

Piwik – Uma alternativa livre ao Google Analytics

Éverton Didoné Foscarini¹ Alexandre Albino Marchi¹

¹Centro de Processamento de Dados – Universidade
Federal do Rio Grande do Sul (UFRGS)
Rua Ramiro Barcelos, 2574 – Portão K – Porto Alegre – RS – Brazil
{foscarini,marchi}@cpd.ufrgs.br

Resumo. Este artigo apresenta o Piwik, um software open source de análise de acessos web. São cobertos aspectos relativos ao seu funcionamento, configurações importantes disponíveis no programa, uma solução para a utilização em sites via HTTPS e um caso de sucesso, referente à contabilização de acessos durante a divulgação do resultado do concurso vestibular da UFRGS.

1. Introdução

O Piwik [Piwik] é um software *open source* de análise de acessos web. Ele assemelha-se ao *Google Analytics* por utilizar o mesmo princípio de funcionamento e por disponibilizar as estatísticas de forma gráfica em uma interface web.

O Piwik é implementado em PHP e pode ser instalado em um servidor Apache+MySQL [Apache] [Mysql] dentro da instituição de forma a manter as estatísticas referentes ao acesso das páginas web em uma base de dados privada.

Uma instalação do Piwik pode gerar estatísticas de dezenas e até centenas de sites dentro da instituição, podendo separá-los em diferentes perfis de acordo com o *VirtualHost* ou no caso de serem armazenados em servidores web diferentes.

2. Instalação e utilização

A instalação do software é trivial, bastante semelhante à qualquer aplicativo PHP+MySQL. É necessário criar uma base de dados e configurar o Piwik e configurá-lo através do seu assistente de instalação web.

A análise do acesso web é feita através da inclusão de código JavaScript em todas as páginas de um determinado site ou no seu *template*, caso exista. Segue exemplo de código gerado pelo Piwik para inclusão em uma página:

```
<!-- Piwik -->
<a href="http://piwik.org" title="Analytics software"
onclick="window.open(this.href);return(false);">
<script type="text/javascript">
  var pkBaseURL = (("https:" == document.location.protocol) ?
    "https://stats.ufrgs.br/piwik/" : "http://stats.ufrgs.br/piwik/");
  document.write(unescape("%3Cscript src='" + pkBaseURL + "piwik.js'
    type='text/javascript'%3E%3C/script%3E"));
</script>
<script type="text/javascript">
  piwik_action_name = '';
  piwik_idsite = 6;
  piwik_url = pkBaseURL + "piwik.php";
  piwik_log(piwik_action_name, piwik_idsite, piwik_url);
</script>
<object>
  <noscript>
    <p>Analytics software </p>
```

```
</noscript>
</object>
</a>
<!-- End Piwik Tag -->
```

Quando o *Browser* de um usuário exibe uma página que contém o código do *Piwik* ele acaba executando um código *JavaScript*, que realiza o *download* de um arquivo armazenado no servidor de estatísticas. O *download* desse arquivo gera entrada no banco de dados com as informações do *Browser* e da configuração do ambiente de trabalho do usuário. A figura 1 ilustra o acesso à página da UFRGS, que gera requisições para dois servidores simultaneamente, o *www.ufrgs.br* que contém a página solicitada pelo usuário e o *stats.ufrgs.br*, que vai coletar as estatísticas do acesso.

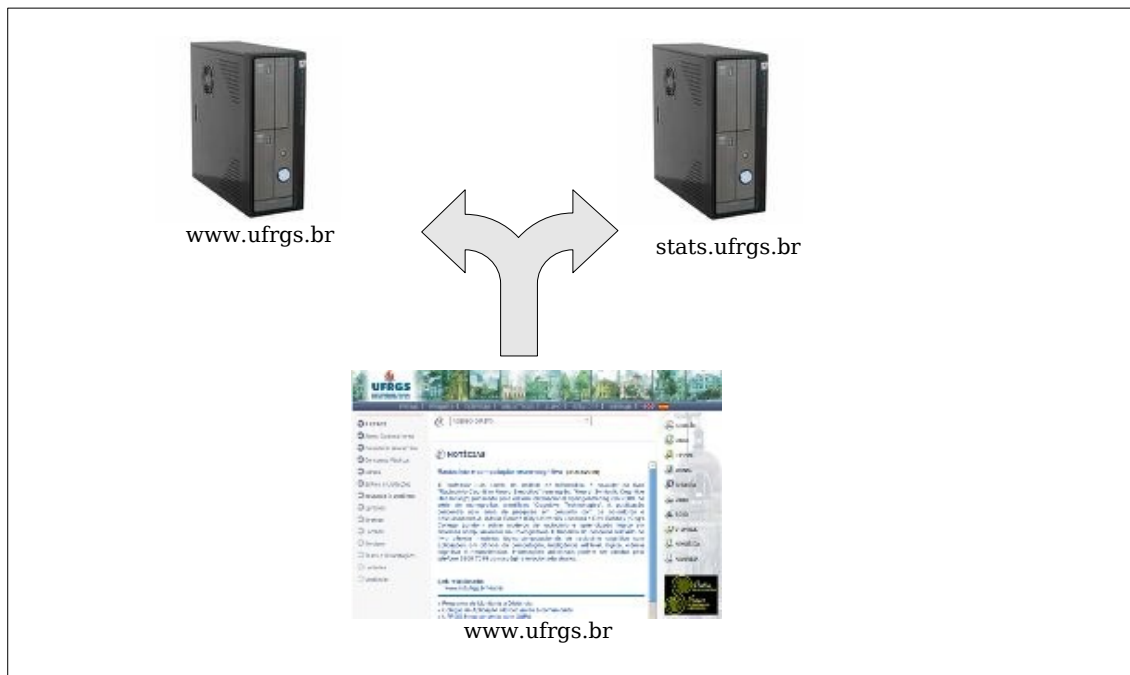


Figura 1: Acesso ao servidor *web* e ao servidor de estatísticas

Através do acesso realizado ao servidor de estatísticas as seguintes informações são coletadas pelo *Piwik*:

- Configurações do computador do usuário
 - Resolução de tela
 - Software de navegação/versão utilizada
 - Sistema Operacional utilizado
 - *Plugins* instalados no navegador
- Informações do acesso
 - Horário da visita
 - Localização do cliente (a partir do endereço IP)
 - Duração da visita e páginas visitadas
 - *Sites* que referenciaram a URL (que geraram tráfego para o *site*)

- Lista de arquivos baixados do *site*
- Sistemas de busca *web* que mais geraram visitas
- Palavras chave utilizadas em sistemas de busca que levaram até o *site*

Essas informações podem ser visualizadas em relatórios com dados diários, mensais e anuais de acesso, permitindo analisar por exemplo quais são os horários de maior e menor acesso a um *site*, a sazonalidade dos acessos, a configuração dos computadores dos usuários e quais são as páginas mais acessadas.



Figura 2: Página inicial do *Piwik*

A figura 2 apresenta a tela inicial do *Piwik*, onde é possível visualizar um resumo das estatísticas de um perfil de *site* analisado.

3. Configurações e otimizações

Determinadas configurações podem ser feitas no arquivo "*config/global.ini.php*" para otimizar o *Piwik*. As mais relevantes são as apresentadas à seguir:

```
[General]
time_before_archive_considered_outdated = 300
```

Tempo em segundos que o *Piwik* aguarda antes de analisar o banco de dados e gerar estatísticas para visualização. Ao usar um valor pequeno (10 por exemplo), o *Piwik* gera estatísticas de acesso em tempo real. Valores pequenos aumentam bastante a carga do servidor, pois é preciso fazer consultas constantes ao banco de dados para gerar as novas estatísticas.

```
[General]
default_day = today
```

Especifica a data de referência do gráfico que é apresentado na tela inicial no sistema. A configuração padrão apresenta os valores do dia anterior.

```
[Tracker]
record_statistics= 1
```

Habilita a gravação das estatísticas no banco de dados. Pode ser mudado para “0” de forma a possibilitar a atualização do sistema ou uma manutenção programada do banco de dados. O Piwik deixará de fazer conexões com o banco de dados, diminuindo a carga do servidor.

4. Estatísticas de páginas que utilizam HTTPS

A instalação do *Piwik* em um servidor exclusivo permite a centralização das estatísticas da instituição, porém também cria um problema de configuração para sites que utilizam o HTTPS. Caso o acesso à um serviço (*webmail* por exemplo) seja feito via HTTPS enquanto o acesso ao servidor de estatísticas é feito via HTTP, os usuários recebem alertas de que alguns itens da página não foram carregados por uma conexão segura, gerando dúvidas e possivelmente reclamações dos usuários.

A utilização de um certificado auto-assinado (ou assinado por uma autoridade local) obrigaria que todos os usuários adicionassem o certificado do servidor de estatísticas na lista de certificados confiáveis de seu *Browser*, gerando também grande número de chamados de suporte e reclamações para a nossa central de atendimento.

A configuração ideal seria a utilização de um certificado SSL válido para o servidor de estatísticas, de forma a possibilitar um acesso transparente para os usuários. Entretanto sabemos que as instituições federais de ensino superior tem algumas dificuldades na obtenção de certificados SSL, de forma que foi preciso criar uma configuração diferenciada.

Foi configurado um *proxy* reverso nos servidores *web* que disponibilizam páginas através de HTTPS para os usuários. Esse *proxy* reverso mascara o acesso ao servidor de estatísticas, disponibilizando um diretório protegido por HTTPS que redireciona o acesso ao servidor de estatísticas, que coletará os dados do computador do usuário. Abaixo está o trecho do código *JavaScript* do *Piwik* com as alterações necessárias em negrito.

```
<!-- Piwik -->
<a href="http://piwik.org" title="Analytics software"
  onclick="window.open(this.href);return(false);">
  <script type="text/javascript">
    var pkBaseURL = (("https:" == document.location.protocol) ?
      "https://webmail.ufrgs.br/piwik/" :
      "http://stats.ufrgs.br/piwik/");
    document.write(unescape("%3Cscript src='" + pkBaseURL + "piwik.js'
      type='text/javascript'%3E%3C/script%3E"));
  </script>
  <script type="text/javascript">
    piwik_action_name = '';
    piwik_idsite = 14;
    piwik_url = pkBaseURL + "piwik.php";
    piwik_log(piwik_action_name, piwik_idsite, piwik_url);
  </script>
  <object>
    <noscript><p> </p>
    </noscript>
  </object>
</a>
<!-- End Piwik Tag -->
```

A configuração do *Apache* consiste na criação de um diretório no *VirtualHost* que está configurado para prover a página via HTTPS usando a diretiva *ProxyPass*:

```
<IfModule mod_proxy.c>
  <Proxy *>
    Order deny,allow
    Allow from all
  </Proxy>
  ProxyPass /piwik http://stats.ufrgs.br/piwik/
  ProxyPassReverse /piwik http://stats.ufrgs.br/piwik/
</IfModule>
```

5. Caso de uso: Divulgação dos resultados do concurso vestibular

A primeira utilização em grande escala do *Piwik* foi feito durante a divulgação dos resultados do concurso vestibular 2009 da UFRGS. O concurso teve aproximadamente 40.000 candidatos, que fizeram as provas em 4 dias, no início de janeiro. O servidor utilizado para armazenar as estatísticas tem 2 processadores *Quad Core Xeon* de 2GHz e 4GB de RAM, utilizando uma configuração do *MySQL* otimizada para bases de dados grandes.

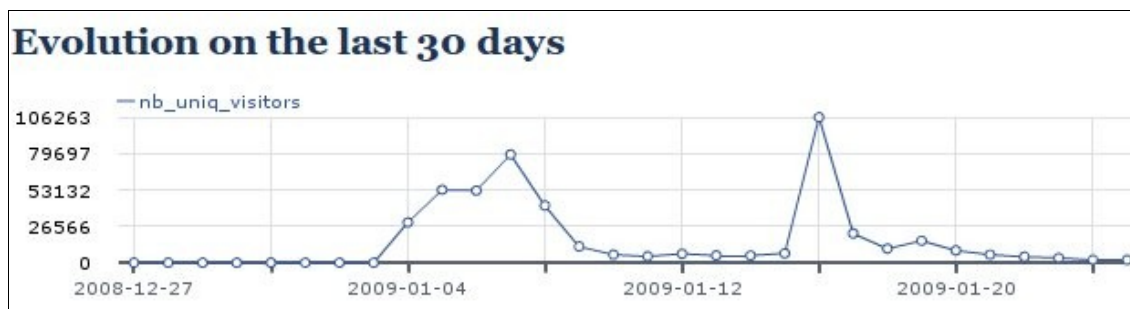


Figura 3: Acessos na página de resultados do vestibular

A figura 3 apresenta o número de acessos ao website de resultados no mês de janeiro. Nos dias de provas (04/01 a 07/01) os gabaritos das provas realizadas eram divulgados durante a tarde, gerando em média 50.000 acessos por dia com picos de 5.000 por hora. No último dia de prova a consulta dos gabaritos chegou à quase 80.000 visitas em 24 horas.

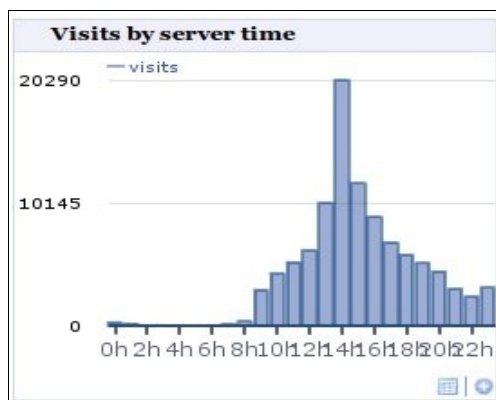


Figura 4: Momento da divulgação do listão

A divulgação do resultado final, em 16/01 gerou mais de 100.000 acessos no dia, com um pico de 20.290 acessos às 14h, momento em que foi divulgada a lista de candidatos.

6. Conclusão

O Piwik têm se demonstrado estável e relativamente escalável. No mês de março de 2009 o Piwik analisou as estatísticas de 6.497.929 pageviews, gerados por 10 sites de alto e médio tráfego hospedados nos servidores da instituição.

Estão sendo mantidos no banco de dados registros de acessos feitos desde novembro de 2008, e a base MySQL em 31/03/2009 tem atualmente perto de 39 milhões de registros, sendo 2.5Gb de dados e 830Mb de índices, ocupando 3.3Gb.

O *Piwik* está sendo usado inicialmente para analisar o acesso web dos serviços de *webmail*, ensino à distância e do sistema de revistas providos pelo CPD. A ferramenta está sendo liberada gradativamente aos usuários que tem seus *sites* hospedados nos servidores institucionais.

Referências

Piwik Web Analytics - <http://piwik.org/>

Apache Web Server - <http://httpd.apache.org/>

MySQL - <http://mysql.com/>