feitas
- d) em média, qual a hora (apenas o campo hora e só o valor inteiro) em que requisições remotas são feitas

6. Adicione uma funcionalidade complementar ao seu programa! Por exemplo, calcular o tipo de requisição mais feita, a média de tamanho das requisições, uma estatística sobre o arquivo mais acessado, etc.