Statement of work

Client: University of Texas at Austin, CIPRES Project. PI: Tandy Warnow

Project: TreeBASE2 Hosting and Database migration

Document history

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draft 1.0</td>
<td>1/13/2009</td>
<td>Hilmar Lapp</td>
</tr>
<tr>
<td>1.0 final</td>
<td>1/28/2009</td>
<td>Hilmar Lapp</td>
</tr>
</tbody>
</table>

Introduction/Background

TreeBASE is a primary repository of published phylogenetic data, containing records from over 2,000 publications, making it a major community resource in evolutionary biology. The NSF-funded Cyberinfrastructure for Phylogenetic Research (CIPRES) project has developed a second-generation database, TreeBASE 2, which allows for a richer set of phylogenetic data and metadata to be stored, and will also support different ways of accessing the data.

The National Evolutionary Synthesis Center (NESCent) is an NSF-funded center with the mission to advance synthetic research to address fundamental questions in evolutionary biology. As one of its core objectives, the Center’s Informatics program promotes availability, sharing, and interoperability of data and software.

CIPRES and NESCent have agreed to move TreeBASE 2 from its current physical host (the San Diego Supercomputing Center (SDSC)) to NESCent prior to the expiration of the CIPRES project. Contingent on renewal of its NSF grant funding, NESCent would host TreeBASE for up to five years. This move entails migrating the current TreeBASE 2 database and web application code from DB2 to PostgreSQL as the database server (RDBMS).

The rest of this document details the tasks and equipment purchases NESCent will carry out under this contract, until Nov 2009. If NESCent’s funding is extended, the new Center grant will become active in Dec 2009, at which time NESCent will take over financial and managerial responsibility for TreeBASE 2. TreeBASE personnel will continue to be responsible for data curation, and a TreeBASE Foundation Board currently being created will provide long-range planning for the database.
Project objectives

The objectives of this project are the following.

i. Move physical hosting of the TreeBASE2 database, middleware, and web-application from SDSC servers to hardware (to be purchased) hosted at NESCen.

ii. Migrate the TreeBASE 2 database, and any software dependencies of the TreeBASE2 middleware and web-application, from DB2 to PostgreSQL as the relational database server.

iii. Provide administration for the TreeBASE hardware, software, and database.

iv. Provide, as needed, technical advice and training to TreeBASE and CIPRES personnel in the PostgreSQL, Hibernate, and Spring technology platforms.

v. Work with TreeBASE and CIPRES personnel to identify technical issues and possible solutions to establish TreeBASE mirror sites.

Requirements

TreeBASE2 hosting:

- Web-application and middleware require a Java application server (JBoss, or possibly Tomcat), supporting at least Java 5, with at least 4GB of physical memory available.
- Database requires a PostgreSQL server instance, v8.3+, with at least 100GBs disk space, and 4GBs of physical memory available for internal PostgreSQL memory pools. The database is not required to, but allowed to, reside on the same physical machine as TreeBASE web or command-line applications.
- TreeBASE web and command-line applications are compatible with a POSIX-compliant Unix operating system.
- TreeBASE developers (and testers) need to be able to test and deploy application and database updates without disrupting the publicly accessible instance of the application or database.

Relational database migration:

- For the purposes of this contract, the migration work is complete when all web-application and middleware software components of TreeBASE2 work on top of a PostgreSQL v8.3+ database as they worked on top of a DB2 database.
- The migrated application and database must pass the same tests as on top of DB2, return the same results for the same queries, and achieve comparable performance within a margin of 30%.
- If PostgreSQL migration degrades performance more than this margin, database performance tuning (PostgreSQL configuration, query optimization, index optimization) is required as part of the migration work.

TreeBASE2 administration:
• The TreeBASE web-application and database is a global resource and must in principle be available online 24/7, except for planned downtime and unplanned outages.

• Planned downtime for system and software maintenance may extend for several hours during the period of 5pm-9am US Eastern Time (GMT-5/-4), and must be announced to the primary TreeBASE contact in time for dissemination.

• Emergency maintenance in order to protect the TreeBASE system or data integrity (such as in response to security vulnerabilities) is possible at any time, and may include a temporary shutdown of all related servers or services.

• The TreeBASE2 online status must be monitored through appropriate software (Nagios) that detects and reports unplanned outages, such as due to hardware, system, or network failure.

• It is sufficient to address unplanned outages at the next working hour window. Working hours are Mon-Fri 9am-5pm US Eastern Time (GMT-5/-4).

TreeBASE2 advice on technology platforms:
• NESCent programming staff will work with TreeBASE and CIPRES staff as suitable to resolve any programming-related issues and questions on technology platforms that NESCent IT is cognizant of.

Risks
Some of this work and the timelines given are contingent on CIPRES/TreeBASE personnel meeting their own milestones that are required to be completed prior to this work commencing.

Specifically, before the migration work (tasks #3-4) can start, at least the data submission and curation interfaces of TreeBASE2 must be functional and tested, the data must have been migrated (and the results verified) from TreeBASE1 to TreeBASE2, and submission to TreeBASE1 must have been discontinued. In order to validate the successful migration of any search-related database access code, the search interface of TreeBASE2 must also be functional and tested. To ensure meeting the performance requirements, appropriate metrics for comparison must first be established for the current TreeBASE2 host and database.

At the time of writing, all of these dependencies were expected to complete prior to March 31, 2009.

The TreeBASE2 schema, code base, or data may harbor unanticipated problems, especially in areas not inspected prior to commencement of this contract, that could lead to significantly more effort required by NESCent than taken into account for the given timelines, or in the worst case put the fulfillment of the work outside of NESCent’s
abilities. The parties will negotiate in this event whether to change timelines, budget, or any other aspects of the contract, or to terminate the contract.

Hosting of TreeBASE2 at NESCent beyond November 30, 2009, is contingent on NESCent being able to successfully secure continuation funding by that date.

Tasks

1. NESCent will purchase two server machines with identical hardware (with the possible exception of amount of RAM) that satisfies the hardware requirements. One of these will serve as production server, and the second as development and testing server. NESCent will add these machines to the NESCent server rack, install all necessary software, and do any additional system setup needed.

2. NESCent will create and manage user accounts for developers as necessary on the development server, and develop and implement a production deployment procedure, including for rolling out application and database schema upgrades.

3. NESCent will work with CIPRES/TreeBASE personnel to migrate the TreeBASE2 database schema and content from the current DB2 backend to PostgreSQL. A TreeBASE2 database instance will be brought up on PostgreSQL on the NESCent servers. An executable and as much as possible automated procedure will be developed for migrating the data. Ultimate responsibility for the success of this step rests with NESCent; however, CIPRES/TreeBASE personnel will assist to the best of their abilities.

4. NESCent will work with CIPRES/TreeBASE personnel to migrate all database access within the TreeBASE2 web-application and middleware codebase to be compatible with PostgreSQL. Together with the database migration, this will result in TreeBASE2 instance on NESCent-hosted servers that works on top of a PostgreSQL backend. Ultimate responsibility for the success of this step rests with NESCent; however, CIPRES/TreeBASE personnel will assist to the best of their abilities.

5. NESCent will provide continued server, web, and database administration for the TreeBASE servers until Nov 2009, according to requirements. This includes integration into NESCent’s monitoring and backup infrastructure.

6. From Jan - Sep 2009, NESCent will provide expertise and technical advice on an as-needed basis to CIPRES/TreeBASE personnel on all technology platforms employed by the TreeBASE2 architecture that NESCent IT staff is cognizant of. In particular, this includes PostgreSQL, Hibernate, Spring, Java servlet and web-service programming, and JBoss and Tomcat application deployment.

7. NESCent will work with CIPRES/TreeBASE personnel to identify technical requirements, obstacles, and solutions for mirroring TreeBASE2 at other institutions. At a minimum, this will result in a document listing host
requirements, TreeBASE2 software requirements, technical issues including unmet software requirements, and possible solutions for these issues.

Timelines

The numbers below correspond to task numbers. Timelines are contingent on TreeBASE2 development milestones by the CIPRES personnel (see Risks).

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Duration</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Hardware purchased and installed</td>
<td>4/24/2009</td>
<td></td>
</tr>
<tr>
<td>#2 Development and deployment ready</td>
<td>5/8/2009</td>
<td></td>
</tr>
<tr>
<td>#3 Schema and content migrated to PostgreSQL</td>
<td>5/22/2009</td>
<td>The final migration may be later.</td>
</tr>
<tr>
<td>#4 Software migrated to PostgreSQL</td>
<td>6/5/2009</td>
<td></td>
</tr>
<tr>
<td>#5 Monitoring and backup integration</td>
<td>5/15/2009</td>
<td></td>
</tr>
<tr>
<td>#6 Technical consultation provided</td>
<td>On-going</td>
<td>Jan – Sep 2009.</td>
</tr>
<tr>
<td>#7 Mirroring documentation complete</td>
<td>Sep 2009</td>
<td></td>
</tr>
</tbody>
</table>

Cost estimates

<table>
<thead>
<tr>
<th>Budget item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware: at least 2x 4 CPU cores, 4GB RAM, 250GB disk space (*)</td>
<td>$5,700</td>
</tr>
<tr>
<td>Migration to PostgreSQL (consulting)</td>
<td>$2,900</td>
</tr>
<tr>
<td>Technology platform advice (consulting)</td>
<td>$3,700</td>
</tr>
<tr>
<td>System, database, and web administration</td>
<td>$7,300</td>
</tr>
<tr>
<td>Backup tapes and off-site storage</td>
<td>$500</td>
</tr>
<tr>
<td>Project management</td>
<td>$2,900</td>
</tr>
<tr>
<td><strong>Project Total</strong></td>
<td><strong>$23,000</strong></td>
</tr>
</tbody>
</table>

(*) NESCent is testing virtualized servers based on a resource “cloud”. If the tests are successful, TreeBASE2 will be hosted on virtual servers, yielding up to 50% more resources available to it at the same cost.

Payment is due in total at commencement of the contract.

Signatures

University of Texas, Austin 

[Name, Title]

NESCent

[Name, Title]

Date: ______________________________ Date: ______________________________