

Report on Software Decision

Author	Santiago Chumbe
Workpackage	1
Task	1.4
Date	3 - October - 2000
Version	1.0
Summary	A brief report to explain why we have chosen the " <i>infernal trio</i> " Apache/ MySQL/PHP to deploy the software package for FAILTE, and why we have discarded the ROADS approach.

1. Introduction

When we started to research for the right software to develop the FAILTE database/catalogue, we noticed that we have to choose between two options:

- A search/retrieval software for Internet resource catalogues based on a flat database system;
- A search/retrieval system based on a Relational Database Management System (RDBMS) [1] designed for putting databases on the web.

There is specific software already developed to deploy search/retrieval systems for the Internet. Some are true search engines, which crawl the web creating their listings automatically (such as HotBot [2]), or some are directories, which look for matches only in their listings previously created (such as Yahoo[3]). In this context, the RDN recommends the use of the JISC funded software ROADS [4] for creating RDN Internet resource catalogues. The ROADS software is a free Yahoo-like system, written in Perl/UNIX. It includes features for linking distributed cooperative databases together using the WHOIS++ protocol, and the Common Indexing Protocol (CIP). ROADS is partially used by EEVL.

On the other hand, in a relational database, data can be logically viewed as tables (relations) comprised of rows (records) and columns (fields). All access to and manipulation of data in a RDBMS is accomplished via Structured Query Language (SQL). This data is put in the web by means of a database API. There are many well-established RDBMS providers. However, as we have the constraint that no budget has been allocated for buying software, our natural choice was MySQL [5]. The grounds for this is that MySQL is very fast, reliable, and easy to use. It also has a very practical set of features developed in very close cooperation with its users. MySQL was originally developed to handle very large databases much faster than existing solutions and has been successfully used in highly demanding production environments for several years (see Appendix A). Though under constant development, MySQL today offers a rich and very useful set of functions. The connectivity, speed, and security make MySQL highly suited for accessing databases on the Internet. Furthermore, it is free.

In conclusion, we merely have to choose between ROADS version 2.3 and MySQL. (In this document, we will use the term MySQL to signify the RDBMS MySQL version 3.22.32 as well as all the technology needed to put a MySQL database in the web).

Since FAILTE will benefit from the use of EEVL hardware and servers, it was not necessary to spend time evaluating Operating Systems and Servers. From the beginning it was assumed that FAILTE would be installed in a Sun Solaris UNIX system and it will use an Apache server.

2. Selection: ROADS vs. MySQL

Our strategy of selection was based on:

- evaluation of the performance and service delivered by both ROADS and MySQL against a defined criteria of selection and;
- analysis of the feedback gathered from actual users of ROADS and MySQL

If we recall that the main purpose of FAILTE is to create a database which is searchable online via a WWW interface, we can build up our criteria of selection concentrating our attention on how robust and efficient ROADS and MySQL are delivering database access via the Internet. Therefore, our criteria are based in an assessment of the following issues:

1. **Scalability:** We want to know what will happen to the efficiency of ROADS and MySQL when the database size grows and/or the Internet requests increase to the point where it gets hundreds of hits per day.
2. **Portability:** To assess the capability of ROADS and MySQL to be installed and used on UNIX platforms without modification.
3. **Performance:** This involves mainly quality factors of the software operation such as speed, reliability, security, user-friendly degree, etc.
4. **Technical Support:** Good technical support will help the FAILTE team in trouble shooting during the database system creation process and, it will guarantee smoother maintenance work in the future when FAILTE is embedded into the activities of EEVL.
5. **Protocols:** This factor determines conformance with cataloguing standards and the compliance with search/retrieval protocols (Whois++, IMS, Dublin Core, etc.)
6. **Integration:** This aspect will help us to identify which options offer a more transparent integration of FAILTE within the EEVL, LTSN-Eng and EASEIT-Eng services.
7. **Flexibility:** To evaluate how easy is to integrate the source code of ROADS or MySQL and change it to fit the FAILTE needs.
8. **Third part support:** To identify the contributed software (applications, languages, etc.) available for each one of the options (ROADS and MySQL).

2.1. Evaluation of ROADS and MySQL against the criteria of selection

Scalability

According the ROADS documentation and the feedback received from some ROADS users (i.e. OMNI, Quality Internet Resources in Health and Medicine [6]), it is possible to deduce that ROADS is not scalable. It doesn't have any problem in handling a few thousands of records, but it becomes extremely slow if it has to manage more than 4000 records. On the other side, MySQL is particularly good at handling thousands of records (the MySQL creator advice is to use MySQL only for very large database applications).

Portability

ROADS is available for the most important variants of the Unix operating system e.g. Linux, FreeBSD, NetBSD, OpenBSD, SunOS, Solaris, Digital Unix (OSF/1), SGI Irix and HP/UX. MySQL also support almost the same spectrum of UNIX implementations (Linux, HP-UX, IBM AIX, Sun Solaris, SGI Irix, SCO UnixWare, FreeBSD, etc.). ROADS as well as MySQL run on the Apache server. Therefore, from the FAILTE point of view, ROADS and MySQL have the same level of portability.

Performance

The search/retrieval process speed of ROADS and MySQL have been estimated trying some of their most popular web implementations (see Appendixes A and B). There did not seem to be important differences in searching-speed of the two applications for medium size databases (less than 2000 records). Their reliability can be deduced from the trends which show that the number of users of ROADS is not increasing significantly, whilst the number of search engines based on

MySQL is growing rapidly. Therefore, it seems that the market trusts MySQL to a greater extent. Both options offer a similar level of security. In terms of user-friendliness, MySQL is much better than ROADS (*"the user interface of ROADS is one of the worst I've seen"* reported an online catalogue manager [7]).

Technical Support

At the moment ROADS is only maintained by two people [8], and their future is not clear (Martin Hamilton said, *"I believed I've arrived at a compromise which should give people currently using ROADS a way to migrate to industry standard search and retrieval data management technology"*). MySQL has ample free technical support, plenty of documentation, more than 20 mailing lists and the confidence that MySQL is developed by a company under the GPL (GNU General Public License).

Protocols

One of the advantages of ROADS is that it includes a single metadata format, the IAFA template, and a single search and retrieval protocol, Whois++. Also, there are a number of plug-ins which may allow a ROADS implementation to connect with other protocols, such as Zplugin (to provide access to Z39.50 clients), roads2metadc.pl (to output a ROADS Dublin Core record as HTML META tags), etc. MySQL deploys a general-purpose solution. Therefore, it is possible to develop a system able to interact with other services, which are using protocols such as Whois++, Dublin Core or Z39.50. The only challenge will be the development effort needed to implement these standards into FAILTE.

Integration

It is expected that EEVL, LTSN-Eng and EASEIT-Eng will share the FAILTE database. These services will access FAILTE when the end-user is looking for electronic learning and teaching resources or when a new resource is identified. A more detailed description of how FAILTE will be integrated into these services is not available to us at the moment. The only thing we can assume now is that a more flexible tool will give us more integration freedom. Therefore, if ROADS is a more flexible tool than MySQL, then we can expect that ROADS will be a useful choice for the integration of FAILTE into the three mentioned services.

Flexibility

Both ROADS and MySQL, are Open Source Software, which means that the user can customise the code to fit his needs. ROADS have been written in Perl, the MySQL approach can be written in any language, which suits the user (the MySQL database is written in C). Moreover, MySQL has all the advantages of a standard such as SQL and it has proved to combine well with powerful and easy-to-learn languages such as PHP [9,10].

Third part support

A number of plug-ins and tools have been created to enhance the functionality of ROADS [11]. Unfortunately, these enhancements are limited to user-outcomes and mainly use the Perl language. In contrast, MySQL has a lot of contributed software available. It is very likely that most popular web application/languages already supports MySQL. For instance, MySQL can be accessed from an MS-Access user-friendly data entry interface by means of an ODBC driver. In contrast, the building of a communication between ROADS and MS-Access can be very challenge.

We have summarised the evaluation of the criteria issues in order to make this presentation clearer. This summary is shown in Figure 2.1. When an issue is fully

accomplished for the option (ROADS or MySQL) a point is assigned to the appropriate option column (+1). In the opposite case, a point will be subtracted (-1). If the issue could be potentially fulfilled or it doesn't mean a difference between MySQL and ROADS, no points are subtracted or added (0).

Issues	ROADS	MySQL
Scalability	0	0
Portability	0	0
User-friendliness	- 1	0
Searching speed	0	+ 1
Technical Support	- 1	+ 1
Protocols	+ 1	0
Integration	0	0
Flexibility	- 1	+ 1
Third part support	- 1	+ 1
Summary	- 3	+ 4

Table 2.1. Summary of the evaluation of the criteria issues

2.2. Analysis of the feedback from users of ROADS and MySQL

There are a number of requirements that ROADS users are willing to be implemented in ROADS. Most of them are related with user-interface and template creation process. The main complaint of the ROADS users is related with its cataloguing interface. There are other comments reported from ROADS users, such as:

- There is not any default way of using wildcards in ROADS searches
- ROADS does not provide the necessary index files and some problems with the addsl.pl script have been reported in this context.
- Its cataloguing interface is difficult to use. For instance, there is not a way of preventing the user from jumping to the top of the "create record" form every time they use an authority file, which can be very annoying when the users have to scroll down every time they use an authority file.
- In fact, the Web based template editor out to be completely re-written, as it's very hard to use.
- It doesn't provide search interfaces to allow the cataloguer to track how end-users search the catalogue
- It doesn't provide facilities to perform administrative functions such as making a record as reviewed, pending, don't include, etc.
- It is not prepared to build up relationships between different records
- Its technology is not updated with the technological developments.

Actually since the ROADS Project ended, only two people has been involved in the maintenance and user support, who have manifested that it is difficult to maintain and to add new features to ROADS. Many users of ROADS have serious plans to migrate to industry standard search/retrieval management systems. For instance the Art, Design, Architecture & Media Information Gateway (ADAM) [12] has switched entirely from ROADS to Index+. Some users, such as the National Agricultural Library (NAL) [13] of USA, are trying to replace the flat files system in ROADS with MySQL.

There is information claiming that the very ROADS team is working to produce a new ROADS package with the MySQL database. In fact some ROADS users have already created a mass of Perl scripts to access MySQL databases (see for instance [14]).

Finally about ROADS:

- A strength of ROADS that we may acknowledge is that it makes searching in flat files easier and it incorporates the Whois++ server protocol for sharing records.
- A general opinion that we have obtained from users, who still using ROADS is that if they were to do all over again, they would consider not using ROADS.

On the other hand, reports on the strong and the weak points of MySQL have been identified; for instance we have found the following comments:

- MySQL is extremely good for logging, when you do many connects (connect is very fast), for selecting and inserting records at the same time, when selects and updates are done by using index keys, when you use many tables without long conflicting locks and when you have big tables (MySQL uses a very particular and compact table format).
- Things to avoid with MySQL: Using variable-length columns (you get more fragmentation of a table due to the differing sizes of the records), updating to a table with deleted rows, joining tables without using index keys or keys which are not unique enough, having on things you can have in a conditional clause.

In general, from the user comments we can deduce that MySQL gives more robust databases. Flat databases are made of sequential associative arrays, which cannot handle as many records as MySQL. The relational structure of MySQL databases allows us to speed-up the searching by indexing big tables without scalability problems.

3. Conclusion

MySQL is better than ROADS to deploy the FAILTE search/retrieval online system.

4. FAILTE Software Configuration

- Host : Sun Sparc Workstation
- Operating System : UNIX SunOS Solaris spey 5.7
- Web Server : Apache 1.3
- RDBMS : MySQL 3.22.32
- Main API : PHP 4.0.0
- Build date : Aug. 14, 2000

The installation and configuration process of MySQL and PHP has been done in a single day.

5. References

- [1] <http://www.citilink.com/~jgarrick/vbasic/database/rdbms.html>
- [2] <http://hotbot.lycos.com>

- [3] <http://www.yahoo.com>
- [4] <http://www.roads.lut.ac.uk>
- [5] <http://www.mysql.com>
- [6] <http://omni.ac.uk/>
- [7] Parkinson B., BIOME Technical manager, Greenfield Medical Library, Nottingham
- [8] <http://www.roads.lut.ac.uk/v2/Manual/manual-1.html#ss1.20>
- [9] <http://www.php.net>
- [10] <http://www.indexdata.dk/phpyz/>
- [11] <http://www.ukoln.ac.uk/metadata/software-tools/>
- [12] <http://adam.ac.uk/adam/tech/>
- [13] <http://www.nal.usda.gov>
- [14] <http://www.infomotions.com/alex/infrastructure/>

Appendix A

Large search/retrieval services based on MySQL. It lists companies and organisations, which are using MySQL to deploy fast and secure access to large amount of information on the Internet

A.1. Web search engines

- AAA Matilda Web Search
- What's New
- Aladin
- Columbus Finder
- Spider
- Blitzsuche
- Indoseek Indonesia
- Yaboo - Yet Another BOOKmarker
- OzSearch Internet Guide
- Splat! Search
- The Open Source Digital Library System Project
- Yahoo!
- Slashdot: A pro-Linux/tech news and comment/discussion site
- Freshmeat: News about new versions of computer related stuff

A.2. Information search engines concentrated on some area

- SpyLOG ; A very popular web counter site
- All about Linux
- TuCows Network; Free Software archive
- Jobvertise: Post and search for jobs
- The Music Database
- Fotball (Soccer) search page
- TAKEDOWN - wrestling
- The International Lyrics Network
- Musicians looking for other musicians (Free Service)
- AddALL books searching and price comparison
- Harvard's Gray Herbarium Index of Plant Names
- The Game Development Search Engine
- My-Recipe.com; Cookbook at i-run.com
- The Innkeeper Vacation Guides
- The Mac Game Database uses PHP and MySQL
- Research Publications at Monash University in Australia
- Occupational Health & Safety website database (a project for the ECC)
- Bioinformatics databases at the Montreal Children's Hospital using MySQL

A.3. Online magazines

- Spoiler Webzine.
- Daily news about Linux in German language
- Betazine - The Ultimate Online Beta Tester's Magazine
- Computer Currents Magazine

- Linxtoday
- 32Bits Online: because there's more than one way to compute

A.4. Web sites backed by a MySQL database

- Amsterdam Airport Schiphol
- CD database
- Used Audio Gear Database
- Musical note-sheets
- AZC.COM's Feature Showcase
- Course Search
- Northerbys Online Auctions
- Bagism - A John Lennon fan page
- US Folk art broker
- Mail reading on the web
- Free home pages on www.somecoolname.mypage.org
- Der Server f@"ur Schulen im Web (In German)
- Auldhaefen Online Services
- CaryNET Information Center
- Dataden Computer Systems
- Andr'emuseet (In Swedish)
- HOMESITE Internet Marketing
- Jade-V Network Services
- Weather World 2010 Technical Credits
- About The Gimp plugin registry
- Java tool Archiver technical detail (Slightly optimistic about MySQL ANSI-92 compliance)
- Games Domain Cheats Database
- The "Powered By" Page (Kcilink)
- Netcasting
- NBL (Australian National Basketball League) tipping
- CGI shop
- Whirlycott: Website Design
- Museum Tusculanum Press
- Centro Siciliano di Documentazione
- Quake statistics database
- Astroforum: Astrologie and related things (in German)
- OpenDebate - Interactive Polls & Open Discussion
- Online chemical dissertation server
- FreSch! The Free Scholarship Search Service
- Stockholm Pinball Locator
- HEK A construction company
- Elsevier Bussines Information
- Medical Links (Using ColdFusion and MySQL)
- Search for jobs & people at JobLink-USA
- Competition Formation Skydiving
- E-commerce and internal accounting
- Denmark's leading business daily newspaper B@o{rsen}
- The Internet NES Database
- Travel agency in Prague in 3 languages
- Linkstation
- Searchable online database at Peoplestaff

- A searchable database system for horse classified ads
- The Poot site
- "Playin' in the LAN"; a network monitoring suite
- U.S. Army Publishing Agency
- Realestate handling in Yugoslavia
- PIMS; a Patient Information Management System
- Pilkington Software Inc
- A Vietnam Veteran's Memorial (The Wall) database.
- Gamer's Union specializes in auctions of used & out of print gaming material
- A daily bulletin at Monterey High school
- Community-owned site serving Lake Washington's Eastside residents and businesses
- French bowling site.

Appendix B

Large search/retrieval services based on ROADS. ROADS is being used in a number of places and for a bewildering variety of purposes. Here are some of the leading examples of services based on ROADS.

- Alex - the Alex Catalogue of Electronic Texts
- BeCaL - the Belief, Culture and Learning Information Gateway
- Biz/ed - Business Education online - Internet Catalogue
- CRUISE - CRanfield University Internet Site Explorer
- History On-line
- JIME - the Journal of Interactive Media in Education
- Leeds University Library (pick: *databases, selected web sites or electronic journals*)
- the Kuopio University Virtual Library - catalogue of biomedical Internet sites
- NADIR Links - catalogue of links to alternative and left-wing/socialist Internet sites
- NOVAGate - the Nordic Gateway to Information in Forestry, Veterinary and Agricultural Sciences
- Port - the National Maritime Museum's online catalogue
- OMNI - The UK's gateway to high quality biomedical Internet resources
- ROUTES - Resources for Open University Teachers and Students
- SOSIG - The Social Science Information Gateway
- WVEVL - the WasteWater Engineering Virtual Library