

Análise de Falhas de Script Utilizando o Log de Acesso do Apache

Prof. Diogo Silveira Mendonça - CEFET-RJ

diogo.mendonca@cefet-rj.br

12/09/2016

Problema

- Sistemas web falham
- Usuários não reportam sempre as falhas
- Como saber que ocorreram falhas? E onde elas ocorreram?

Logs de Acesso do Servidor Web

- Informações sobre o uso do sistema
- Campos de Dados
 - IP
 - Data e Hora do Acesso
 - URL
 - Código HTTP
 - Bytes Transferidos
 - Página de Origem (Referer)
 - Agente de Usuário

Logs de Acesso do Servidor Web

- Códigos HTTP dão indícios sobre o resultado de cada requisição
- HTTP 200 – Acesso OK
- Em especial o HTTP 500 – Internal Server Error é retornado por falhas abrupta em software (exceções/erros não tratados)

Estudos de Caso

- Caso 1 (vou apresentar hoje)
 - Sistema de Gestão Financeira Empresarial
 - Comercializado no modelo SaaS na Web
 - 1 ano de logs
 - Dados de 2015
- Casos 2 e 3 (não serão apresentados)
 - Dentre 6 sistemas web do CEFET-RJ estudados 2 apresentaram problemas
 - 1 mês de logs de acesso
 - Dados de 2016

Estudos de Caso 1

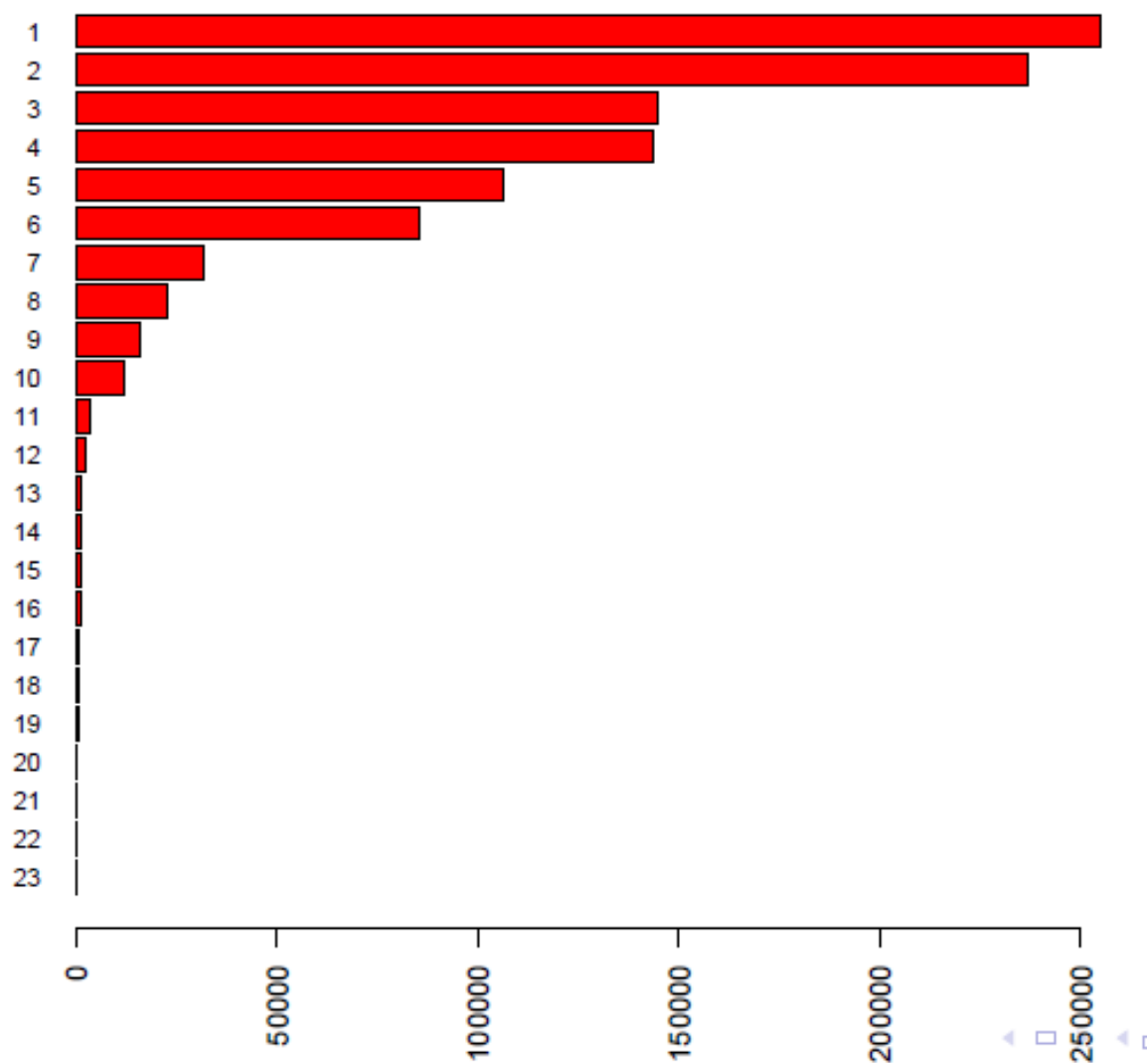
- Removidos dados de acesso a URLs estáticas
- Somente acessos a URLs que apontam para scripts
- 1.357.203 registros no log

Códigos HTTP

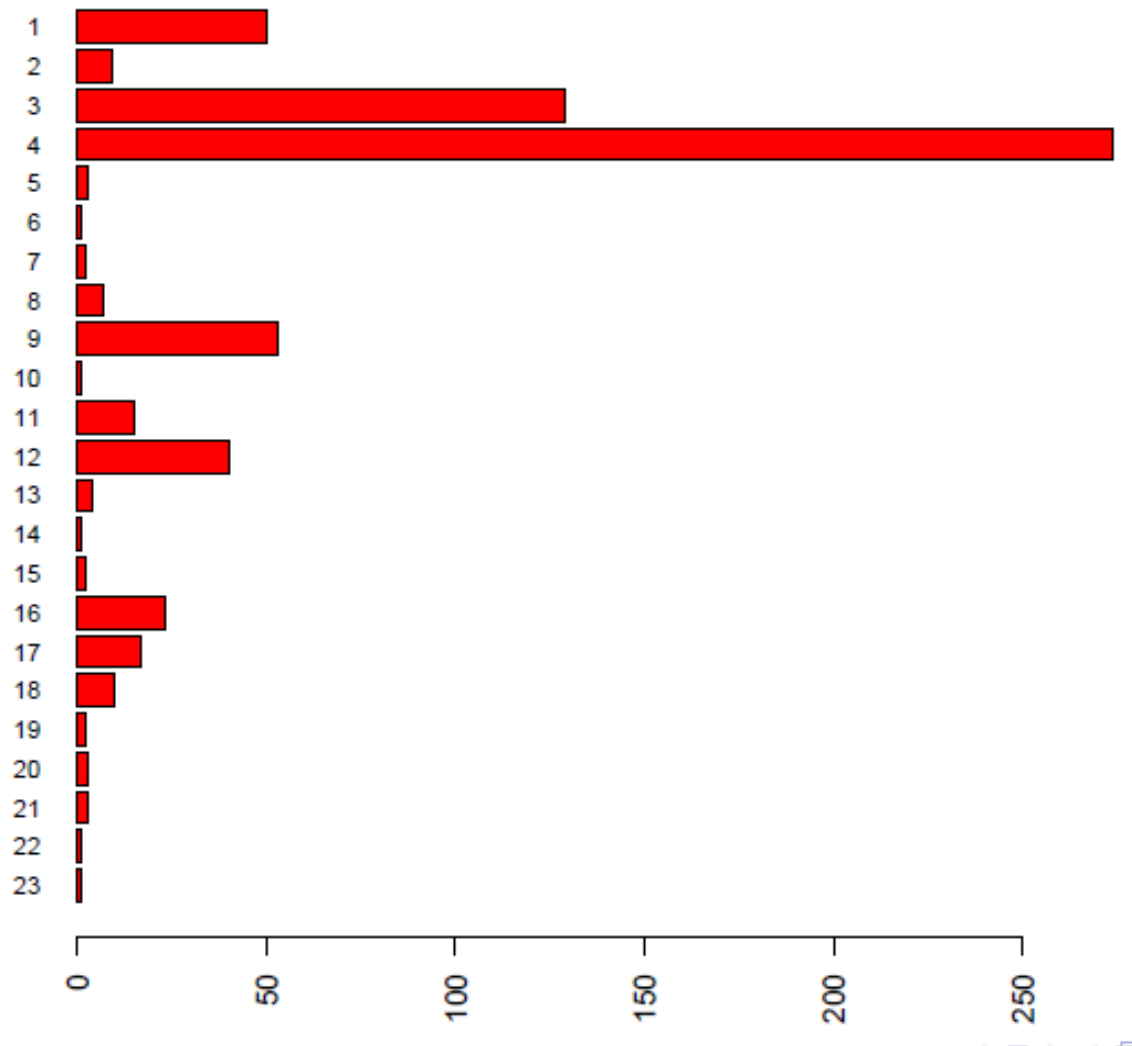
HTTP Code	Hits	% Total
200	1355524	99.876
302	978	0.072
403	33	0.002
500	668	0.049

- ▶ 200 - Normal Access
- ▶ 302 - Found (Redirect)
- ▶ 403 - Forbidden
- ▶ 500 - Internal Server Error - Script Error

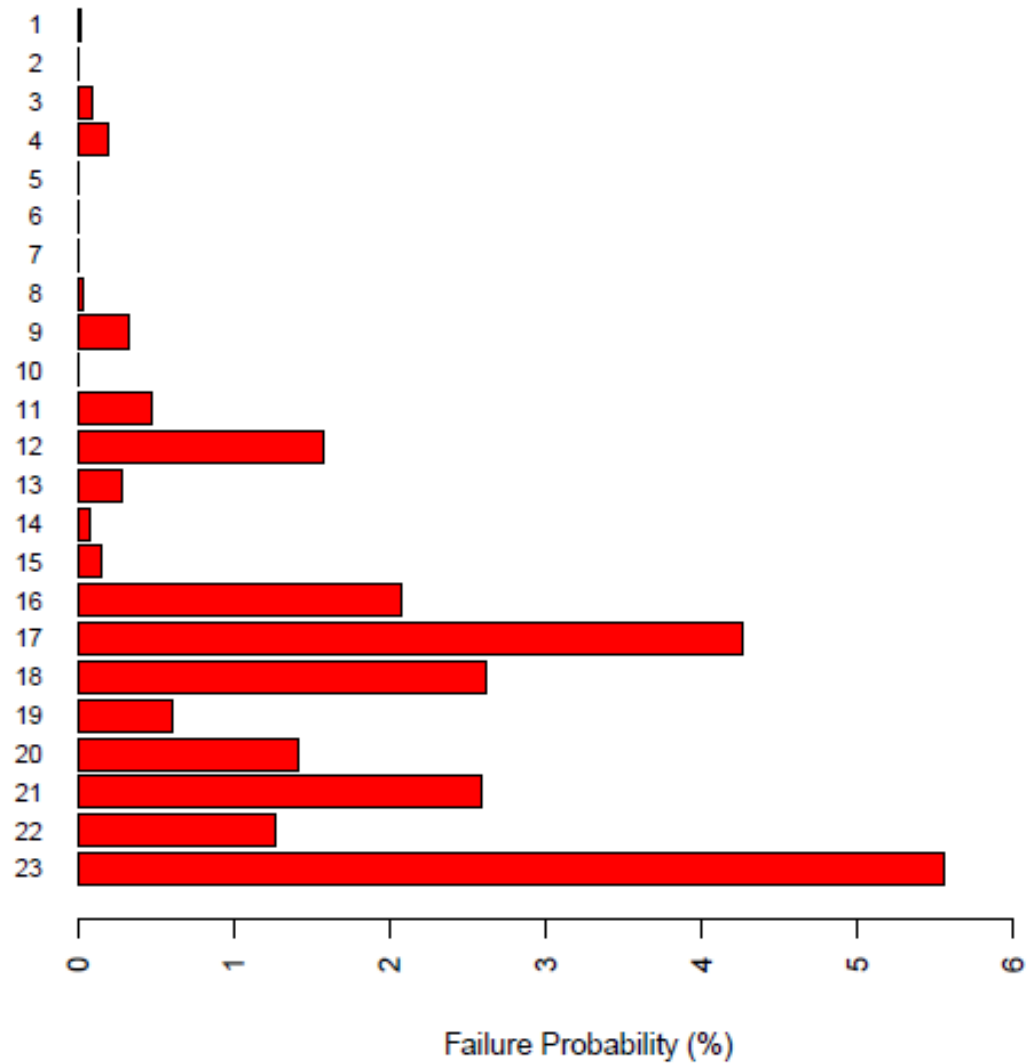
Todos os Acessos em URLs que apresentaram pelo menos um HTTP 500 (78,56% do total)



Acessos que retornaram HTTP 500



Probabilidade de falha por HTTP 500



Conclusões que podemos tirar dos logs

- Pouco acesso e alta probabilidade de falha – os testes não foram feitos adequadamente
 - Falhas grosseiras
 - Teste mais abrangente
- Muitas falhas, mas probabilidade de falhas baixa – Problemas específicos que devem ser procurados
 - Falhas específicas
 - Teste mais específico

Resultados e Trabalhos Futuros

- Pacote para limpar e transformar o arquivo de log de acesso para data frame R publicado no CRAN (ApacheLogProcessor) - LGPL
- A fazer (e precisando de alunos):
 - Análise dos parâmetros passados via GET (quais parâmetros exercitam as falhas)
 - Análise dos logs de erro para identificação dos defeitos
 - Artigo

Dúvidas, sugestões?

Contato:

diogo.mendonca@cefet-rj.br