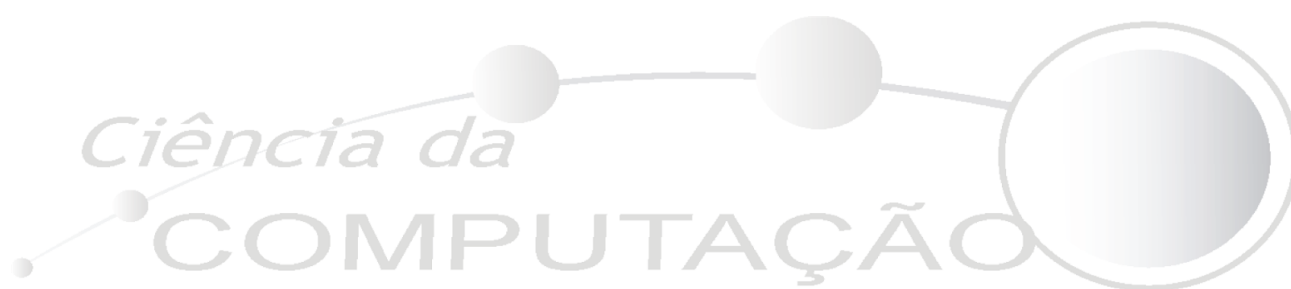


Relatório do Trabalho de Simulação

Processamento Paralelo - 2º Sem. 2017

Rogério Eduardo Pereira
Hurá Bittencourt Rocha e Silva
Igor de Souza Leal



Sumário

1. Introdução
2. Funcionamento do Caso de Uso
3. Nomenclaturas
4. Descrição do modelo físico
 - 4.1. Logs de Acesso - Registro em imagens e seu funcionamento
 - 4.2. Logs de Cliques - Registro em imagens e seu funcionamento
5. Fitness
6. Descrição do modelo lógico
7. Perguntas pertinentes
8. Estratégia de execução
9. Respostas às perguntas
 - 9.1. Pergunta 1
 - 9.2. Pergunta 2
 - 9.3. Pergunta 3
 - 9.4. Pergunta 4
 - 9.5. Pergunta 5
10. Conclusão

Introdução

Este trabalho tem por objetivo simular um sistema de gravação de logs de impressões (acessos) e cliques em um serviço de distribuição de anúncios. O sistema roda em um servidor provisionado pela Amazon, chamado EC2, que contém todo o código responsável pela entrega do anúncio.

Funcionamento do caso de uso

Quando uma pessoa acessa um site que entrega nossos anúncios, um script, que está no site acessado, chama um arquivo php que está no servidor EC2, que então monta o anúncio para o usuário visualizar. Quando o anúncio é montado, um log de impressão (acesso) é gerado.

Um outro serviço da Amazon, chamado Kinesis, funciona como uma fila. Ele armazena cada log gerado. Após essa fila atingir o tamanho de 5MB, o Kinesis gera um arquivo de logs e salva em um banco de dados chamado S3, também da Amazon.

Em nosso caso de uso, teremos duas situações tanto para a geração de logs na fila Kinesis, como para o armazenamento no banco de dados S3:

- Para uma impressão (acesso). Quando um usuário acessa um site e visualiza um anúncio.
- Para um clique. Quando um usuário acessa um site, visualiza um anúncio e resolve clicar no anúncio.

Nomenclaturas

Para um melhor entendimento e facilidade, utilizaremos as seguintes nomenclaturas para alguns termos:

- Kinesis *utilizaremos ->* Buffer
- S3 *utilizaremos ->* Banco de dados
- Impressões *utilizaremos ->* Acesso

Descrição do modelo Físico

Quando uma pessoa acessa um site que é cliente do serviço de anúncios, é disparado um anúncio para esse cliente. A taxa de visualização de anúncios é em média 80 visualizações por segundo. É gerado um log em buffer de visualização contendo:

- IP do Usuário
- Timestamp
- Sistema Operacional
- Navegador
- URL
- Agente do Usuário
Contendo informações dos usuários, sistema operacional, etc...
- ID Unitário
- URL de Redirecionamento
- ID da palavra
- Ação
- Plataforma
- ID Afiliado

Cada log de visualização tem em média o tamanho de 4KB. Ao acumular a quantidade de 5MB de logs no buffer é gerado um arquivo contendo todos os logs e então, salvos no banco de dados.

Quando um cliente clica em um anúncio, o clique gera um novo log em buffer contendo as mesmas informações do log de acesso mudando apenas o parâmetro *action*. O processo de salvar o log do buffer de cliques no banco de dados segue a mesma lógica operacional do processo de gravação de logs de acesso.

Cada clique segue uma taxa de 2,3% em proporção às visualizações.

Em um período de 5 minutos foram coletados 24.680 logs de visualizações e 573 logs de cliques.

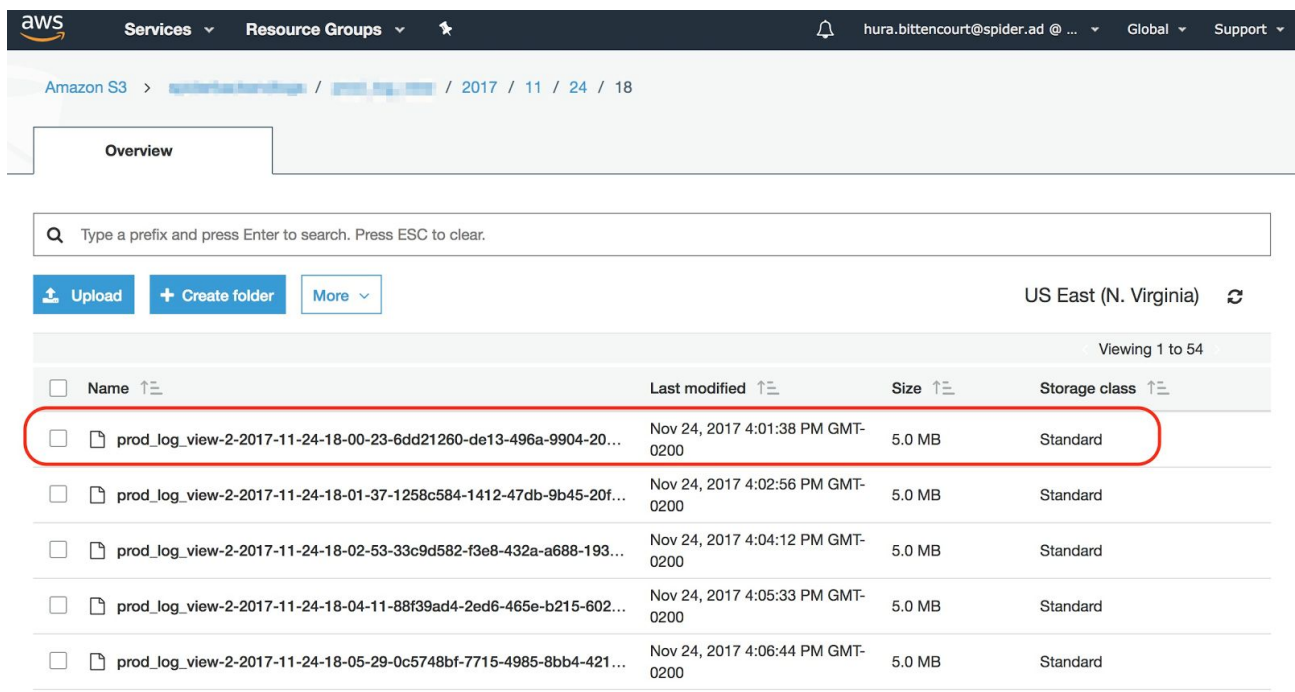
Para salvar um log de 5 MB em disco, leva em média o tempo de 4 segundos.

O arquivo de logs pode ser acessado pelo seguinte link:

[Planilha de Visualização e Cliques](#)

Logs de Acesso - Registros em imagens e seu funcionamento no projeto

Abaixo temos a imagem do serviço da Amazon que salva no banco de dados o arquivo de logs de acesso. O arquivo que está marcado em vermelho contém 5 MB de logs de acesso, na data de 24/11/2017 às 15:01:38 (horário de Brasília), que foi o período de tempo utilizado para estudo e análise deste trabalho:

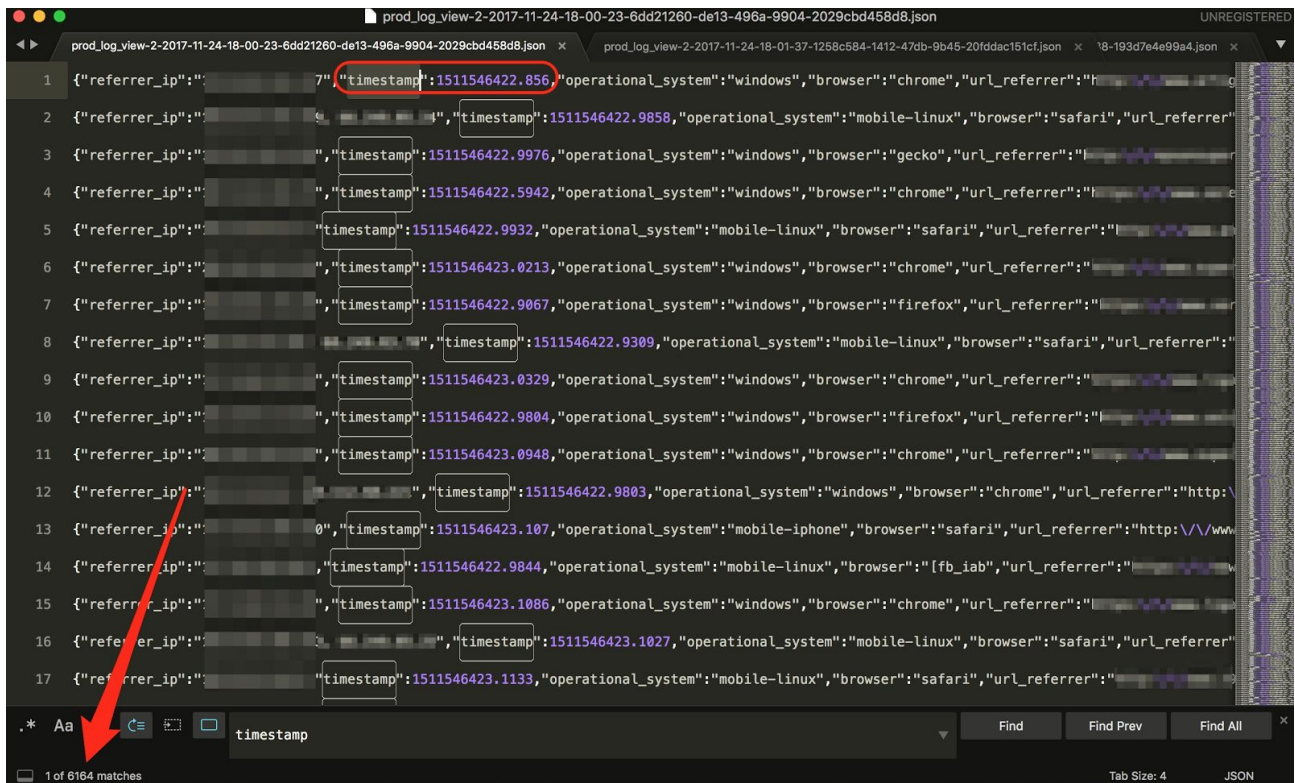


The screenshot displays the AWS S3 console interface. At the top, the navigation bar includes the AWS logo, 'Services', 'Resource Groups', and user information. The breadcrumb trail indicates the path: Amazon S3 > [bucket name] > 2017 / 11 / 24 / 18. The 'Overview' tab is selected. Below the search bar, there are buttons for 'Upload', 'Create folder', and 'More'. The region is set to 'US East (N. Virginia)'. The table lists objects with the following columns: Name, Last modified, Size, and Storage class. The first object is highlighted with a red box.

Name	Last modified	Size	Storage class
prod_log_view-2-2017-11-24-18-00-23-6dd21260-de13-496a-9904-20...	Nov 24, 2017 4:01:38 PM GMT-0200	5.0 MB	Standard
prod_log_view-2-2017-11-24-18-01-37-1258c584-1412-47db-9b45-20f...	Nov 24, 2017 4:02:56 PM GMT-0200	5.0 MB	Standard
prod_log_view-2-2017-11-24-18-02-53-33c9d582-f3e8-432a-a688-193...	Nov 24, 2017 4:04:12 PM GMT-0200	5.0 MB	Standard
prod_log_view-2-2017-11-24-18-04-11-88f39ad4-2ed6-465e-b215-602...	Nov 24, 2017 4:05:33 PM GMT-0200	5.0 MB	Standard
prod_log_view-2-2017-11-24-18-05-29-0c5748bf-7715-4985-8bb4-421...	Nov 24, 2017 4:06:44 PM GMT-0200	5.0 MB	Standard

1. Painel da Amazon - Registro do arquivo de logs de Acesso com 5MB de tamanho.

Abaixo ilustramos como é o arquivo do log de acesso no formato json. Para efeito do projeto, selecionamos somente o parâmetro *timestamp* para contabilizar quantas ocorrências tiveram. Com isso, conseguimos verificar quantos logs em média temos em um arquivo de 5 MB de log, no período de tempo mencionado anteriormente, utilizado no estudo.

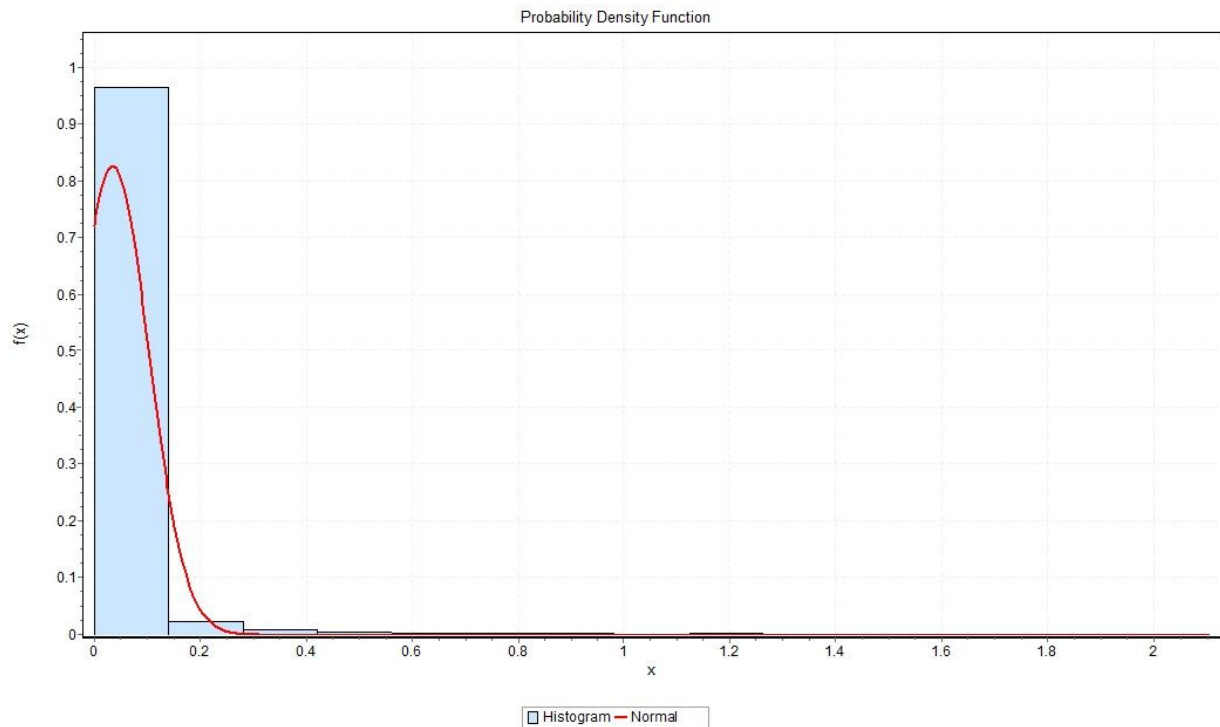


```
1 {"referrer_ip": "...", "timestamp": 1511546422.856, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
2 {"referrer_ip": "...", "timestamp": 1511546422.9858, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
3 {"referrer_ip": "...", "timestamp": 1511546422.9976, "operational_system": "windows", "browser": "gecko", "url_referrer": "..."}
4 {"referrer_ip": "...", "timestamp": 1511546422.5942, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
5 {"referrer_ip": "...", "timestamp": 1511546422.9932, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
6 {"referrer_ip": "...", "timestamp": 1511546423.0213, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
7 {"referrer_ip": "...", "timestamp": 1511546422.9067, "operational_system": "windows", "browser": "firefox", "url_referrer": "..."}
8 {"referrer_ip": "...", "timestamp": 1511546422.9309, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
9 {"referrer_ip": "...", "timestamp": 1511546423.0329, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
10 {"referrer_ip": "...", "timestamp": 1511546422.9804, "operational_system": "windows", "browser": "firefox", "url_referrer": "..."}
11 {"referrer_ip": "...", "timestamp": 1511546423.0948, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
12 {"referrer_ip": "...", "timestamp": 1511546422.9803, "operational_system": "windows", "browser": "chrome", "url_referrer": "http://..."}
13 {"referrer_ip": "...", "timestamp": 1511546423.107, "operational_system": "mobile-iphone", "browser": "safari", "url_referrer": "http://www..."}
14 {"referrer_ip": "...", "timestamp": 1511546422.9844, "operational_system": "mobile-linux", "browser": "[fb_iab]", "url_referrer": "..."}
15 {"referrer_ip": "...", "timestamp": 1511546423.1086, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
16 {"referrer_ip": "...", "timestamp": 1511546423.1027, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
17 {"referrer_ip": "...", "timestamp": 1511546423.1133, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}

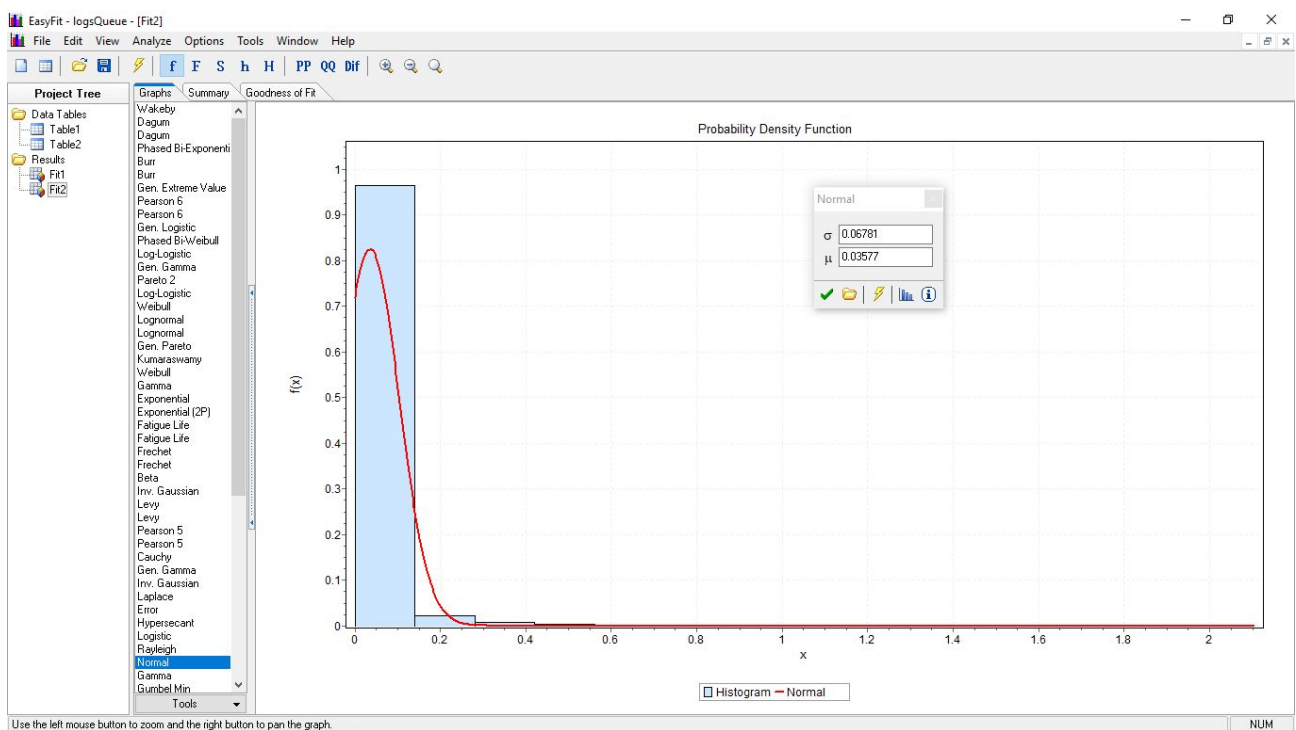
.* Aa timestamp
1 of 6164 matches
```

2. Editor de texto *Sublime* - Acesso a um arquivo de logs de Acesso.

Para calcular a frequência em que cada log de acesso é gerado, ou seja, pessoas que acessam sites que entregam anúncios nosso, utilizamos a média normal, calculada tanto no software *easyfit* quanto no próprio software de planilhas da Google. Na imagem abaixo conseguimos verificar o resultado do programa *easyfit* em relação a média normal:

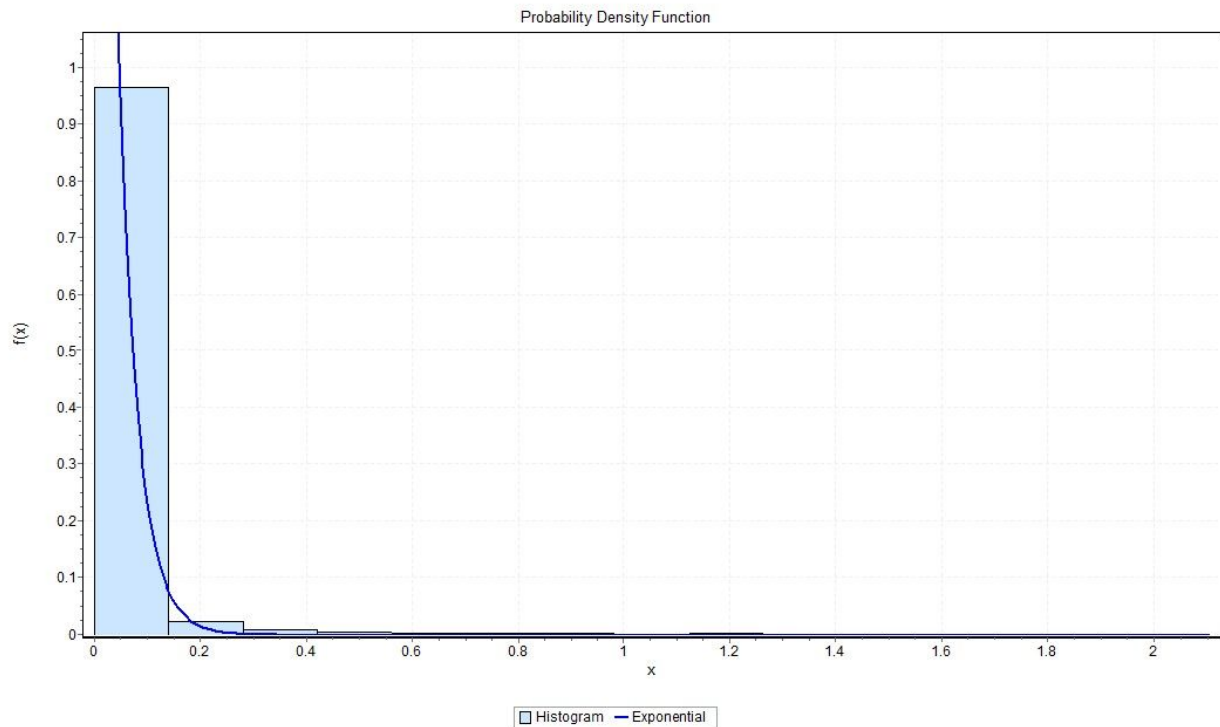


3. Software *easyfit* - Gráfico da média normal dos logs de Acesso.

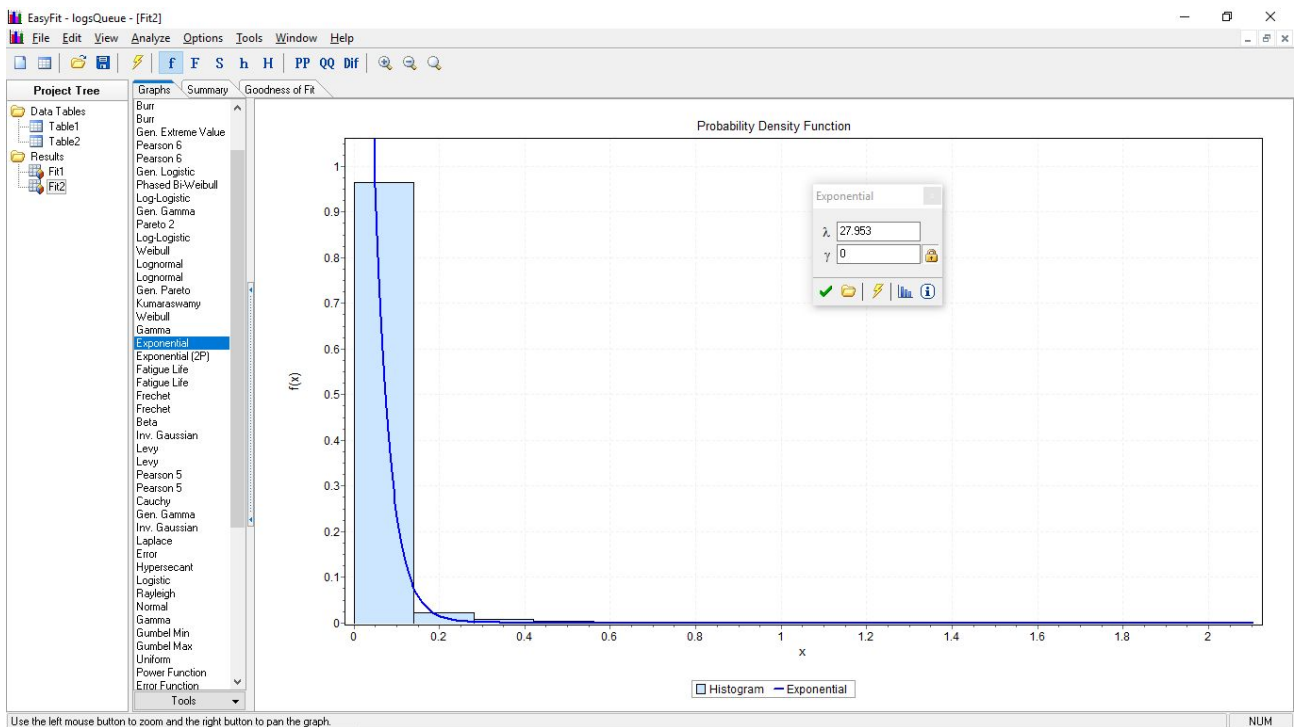


4. Software *easyfit* - Print mostrando resultado da média normal dos logs de Acesso.

Tentamos utilizar a média exponencial no software *easyfit* porém, não obtivemos sucesso no caso dos logs de acesso. Acreditamos que o valor resultante da diferença entre os *timestamp* de cada log era pequeno demais para o cálculo. Abaixo temos as imagens ilustrando os resultados:



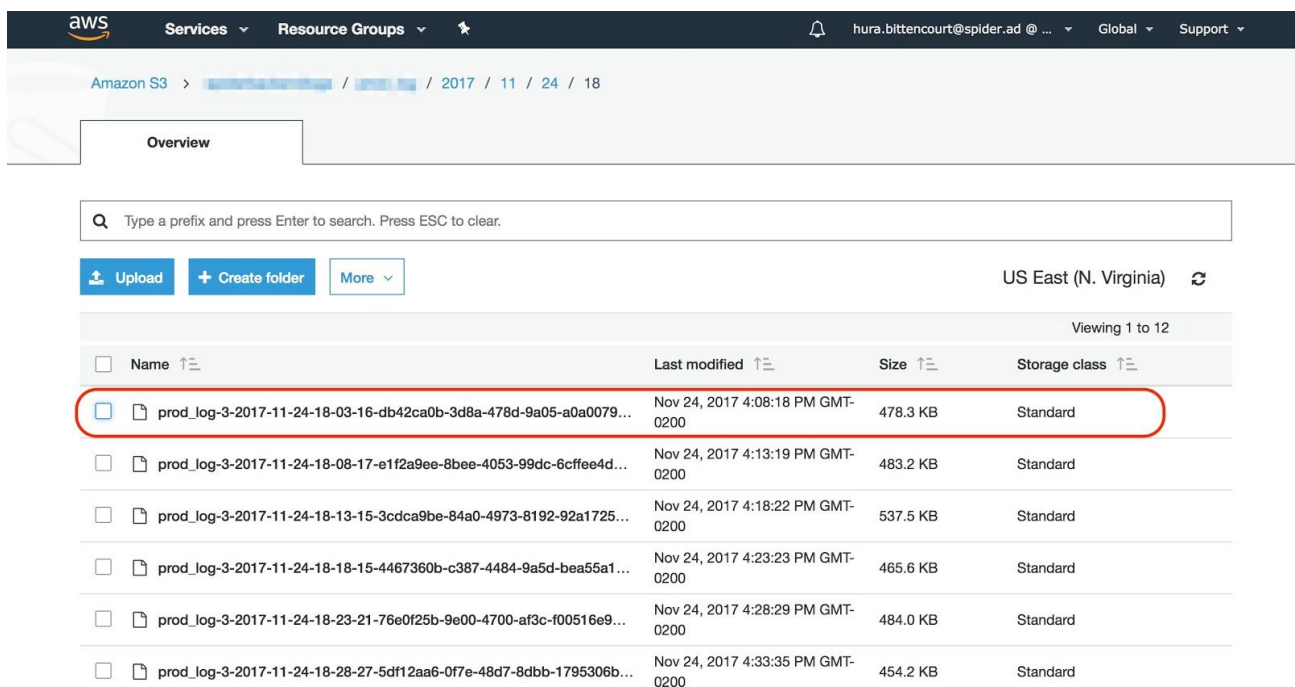
5. Software *easyfit* - Gráfico da média exponencial dos logs de Acesso.



6. Software *easyfit* - Print mostrando resultado da média exponencial dos logs de Acesso.

Logs de Clique - Registros em imagens e seu funcionamento no projeto

Abaixo temos a imagem do serviço da Amazon que salva no banco de dados o arquivo de logs de clique. O arquivo que está marcado em vermelho contém 478.3 KB de logs de clique, uma vez que, no caso do buffer não atingir os 5 MB, a cada 5 minutos ele salva no banco de dados o que tiver no buffer. O registro foi na data de 24/11/2017 às 15:08:18 (horário de Brasília), utilizado para estudo e análise dos logs de clique desse trabalho:

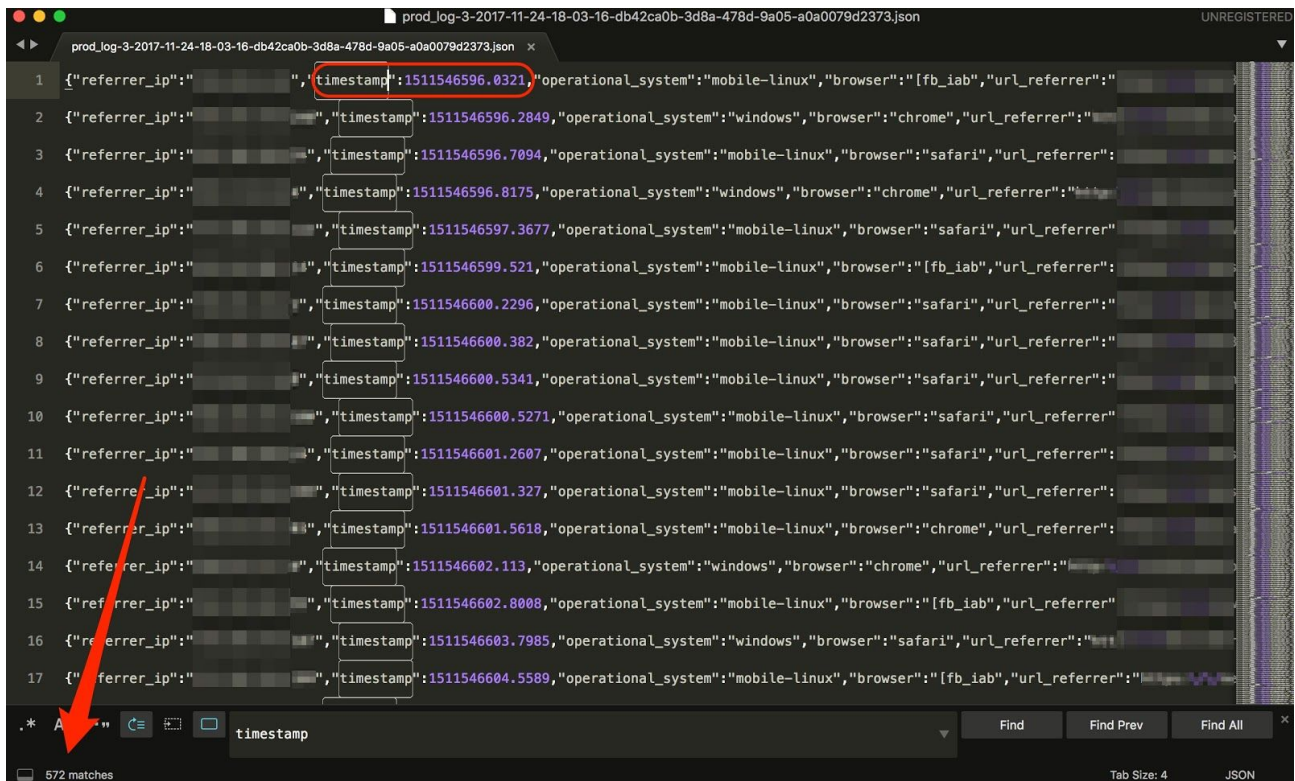


The screenshot shows the AWS S3 console interface. At the top, there's a navigation bar with the AWS logo, 'Services', 'Resource Groups', and a user profile. Below this, a breadcrumb trail indicates the current location: 'Amazon S3 > [bucket name] / [folder name] / 2017 / 11 / 24 / 18'. A tab labeled 'Overview' is selected. A search bar is present with the placeholder text 'Type a prefix and press Enter to search. Press ESC to clear.' Below the search bar, there are buttons for 'Upload', 'Create folder', and 'More'. The region 'US East (N. Virginia)' is displayed. The main content area shows a table of objects. The first row is highlighted with a red box. The table has columns for 'Name', 'Last modified', 'Size', and 'Storage class'.

Name	Last modified	Size	Storage class
prod_log-3-2017-11-24-18-03-16-db42ca0b-3d8a-478d-9a05-a0a0079...	Nov 24, 2017 4:08:18 PM GMT-0200	478.3 KB	Standard
prod_log-3-2017-11-24-18-08-17-e1f2a9ee-8bee-4053-99dc-6cfee4d...	Nov 24, 2017 4:13:19 PM GMT-0200	483.2 KB	Standard
prod_log-3-2017-11-24-18-13-15-3cdca9be-84a0-4973-8192-92a1725...	Nov 24, 2017 4:18:22 PM GMT-0200	537.5 KB	Standard
prod_log-3-2017-11-24-18-15-4467360b-c387-4484-9a5d-bea55a1...	Nov 24, 2017 4:23:23 PM GMT-0200	465.6 KB	Standard
prod_log-3-2017-11-24-18-23-21-76e0f25b-9e00-4700-af3c-f00516e9...	Nov 24, 2017 4:28:29 PM GMT-0200	484.0 KB	Standard
prod_log-3-2017-11-24-18-28-27-5df12aa6-0f7e-48d7-8dbb-1795306b...	Nov 24, 2017 4:33:35 PM GMT-0200	454.2 KB	Standard

7. Painel da Amazon - Registro do arquivo de logs de Clique.

Abaixo ilustramos como é o arquivo do log de acesso no formato json. Para efeito do projeto, selecionamos somente o parâmetro *timestamp* para contabilizar quantas ocorrências tiveram. Com isso, conseguimos verificar quantos logs em média temos em um arquivo de 5 MB de log, no período de tempo mencionado anteriormente, utilizado no estudo.

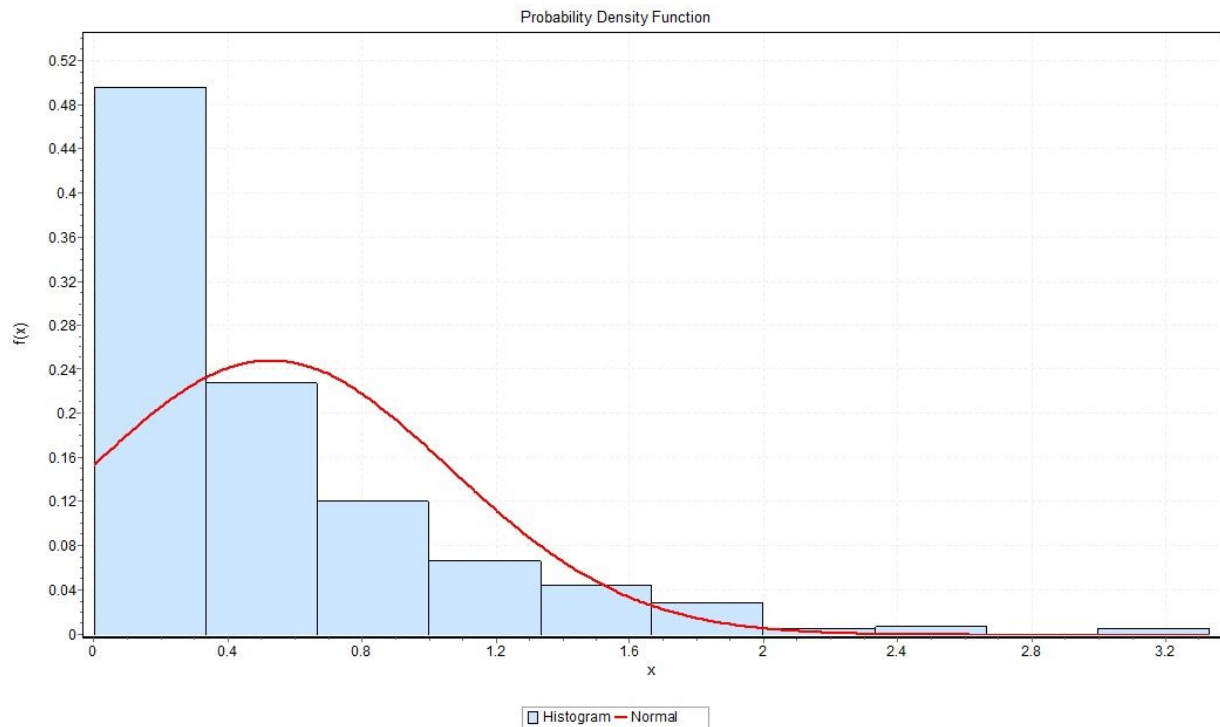


```
1 {"referrer_ip": "...", "timestamp": 1511546596.0321, "operational_system": "mobile-linux", "browser": "[fb_iab", "url_referrer": "..."}
2 {"referrer_ip": "...", "timestamp": 1511546596.2849, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
3 {"referrer_ip": "...", "timestamp": 1511546596.7094, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
4 {"referrer_ip": "...", "timestamp": 1511546596.8175, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
5 {"referrer_ip": "...", "timestamp": 1511546597.3677, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
6 {"referrer_ip": "...", "timestamp": 1511546599.521, "operational_system": "mobile-linux", "browser": "[fb_iab", "url_referrer": "..."}
7 {"referrer_ip": "...", "timestamp": 1511546600.2296, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
8 {"referrer_ip": "...", "timestamp": 1511546600.382, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
9 {"referrer_ip": "...", "timestamp": 1511546600.5341, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
10 {"referrer_ip": "...", "timestamp": 1511546600.5271, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
11 {"referrer_ip": "...", "timestamp": 1511546601.2607, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
12 {"referrer_ip": "...", "timestamp": 1511546601.327, "operational_system": "mobile-linux", "browser": "safari", "url_referrer": "..."}
13 {"referrer_ip": "...", "timestamp": 1511546601.5618, "operational_system": "mobile-linux", "browser": "chrome", "url_referrer": "..."}
14 {"referrer_ip": "...", "timestamp": 1511546602.113, "operational_system": "windows", "browser": "chrome", "url_referrer": "..."}
15 {"referrer_ip": "...", "timestamp": 1511546602.8008, "operational_system": "mobile-linux", "browser": "[fb_iab", "url_referrer": "..."}
16 {"referrer_ip": "...", "timestamp": 1511546603.7985, "operational_system": "windows", "browser": "safari", "url_referrer": "..."}
17 {"referrer_ip": "...", "timestamp": 1511546604.5589, "operational_system": "mobile-linux", "browser": "[fb_iab", "url_referrer": "..."}

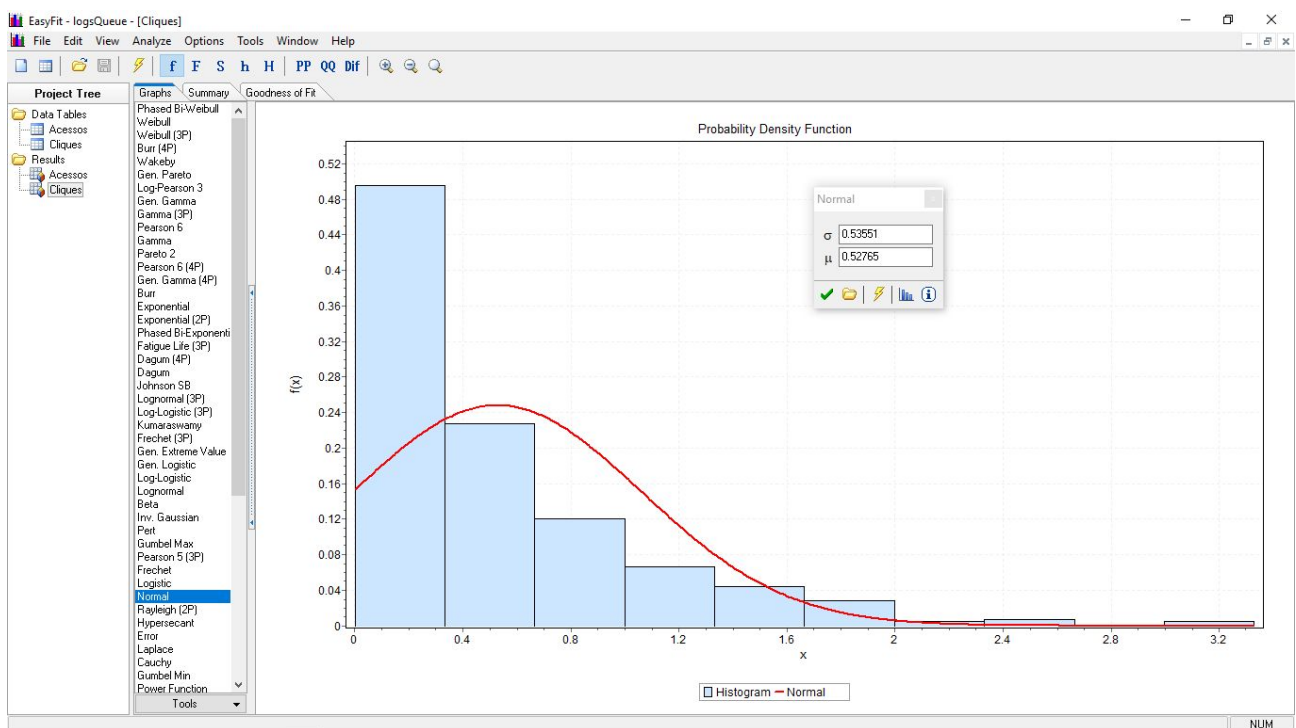
.* A timestamp
572 matches
Tab Size: 4
JSON
```

8. Editor de texto *Sublime* - Acesso a um arquivo de logs de Clique.

Para calcular a frequência em que cada log de clique é gerado, ou seja, pessoas que acessam sites que entregam anúncios nosso e clicam nesse anúncio, utilizamos a média normal, calculada tanto no software *easyfit* quanto no próprio software de planilhas da Google. Abaixo podemos verificar o resultado do programa *easyfit* em relação a média normal:

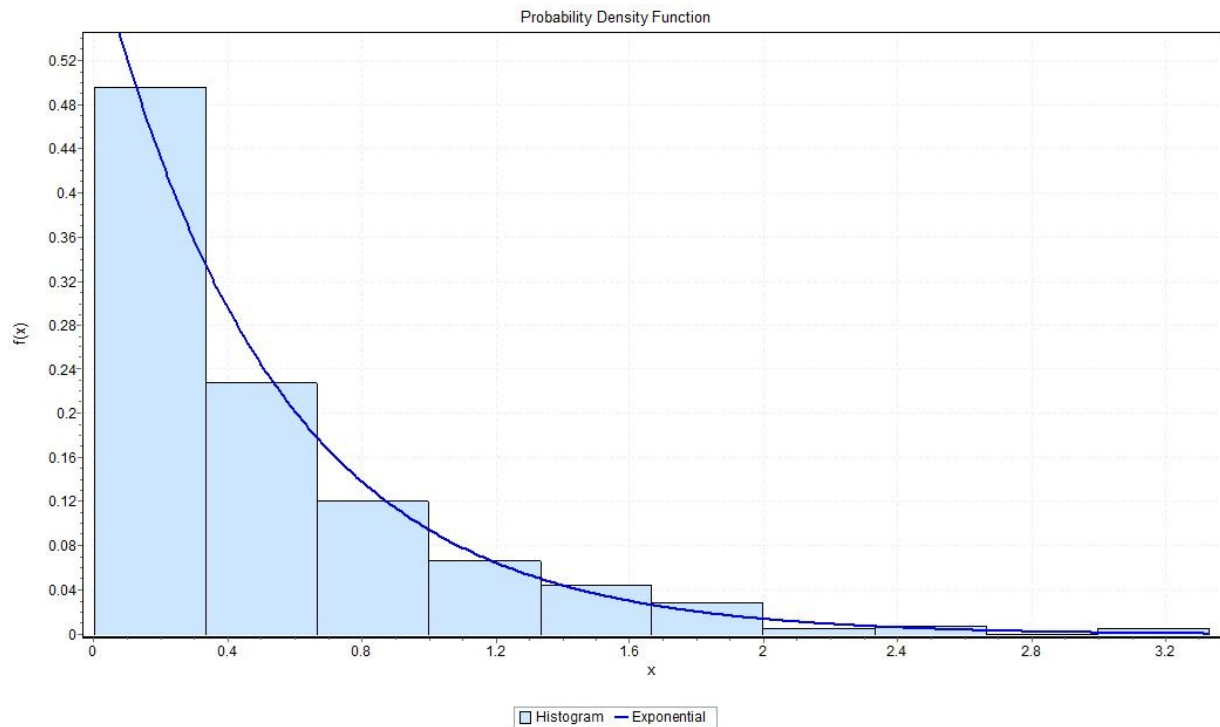


9. Software *easyfit* - Gráfico da média normal dos logs de Clique.

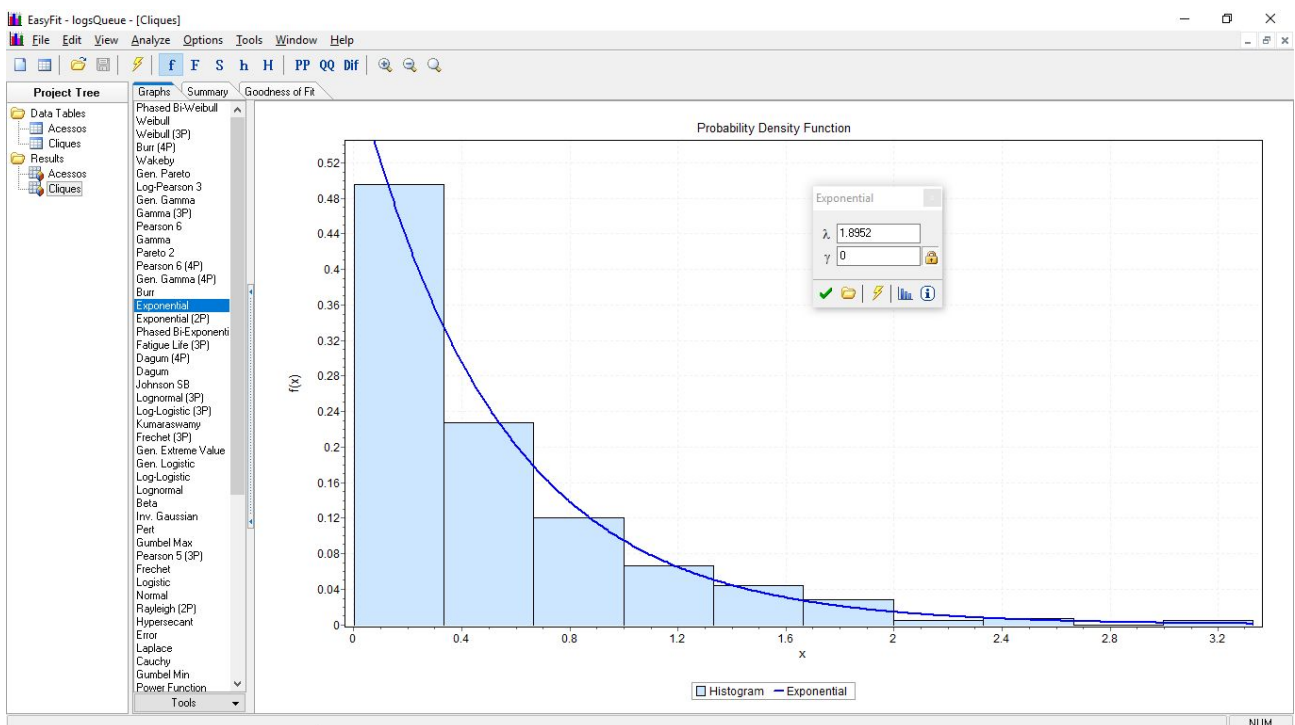


10. Software *easyfit* - Print mostrando resultado da média normal dos logs de Clique.

Tentamos utilizar a média exponencial no software *easyfit* também para o caso de logs de clique porém, não obtivemos sucesso. Acreditamos que o valor resultante da diferença entre os *timestamp* de cada log era pequeno demais para o cálculo. Abaixo temos as imagens ilustrando os resultados:



11. Software *easyfit* - Gráfico da média exponencial dos logs de Clique.

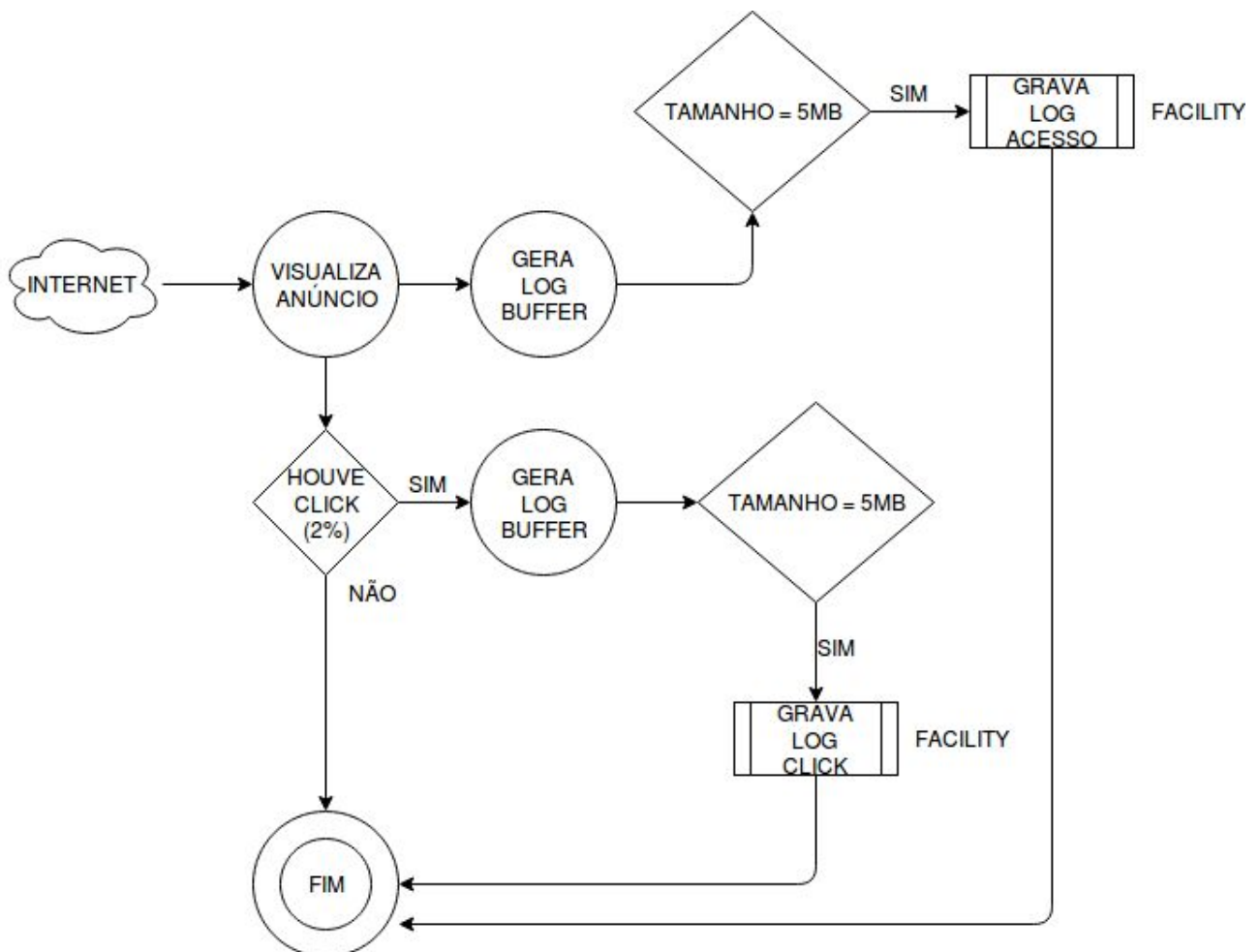


12. Software *easyfit* - Print mostrando resultado da média exponencial dos logs de Clique.

Fitness

O resultado completo do software *easyfit*, tanto do Summary como do Goodness of Fit podem ser encontrados no arquivo relatorios/easyfitness.html que foi enviado com os arquivos do trabalho.

Descrição do modelo Lógico



13. Modelo lógico do Caso de Uso utilizado.

Perguntas Pertinentes

Durante a simulação tínhamos por objetivo responder às seguintes questões:

- O desempenho seria melhor em caso de eliminar o buffer e salvar diretamente no banco de dados?
- Qual seria o melhor tamanho de buffer para a quantidade atual de acessos?
- Se a taxa de acessos aumentar consideravelmente qual seria o melhor tamanho do buffer a ser salvo?
- Se a taxa de cliques aumentar consideravelmente qual seria o melhor tamanho do buffer a ser salvo.
- Se, caso desse problema com a velocidade de gravação do disco, seria muito prejudicado o desempenho geral?

Estratégia de Execução

Foi definido como uma facility a ação de salvar o log no banco de dados e não da quantidade de logs no buffer.

Para responder a questão 1 e 2 é preciso alterar o valor da variável *qtdParaGravarLog* (linha20). Se for eliminar o buffer alterar para 1, se for simular o melhor desempenho do buffer alterar para o valor correspondente à quantidade de logs no buffer.

Para responder a questão 3 é preciso diminuir ou aumentar o valor da variável *tempoAcessoSite* (linha 11) (valor entre 0 e 100), e a variável *qtdParaGravarLog* (linha20), conforme explicado para as questões 1 e 2.

Para responder a questão 4 é preciso diminuir ou aumentar o valor da variável *taxaCliques* (linha 19) (quanto menor o valor, maior a taxa de acesso), e a variável *qtdParaGravarLog* (linha20), conforme explicado para as questões 1 e 2.

Para responder a questão 5 é preciso alterar o valor da variável *tempoGravarLog* (linha 16) (valor em segundos)

Respostas às Perguntas

Pergunta 1

Não, o desempenho é pior quando elimina o buffer.

Pergunta 2

O buffer com o tamanho de 5MB tem um desempenho aceitável, ocupando 26% do tempo do servidor. Se aumentasse o valor do buffer, o tempo ocupado com o processo de gravação dos logs diminuiria, mas em contrapartida, o desempenho do servidor poderia ser prejudicado, uma vez que estaria gravando um log de 10MB, podendo causar uma lentidão de respostas devido a rede ocupada com a transmissão dos dados.

Pergunta 3

Dobrando a quantidade de acesso o tempo de uso do servidor dobra, sendo necessário reduzir pela metade o tamanho do buffer para manter o mesmo desempenho.

Pergunta 4

Atualmente o tempo utilizado pelo servidor para gravar a taxa de cliques é de 0,6%, para começar a prejudicar o desempenho do servidor seria necessário uma taxa de cliques acima de 38%, causando um aumento de 10% do tempo de uso do servidor.

Pergunta 5

A relação entre o tempo de gravação é diretamente proporcional ao desempenho, dobrando o tempo, o tempo ocupado do servidor dobra.

Até 12 segundos para gravação o servidor nas condições atuais não gera filas para a facility de gravação do buffer no banco de dados. Levando 13 segundo é gerado uma fila de 4 processos para ser gravados.

Aumentando para 14, a fila fica com 26 itens. Aumentando para 20 segundos, a fila atinge o tamanho de 179 itens a serem gravados.

Conclusão

Podemos concluir que nesse caso do uso, o serviço se encontra bem otimizado. Tanto o tamanho de armazenamento do buffer quanto o tempo para se gravar em disco no banco de dados, perante a quantidade de pessoas que acessam sites com nossos anúncios, são ideais para a situação atual.