Docs/EPrintsStatsInstall

This document explains how to set up the ePrintsStats package that was developed at the School of Computing, University of Tasmania, and later modified extensively at the University of Otago. Please note that the package is based on code that was developed at University of Melbourne and we acknowledge their contribution to the final package.

The instructions should be fairly generic and enable you to set up the stats package in any environment (for example, I originally wrote these instructions while attempting to install the software under Mac OS X). You should change paths as appropriate to match your local conditions.

1. Prerequisites
2. Download and unpack the stats package
3. Configure ePrintsStats for your site
4. Load the database schema
5. Parse the eprints web server log files
6. Add crontab entries to update tables
7. Configure HTML include files for your site
8. Configure EPrints
9. [Optional] Set up GeoIP tracking for your local intranet
10. Check that your site works
11. Create link(s) to ePrintsStats on your EPrints site

1. Prerequisites

- Web server (Apache 1.3 or 2.0 should both work)
- Installed ePrints
- MySQL database engine (4 or 5)
- PHP (works with 4.3/4.4, should work with 5)

1.1. Installing GeoIP

ePrintsStats used to use the `Geo::IPfree` database, but this is now very out of date, and has been largely superseded by MaxMind's free GeoIP database, which is updated monthly (a commercial version is also available). They appear to provide a GeoIP module for most things, including PHP, Perl and Apache 2. The best performing approach would be to install the `mod_geoip` Apache module then access that from PHP, but that only works under Apache 2 (and also does not seem to install under Mac OS X for unknown reasons). Alternatively, there is a pure PHP module which is easy to install. Both options

are described below.

Resources:

- various APIs: ⊕ http://www.maxmind.com/app/linux
- free country database: ⊕ http://www.maxmind.com/app/geoip_country

1.1.1. Install GeoIP library (including database)

Download the source from ⊕ http://www.maxmind.com/app/c, then do the usual:

```
% ./configure
% make
% sudo make install
```

The libraries are installed in `/usr/local/lib`, the headers in `/usr/local/include`, the binaries in `/usr/local/bin` and the GeoIP database file in `/usr/local/share/GeoIP`. Change the `configure` parameters if you want different locations.

1.1.2. Install mod_geoip (Apache 2 only)

Download the module from ⊕ http://www.maxmind.com/app/mod_geoip, then:

```
% apxs -i -a -L/usr/local/lib -I/usr/local/include -lGeoIP -c mod_geoip
.c
```

Note that this doesn't seem to work under Mac OS X, so you will have to use the pure PHP module if you intend to deploy on Mac OS X. Similarly, if you are using Apache 1.3, you have no choice but to use the pure PHP module. See the next section.

1.1.3. Install PHP module (works in all cases)

Download the module from ⊕ http://www.maxmind.com/download/geoip/api/php/geoip.inc and put it somewhere in PHP's `include_path` (for example, `/usr/local/lib/php`). No other configuration is required.

1.1.4. Ongoing maintenance

The free database is updated once per month. Download it from ⊕ http://www.maxmind.com/app/geoip_country and drop it into `/usr/local/share/GeoIP` (or wherever you installed it), replacing the existing copy. Everything should continue working transparently. You could set up a cron job to do this.

2. Download and unpack the stats package

The package is available for download via the URL: 🌐
http://eprints.comp.utas.edu.au:81/downloads/ePrintsStats.zip. Extract the contents of the
zip-compressed archive to an appropriate installation location (e.g.,
`/usr/local/eprints`).

3. Configure ePrintsStats for your site

Edit the configuration items listed in the files below as necessary. All items marked with a
◆ **must** be changed regardless (they are either security related things like passwords or
fundamental configuration items). Other items only need to be changed if the defaults
don't suit (e.g., if you want to run the database on a separate machine, you will need to
change the server names). Examples of possible values are given in [brackets].

3.1. config/inc.vars.es.php

The following are all elements of the `$GLOBALS["config_vars"]` array:

- `connections["sqlpass_public"]` ◆ is the password for the statistics
  database.
- `connections["sqlserver"]` is the server for the statistics database.
  [`localhost`]
- `filelocation["base_install_dir"]` ◆ is ePrintsStats' installation
  directory. [`/usr/local/eprints/ePrintsStats`]
- `general["admin_name"]` and `general["admin_email"]` are the
  name and email of the ePrintsStats administrator.
  `general["support_name"]` and `general["support_email"]` are
  the same thing for the support contact. [`admin@eprints.example.com`]
- `general["eprints_location"]` ◆ is the URL for the target EPrints
  installation. [`http://eprints.example.com`]
- `local_networks` lists the "country" names and corresponding codes for all IP
  ranges that you wish to track separately (e.g., local intranet traffic). See section 9 for
  details.

3.2. scripts/eprints-usage.php

- `$gi` uses the path to the GeoIP database. You will need to change this if you
  installed it in a non-standard location.
- `$log_dir` ◆ is the directory containing the Apache logs. This will probably be

something like `/var/log/httpd/`, but this will vary depending on your OS and Apache installation (e.g., under Mac OS X with Fink, it's `/sw/var/apache2/logs/`).

- `$log_file` ⬦ is an associative list that maps EPrints repositories to Apache log files. [`'example_eprints' => 'access_log'`]
- `$sqlserver` is the server for the statistics database. [`localhost`]
- `$sqlpass` ⬦ is the statistics database password.
- `$sqlserver2` is the server for the EPrints database.
- `$sqluser2` ⬦ is the username (= archive name, usually) for the EPrints database. [`example_eprints`]
- `$sqlpass2` ⬦ is the password for the EPrints database.
- `$local_IPs` is an array of local IP addresses that can be used to separate accesses from machines in your local intranet from the rest of the world. If you don't need to do this, just set the array to empty. See section 9 for more detail.

Some of these changes also need to be made to `scripts/fix-countries.php`.

3.3. sql/db.schema.eprintstats.sql

- There are two statistics database passwords at the end of the file for the "public" and "private" ePrintsStats users. ⬦

3.4. vhost/apache.conf

Change the various paths to match the local environment, for example: ⬦

```
Alias /es "/usr/local/eprints/ePrintsStats/htdocs"
  ...
  php_admin_value error_log "/usr/local/eprints/ePrintsStats/logs/php.e
rror.log"
  php_admin_value open_basedir "/usr/local/eprints/ePrintsStats:/usr/lo
cal/lib/php"
  php_admin_value include_path ".:/usr/local/eprints/ePrintsStats/inclu
des:/usr/local/eprints/ePrintsStats/config:/usr/local/lib/php"
  ...
```

The PHP library paths are the most likely to vary from one OS to another (e.g., under Mac OS X with Fink it's `/sw/lib/php4`). You **must** include the path to the `ePrintsStats/includes` directory, otherwise nothing will work!

4. Load the database schema

Load the mysql schema from `sql/db.schema.eprintstats.sql`:

```
% mysql -u root -p < db.schema.eprintstats.sql
```

5. Parse the eprints web server log files

Parse the log files by running:

```
% php scripts/eprints-usage.php
```

This will extract the access data from the log files, determine the correct country information, and then load this into the stats database.

5.1. Special note when upgrading from older versions of ePrintsStats

If you are upgrading from an older version of ePrintsStats that used the obsolete `Geo::IPfree` module, you may wish to run:

```
% php scripts/fix-countries.php
```

to ensure that your existing country data is fully up to date. This is entirely optional, however, and should only ever be done **once** (IP ranges may move around, thus causing the country stats to change over time, which isn't what you want).

6. Add crontab entries to update tables

Run the usage script regularly to update the usage tables by adding it to the script owner's crontab. For example the crontab entry might be:

```
30 2 * * * /usr/local/bin/php -n /usr/local/eprints/ePrintsStats/script
s/eprints-usage.php
```

7. Configure HTML include files for your site

Edit at least these HTML include files to match your site. You might for example use the headers, footers, and style sheets of your EPrints site.

- `includes/inc.html.header.es.php`
- `includes/inc.html.footer.es.php`

8. Configure EPrints

- ◈ ◈ **EXTREMELY IMPORTANT!** ◈ ◈ Edit `ArchiveConfig.pm` for each archive, and add `'/es'` to `rewrite_exceptions`. It you do not do this, ePrintsStats simply **will not work** (causing much swearing and frustration with Apache). If you cannot find the `rewrite_exceptions` configuration option, you probably have an older version of ePrints, and this warning is therefore irrelevant.

- Edit `apachevhost.conf` for each archive and add the following line:

```
Include /usr/local/eprints/ePrintsStats/vhost/apache.conf
```

  Note that ePrintsStats also doesn't appear to work if this `Include` appears outside of the archive's `<VirtualHost>` directive.

9. [Optional] Set up GeoIP tracking for your local intranet

We find it valuable to be able to separate internal accesses from our local intranet from those elsewhere in our country. To do this you will need to know the IP address range(s) used by your institution, and design a small logo in png format. Address ranges are specified in the `$local_IPs` array in `scripts/eprints-usage.php`. Each entry in this array is a list of IP addresses corresponding to a particular "country" that you want to track (for example, Otago University differentiates between local Intranet traffic and repository admin traffic). Assign a different country code for each range and create flag icons as appropriate. ISO 3166-1 specifies the range XA through XZ as "user-assignable", so you can use codes from there as necessary (for example, use "XA" for admin traffic and "XI" for intranet traffic).

Each address range key is the name that will appear in the statistics database (the "country name"), followed by a comma, followed by the appropriate ISO 3166-1 country code as noted above. Each entry in the range is either a single IP address, or an array specifying inclusive lower and upper bounds for a contiguous IP address range (see example below). All IP addresses must be converted to long values using the ip2long() function before being stored.

Note that address ranges may overlap. The script will use the first range that matches a given IP, so list the ranges in the correct order of precedence for your needs.

If you're not interested in separating out intranet traffic, then simply set the array to empty.

Example:

```
$local_IPs = array(
    'Repository Admin,XA' => array(
        ip2long('192.168.1.5'),
        ip2long('192.168.1.22'),
        array(
            'lower' => ip2long('192.168.1.30'),
            'upper' => ip2long('192.168.1.35'),
        ),
    ),
    'Our Intranet,XI' => array(
        array(
            'lower' => ip2long('192.168.1.0'),
            'upper' => ip2long('192.168.255.255'),
        ),
    ),
);
```

In this example, 'Repository Admin' covers the IP addresses 192.168.1.5, 192.168.1.22 and the range 192.168.1.30 to 192.168.1.35, inclusive. 'Our Intranet' covers the range 192.168.1.0 to 192.168.255.255, inclusive. A machine will only match the 'Our Intranet' range if it first fails to match the 'Repository Admin' range.

You will also need to set up corresponding entries in $GLOBALS["config_vars"]["local_networks"] in scripts/inc.vars.es.php, so that the ePrintsStats web pages can trap these country codes and display them correctly. For the example above, the settings would be:

```
        "local_networks" => array(
            "XA" => "Repository Admin",
            "XI" => "Our Intranet",
        ),
```

10. Check that your site works

Depending on the set up of the alias for this project you can browse to it using either http://<yourhost>/es/ or http://<yourhost>/es/index.php.

11. Create link(s) to ePrintsStats on your EPrints site

Decide where you want to provide access to ePrintsStats on your EPrints web server. For example, you might want to put a link on your EPrints home page or in the EPrints footer or both. Have a look at these sites to see how they did it: 🌐 http://eprints.comp.utas.edu.au:81/ or 🌐 http://eprints.otago.ac.nz/.

You could insert something similar to the following HTML into your EPrints home page:

```
<p><a href="http://<yourhost>/es/index.php?action=show_detail_date;rang
e=4w">Detailed statistics</a>
   for individual ePrints.</p>
```

and/or insert the following HTML into your EPrints footer:

```
<a href="/es/index.php?action=show_detail_date;range=4w">Detailed Stati
stics</a> |
```

Remember that you will need to regenerate your EPrints views and abstracts and perhaps even restart your EPrints server for the changes to be fully made.

last edited 2006-01-19 16:13:16 by NigelStanger