

Diário de Bordo

Análise das requisições para o servidor NASA Kennedy Space Center WWW na Flórida por um período específico.

Maicon Henrique Cunha

maiconhenrique@hotmail.com

www.maiconcunha.com



Fonte dos Dados



Site: <http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html>



Dados:

ftp://ita.ee.lbl.gov/traces/NASA_access_log_Jul95.gz
ftp://ita.ee.lbl.gov/traces/NASA_access_log_Aug95.gz

Questões Principais.

Determinar:

1. Número de hosts únicos.

2. O total de erros 404.

3. Os 5 URLs que mais causaram erro 404.

4. Quantidade de erros 404 por dia.

5. O total de bytes retornados.

Estratégia Adotada



1. Utilizar a plataforma [Databricks Community Edition](#) para realizar um primeiro processo de extração e transformação dos dados.



2. A partir desse pré-processamento obter um arquivo para utilização no software Tableau.



3. Realizar um segundo processo de transformação dos dados no Tableau.



4. Construir um único dashboard no Tableau onde seja possível visualizar todas as respostas das questões anteriores além de possibilitar outros insights sobre o banco de dados.

Linguagens, plataformas e softwares utilizados



DATABRICKS



PYTHON



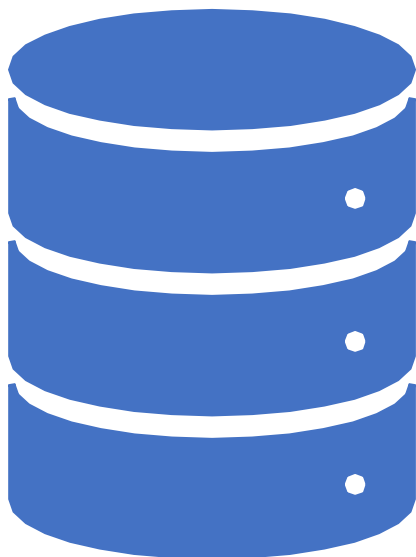
SPARK



TABLEAU



POWER POINT

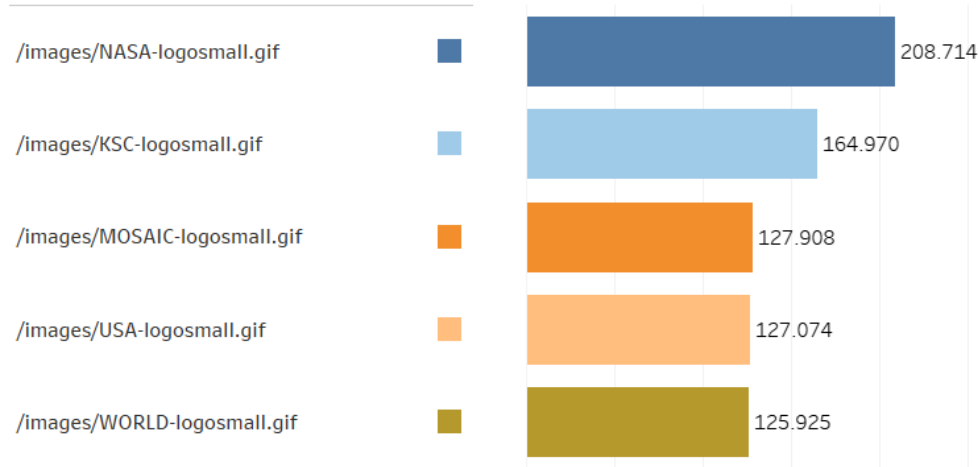


Tour pelos Arquivos

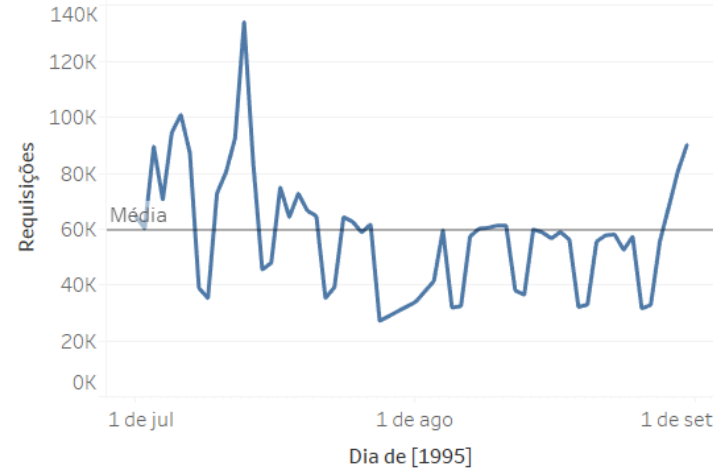
- Versão html do pré-processamento realizado na plataforma Databricks: [preprocess_logs_nasa.html](#)
- Dashboard com as respostas desenvolvido Tableau: [dashboard tableau.twbx](#)
- Link para o dashboard anterior desenvolvido no Tableau em versão web [aqui](#).
- Observação: não realizei o upload do arquivo csv utilizado para construir os dashboards por ser muito grande. Caso seja de interesse posso fazer o upload desse arquivo.

Análise Logs NASA - Jul/Ago 1995

Url's mais requisitadas



Requisições por Dia	
01/07/95	64.714
02/07/95	60.265
03/07/95	89.584
04/07/95	70.452
05/07/95	94.575
06/07/95	100.960
07/07/95	87.233
08/07/95	38.867
09/07/95	35.272
10/07/95	72.860
11/07/95	80.407
12/07/95	92.536
13/07/95	134.203
14/07/95	84.103
15/07/95	45.532



Selecione o Código

- ☒ (Tudo)
- ☒ 200
- ☒ 302
- ☒ 304
- ☒ 400
- ☒ 403
- ☒ 404
- ☒ 500
- ☒ 501

Bytes Total

65.524.314.915

Total de Hosts Únicos

137.932

Requisições por Código

200	3.100.524
304	266.773
302	73.070
404	20.899
403	225
500	65
501	41
400	15
Total geral	3.461.612

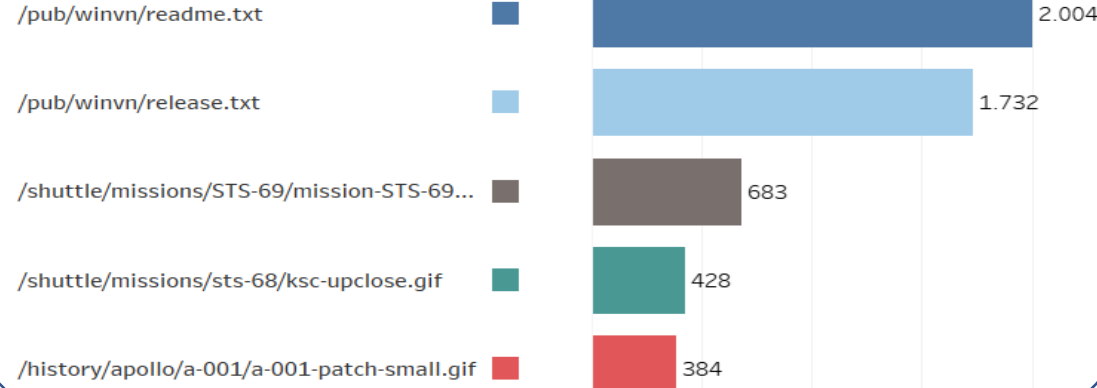
Resposta 5

Resposta 1

Dashboard: Tableau (tudo)

Análise Logs NASA - Jul/Ago 1995

Url's mais requisitadas



Selecione o Código

- ☐ (Tudo)
- ☐ 200
- ☐ 302
- ☐ 304
- ☐ 400
- ☐ 403
- ☒ 404
- ☐ 500
- ☐ 501

Bytes Total

0

Total de Hosts Únicos

8.268

Requisições por Código

404	20.899
Total geral	20.899

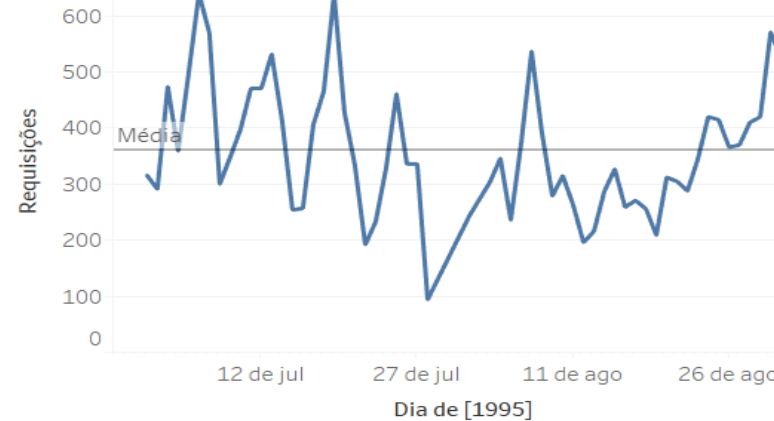
Resposta 3

Resposta 2

Resosta 4

Requisições por Dia

01/07/95	316
02/07/95	291
03/07/95	474
04/07/95	359
05/07/95	497
06/07/95	640
07/07/95	570
08/07/95	300
09/07/95	348
10/07/95	398
11/07/95	471
12/07/95	471
13/07/95	532
14/07/95	413
15/07/95	254



Dashboard: Tableau (erro 404)

Observações sobre o dashboard



É possível visualizar a versão web desse dashboard sem que tenha sido instalado o Tableau e não é necessário licença. Qualquer pessoa com [esse](#) link pode acessar o conteúdo.



O dashboard possibilita responder questões além das formuladas.



O dashboard está otimizado para desktops, mas é possível visualizar também em tablet e celular.



Clicando sobre um dia específico o dashboard é filtrado por esse dia.



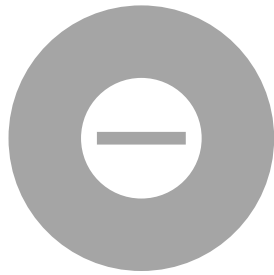
Clicando sobre uma url específica, o dashboard é filtrado por essa url.

Considerações Finais

A seleção das ferramentas utilizadas levou em consideração:



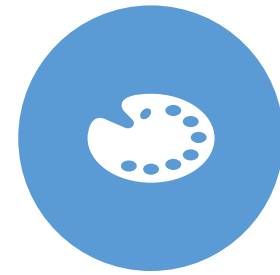
DIMENSÃO DOS
ARQUIVOS



FACILIDADE DE
ACESSO



CONHECIMENTO
PESSOAL



ESTÉTICA