

# 优化和规模化 Piwik

## – 如何获得高流量性能

Piwik 系列中文文档

在 [wenku.goodu.info/piwik](http://wenku.goodu.info/piwik) 购买，可享受 3 个月技术支持服务

***InnovateDigital***

[www.InnovateDigital.com](http://www.InnovateDigital.com)

2012-9

## 目录

1. 服务器的推荐配置和技巧.....	3
2. 优化 Piwik – 配置.....	4
3. Piwik 的其他高性能建议.....	5
3.1. 设置可忽略的 URL 参数.....	5
3.2. 负载均衡.....	5
3.3. 一些经验数据.....	6
4. 如何设置自动归档.....	7
4.1. Linux/Unix: 如何设置 crontab 自动归档报告.....	7
4.2. Windows: 如何使用 Windows 任务计划设置自动归档.....	8
4.3. CPanel: 如何使用 CPanel 设置 cron 脚本.....	8
4.4. 如果网站不支持 cron 任务, 使用 Web Cron.....	9

本指南还处于不断完善中，为了使 Piwik 能支撑数以百万计的页面访问量，本文档汇总了 Piwik 权威用户提供的建议和技巧。

Piwik 可支持每月数以百万计的页面访问量，以及成千上万的注册网站。如果您的访问量是每天几百个网页访问量，Piwik 可“开箱即用”，正常工作。

如果每天有成千上万的访问量或更多，请阅读下面内容。

## 1. 服务器的推荐配置和技巧

- n 如果跟踪高流量网站，推荐将 Piwik 安装在一个专用的高性能服务器上。
- n Piwik (php/Mysql) 服务器的内存非常重要，越多越好：对于处理大型网站的报告，推荐 Piwik 的最小内存是 4Gb，RAM 越多越好！
- n 使用 PHP 缓存是高性能的关键。推荐使用 [APC](http://en.wikipedia.org/wiki/List_of_PHP_accelerators#Alternative_PHP_Cache_28APC.29)（Alternative PHP Cache）[http://en.wikipedia.org/wiki/List\\_of\\_PHP\\_accelerators#Alternative\\_PHP\\_Cache\\_28APC.29](http://en.wikipedia.org/wiki/List_of_PHP_accelerators#Alternative_PHP_Cache_28APC.29)。
- n 推荐使用轻量级 web 服务器，例如 [Nginx](#) 或 [lighttpd](#)，比 Apache 或 IIS 更快和更有效。
- n Piwik 性能严重依赖于 MySQL 服务器的性能，因为所有 Piwik 的数据都存放在 MySQL 中并使用 MySQL 查询进行汇总。建议使用 [Mysql Tuning Primer](https://launchpad.net/mysql-tuning-primer)（<https://launchpad.net/mysql-tuning-primer>）或 [Mysql Tuner](https://github.com/rackerhacker/MySQLTuner-perl#readme)（<https://github.com/rackerhacker/MySQLTuner-perl#readme>）工具优化 Mysql 设置。

## 2. 优化 Piwik – 配置

- n 默认情况下，Piwik 在 Real Time（实时）(<http://piwik.org/docs/real-time/>) 方式下运行。但是当跟踪大网站时，强烈建议禁用“real time”报告，例如，对当天的报告，每 1-6 个小时处理一次。采用这种方式，访问用户界面只能看到预处理过的报告。

注意：*Real Time Visitors widget*（实时访问者小窗口）和 *Visitors Log*（访问者日志）仍然是实时的。只是，其他“today”（今天）报告不包含最新的访问和页面访问量。

- n 建议删除陈旧的日志，例如删除 6 个月前的日志，使数据库大小能更有效管理。
- n 更新 Piwik 到最新版本。有时，Piwik 的新版本包含大数据表的 SQL 模式更新（如增加新字段，新索引等）。模式的更新将耗费几分钟或几小时。建议通过 shell/命令行方式执行。

## 3. Piwik 的其他高性能建议

### 3.1. 设置可忽略的 URL 参数

获得优秀性能的一个重要因素是减少跟踪唯一 URL 的数量。请确保设置了不想统计的可忽略的 URL 参数（例如会话 ID 参数）。跟踪的唯一 URL 越少，Piwik 的性能越好。

### 3.2. 负载均衡

对于高级用户，可采用负载均衡方式，在负载均衡器后面使用多个服务器。

Piwik 将所有跟踪数据，报告，用户，设置等信息保存在共享的 mysql 数据库中。Piwik 也需要读/创建文件用于存储信息，缓存，临时文件等：

- n 每个请求都会用到 Piwik 的 `config/config.ini.php`。它包含数据库访问，启用的插件以及各种设置。  
该文件只能由超级用户在管理界面更改，例如插件，超级用户口令，smtp 设置等等。)，该文件必须在所有 Piwik 服务器之间保持同步。
- n `tmp/*` 包含零食和缓存的文件，例如已编译的模板，生成的 PDF，跟踪缓存等。该目录可以放心删除，Piwik 会自动重新编译模板和重新创建缓存。
- n 默认情况下，Piwikby 使用文件会话存储，通常存放在 `tmp/sessions`，或全局会话目录（查看 `php.ini`）。  
一个配置可设置启用数据库存储。在文件 `config/config.ini.php` 中，  
[General] 区块内，增加下面内容（如果不存在）：

```
session_save_handler = dbtable
```

注意:如果有很多并发仪表盘用户，需要增加客户端对数据库服务器的最大连接数。

如果 Piwik 使用在负载均衡器后，请按下面步骤操作：

- 1) 同步所有服务器上的 `config/config.ini.php` – 确保所有服务器上的 `config/config.ini.php` 文件完全相同（可使用 `rsync`）。
- 2) 删除每次新部署中 `tmp/*` 下的内容（例如，当升级 Piwik 或插件时）。
- 3) 启用数据库会话存储
- 4) 强烈推荐启用 Piwik 的 SSL。

至此，Piwik 应该可在负载均衡环境中运行。例如将 Piwik 负载均衡到三个服务器上：负载均衡器在 2 个 web server 之前，这两个 Web Server 跟踪/查询

1 个主 mysql 数据服务器。进一步优化，可以再增加一个 UI web server 处理所有 API 请求，UI 请求和运行 cron 归档。这个服务器也可以再加一个服务器并做负载均衡。

通常，与主 DB 服务器（多核，8G 内存，快速磁盘访问）相比 web server 相对便宜。主数据库服务器可扩展到支持每天超过 1M 页面访问量。

### 3.3. 一些经验数据

如果网站每天的页面浏览量大约是 500k 或更多，(每月多于 15M 网页访问量)，Piwik 可能会达到 php 内存的最高限制，或 MySQL 或硬件的限制。但是确实有一些用户可跟踪 1b 的页面访问量。

粗略估计每 5M 页面访问量，Piwik 需要大约 1GB MySQL 数据库空间。如果您的网站每天跟踪 100k 页面访问量，一年后数据库大约是 7GB。

## 4. 如何设置自动归档

如果网站的访问量每天多于上千次，Piwik 处理数据的时间就需要几分钟。避免等待的最佳办法是在服务器上设置 cron 作业，使得每小时自动处理数据。

### 4.1. Linux/Unix: 如何设置 crontab 自动归档报告

Crontab 是基于时间的调度服务。Crontab 需要安装 php-cli 或 php-cgi。可能也需要 ssh 访问服务器：

```
# crontab -e
```

然后写下面行：

```
MAILTO="youremail@example.com"  
5 * * * * www-data /usr/bin/php5 /path/to/piwik/misc/cron/archive.php --  
url=http://example.org/piwik/ > /home/example/piwik-archive.log
```

Piwik 归档脚本将每小时的第五分钟运行。通常在一分钟内结束。对于大型网站（1 万浏览量或更多），归档需要 30 分钟。

参数的详细介绍：

- n `MAILTO="youremail@example.com"` — 如果执行脚本时产生错误，[脚本输出和错误信息将发送到 youremail@example.com](#)。
- n `www-data` 是执行 cron 作业的用户，通常是 web 服务器的用户。
- n `/usr/bin/php5` 是 PHP 执行文件。根据服务器的具体配置和操作系统可能会有所差别。可以在 shell 中执行“which php5”或“which php”找到 php5 执行文件的绝对路径。
- n `--url=http://example.org/piwik/` 是脚本中唯一的必须的参数，必须设置为 Piwik 的基础 URL，例如 [http://analytics.example.org/](#) 或 [http://example.org/piwik/](#)
- n `/home/example/piwik-archive.log` 是脚本输出的路径，如果不需要输出可以设置为 `/dev/null`，脚本输出通常会包含有用的信息，例如归档的网站名，处理时间等。

**‘linux cron’工具的说明：** cron 工具使用两种不同烈性的配置文件：系统 crontab 和用户 crontab。他们之间的差别在于第六个字段。

- n system crontab，第六个字段是执行该命令的用户名。这允许 system crontab 可以任何用户身份执行命令。

**n** user crontab, 第六个字段是执行的命令, 所有命令都以创建该 crontab 的用户身份执行。这是一个重要的安全特性。

如果以用户 crontab 设置 crontab, 请按下面设置:

```
5 * * * * /usr/bin/php5 /path/to/piwik/misc/cron/archive.php --  
url=http://example.org/piwik/ > /dev/null
```

Crontab 作业将每小时的第 5 分钟触发归档处理。这保证当访问 Piwik 仪表板时, 数据是已经处理过的, Piwik 的加载很快。

**测试 cron 命令:** 为确保 crontab 实际正常工作, 在 shell 中以 crontab 用户执行脚本:

```
# su www-data -c "/usr/bin/php5 /path/to/piwik/misc/cron/archive.php --  
url=http://example.org/piwik/"
```

应该能看到脚本的输出, 归档的网站列表, 以及结束时的总结。

## 4.2. Windows: 如何使用 Windows 任务计划设置自动归档

打开 Windows 上的任务计划程序, 单击“所有程序”->“附件”->“系统工具”->“任务计划”。

单击‘添加任务计划’并输入名称, 例如“Piwik Auto Archiving”。单击“日程安排”, 设置为每天, 每小时执行。确认设置并单击选项卡“任务”。

在任务属性中, “运行”输入框中, 输入执行脚本命令。例如:

```
C:\xampp\php\php.exe "D:\www\piwik\misc\cron\archive.php" --  
url=http://piwik.example.org/
```

## 4.3. CPanel: 如何使用 CPanel 设置 cron 脚本

如果使用 CPanel, Webmin 或 Plesk 等用户界面, 设置比较简单。下面是 CPanel 的说明:

1. 登录到 CPanel
2. 单击“Cron Jobs”
3. Email 保持空白
4. 在‘Minutes’设置 00, 其他保持空白

5. 然后粘贴 PHP5 执行文件路径, Piwik misc/cron/archive.php 脚本的路径, 以及 Piwik 基本 URL 参数- `url=http://piwik.example.org/`

下面是一个示例( 在本示例中需要将'yourpanelsitename'修改自己的名称)

```
/usr/local/bin/php -f
/home/yourpanelsitename/public_html/piwik/misc/cron/archive.php --
url=http://example.org/piwik/ > /home/example/piwik-archive-output.log
```

"yourpanelsitename" 往往是您的域的前 8 个字符。

6. 单击 "Add New Cron Job"

Piwik 将每个小时自动处理报告。

## 4.4. 如果网站不支持 cron 任务, 使用 Web Cron

如果可能, 我们强烈建议运行 cron 或任务计划。但是在一些共享的服务器上, 或一些特殊的服务器配置上, 使用 cron 或任务计划很困难。

一些 web 主机让您设置一个 web cron, 这是一个简单的 URL, 主机将在计划时间内自动访问它。您可以在主机界面输入下面 URL:

```
http://your-server.org/path/to/piwik/misc/cron/archive.php?token_auth=XYZ
```

将 XYZ 替换为超级用户的 32 个字符的 token\_auth。可以超级用户登录到 Piwik, 单击页面顶部的 API 链接, 将显示 token\_auth。

可以将 URL 粘贴到您的浏览器中, 等待几分钟, 处理结束后, 查看输出。web cron 应该每小时至少触发一次。