Running Large Scale Astronomy Events on the Web

G. Tulloch[1]– P. Trudel[2] – M. Simmons[3]

1 – RASC Winnipeg Centre, 14 Rainbow Cove, Winnipeg, Manitoba, R2J 3Z4
2 - RASC Winnipeg Centre, Box 79 Grp 2 RR 1, Anola, Manitoba, R0E 0A0
3 – Astronomers Without Borders, 26500 Agoura Rd., Suite 102-681, Agoura, CA 91301 USA

Abstract

The 100 Hours of Astronomy IYA2009 Cornerstone Project was a four-day event designed to bring astronomy to the public around the world.

In this paper we review the 100 Hours of Astronomy event with particular reference to how a large scale event of this nature is tied to the web presence for the event, evolution of the web site, and issues that were experienced in the time preceding the event, during, and following. Of particular relevance was evolution of the technology environment required to support the event, software employed, and how the technical environment responded to the anticipated and unanticipated volume of participants in the events. Finally, the paper examines how lessons learned during the 100HA event can be used to plan for future events during the International Year of Astronomy and beyond.

The 100 Hours of Astronomy Experience

The 100 Hours of Astronomy IYA2009 (IYA2009) Cornerstone Project was a four-day event designed to bring astronomy to the public around the world. Groups planned events ranging from a few hours on one day to a 100-hour marathon, with plans that fit the resources and enthusiasm available. All activities and events during 100HA were aimed at bringing astronomy to a new audience and providing a million people with a `Galileo Moment.`` In addition to groups worldwide publishing their own events via the 100hoursofastronomy.org web site, 100 Hours of Astronomy also conducted several global events promoted through a centralized web site. Science facilities and enthusiasts were encouraged to make whatever plans were appropriate to their resources, and to be creative and innovative in attracting the public to the events in their areas. There were several component projects within the 100HA central program, including:

- Around The World In 80 Telescopes Web Cast - a 24-hour program web cast from within 76 observatories and space telescope control rooms. Each observatory's segment included a brief prerecorded introduction to their facility and the work they do followed by live presentations from astronomers on-site and questions from the public. The web cast was anchored from the headquarters of the European Southern Observatory (ESO) in Garching, Germany.
- 24-Hour Global Star Party - a series of telescope observing events open to the public held by amateur astronomers throughout the world.
- 100 Hours of Remote Astronomy - several remotely-operated telescopes offered free time to allow the public to sample the capabilities of this emerging technology.
- Sun Day - a day organized by the IYA2009 Solar Physics Task Group that supported worldwide activities focusing on the Sun to the public, particularly through public observing events.

In the following sections we review the progress of the web site during the ramp up to the event,

how the event itself transpired, and post-event activities.

Before the Event

In January of 2009 the 100hoursofastronomy.org website transitioned from a simple HMTL based site hosted to a LAMP (Linux/Apache/MySQL/PHP) based site implemented on the Joomla 1.5x content management system to enable content editing via web browser by the 14-member 100HA Task Group. This group was responsible for organizing the worldwide effort on behalf of the International Astronomical Union.

Two volunteer web masters were engaged through a general request for assistance, one (GT) predominately responsible for hosting and configuration of servers and network, and one (PT) responsible for content and Joomla configuration and customization. Part of the core 100HA web site was an event registration function, EventList, deployed as a module to Joomla. When modifications were required a vendor was engaged to modify the code to include Google Maps functionality and generally enhance the application for use in registration and display of 100HA events worldwide. The initial hosting for the site was via an inexpensive, mass market web hosting company located in the United States.

As site traffic increased, a preplanned move to a dedicated hosting environment was made, initially to a single server hosted at a data center in Chicago USA. Configuration of this server began in early March 2009 and the site was transitioned to the new server March 17. There was some concern that the change to the Domain Name record for the site to change the IP address from the old to the new server would take a while to propagate through the Internet due to caching of the address. Both sites remained active during the changeover, and the databases were compared after a few days to catch any data that was entered on the old site before the new site address propagated. However, only 3 events were registered during the interim period and manually transferred to the new site, so propagation delays were very minimal, likely because most users were new to the site and had not cached the old IP address.

During the 100HA Event

The new server operated at very light loading until the commencement of the 100HA event. The technical team hypothesized that load on the server would likely increase to approximately 25,000 visits per day during the event, but most of the load would be reading information rather than entering of new events. As such, the number of visits would be higher, but the actual load per visit on the server would be lower as read activity (viewing web pages produced from the database) was expected to be much more efficient than write activity (e.g. adding new events to the database.) Because of minimal loading of the server at approximately 5000-6000 visits per day, it was judged that the site could remain on a single server dual core server with 4GB of RAM and handle the load. Plans were made to add a second server to host the database if required, and inquiries were made on the method by which multiple servers could be set up on a load balanced network to handle large loads, but none of these plans were put into effect to minimize costs.

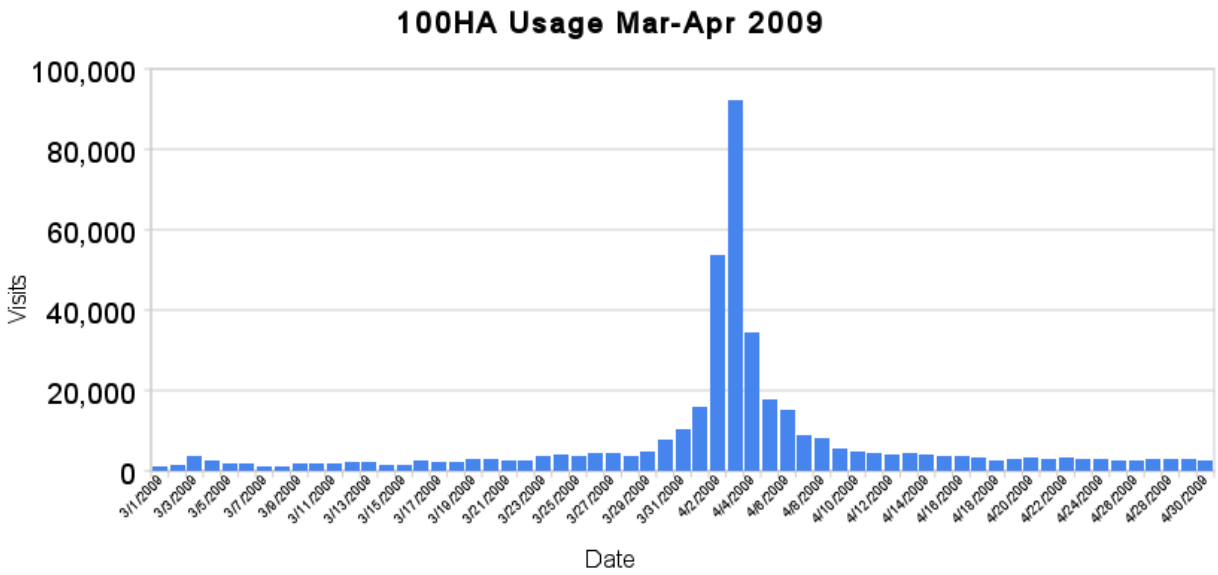

Figure 1 - Chart comparing visits across March 1 - Apr 30 2009

On April 2nd and 3rd, on commencement of the Around The World in 80 Telescopes web cast, utilization of the site skyrocketed, with 32,000 visits on the 2nd growing to 98,000 on the 3rd of April, mostly attributable to visitors accessing the site to view the web cast. Since the actual web cast content was hosted on a partner web streaming provider, previously the web cast was not expected to introduce a heavy load on the 100HA server. Once the user had linked to the streaming provider web site the user would have no further interaction with the 100HA web site. This assertion proved to be in error, since as constructed all users accessing the web cast needed to load the 100HA front Joomla page every time they accessed the web cast. The content editors embedded a viewer for the video directly on the 100HA web site main page which ensured that traffic on the 100HA main page was inadvertently maximised. To exacerbate matters, during the web cast issues with the streaming video provided caused users to "reload" the page often to unfreeze the video.

The web cast therefore created an additional huge volume of traffic on the 100HA site as many hundreds (possibly thousands) of people, near-simultaneously pressed the refresh buttons on the browsers, forcing the web server to re-transmit the 200+ kilobyte Joomla main page. Most of this retransmission would have been pulled from Joomla's built-in cache, however substantial parts of the site are built dynamically on the fly, and therefore forced Joomla to go back to the database for a significant amount of content. While the web cast was hosted on Ustream (a sponsor who provides web casting) and thus the majority of the data being transmitted was obtained from another site, repeated transmission of the embedded viewer on the main page of the 100HA web site front page introduced extreme unanticipated load on the server, which caused performance to plummet. Once server CPU and memory utilization became very large, the MySQL database task on the server was unable to keep up with requests and began returning database errors to users, directly affecting the visitor's experience. In the end, this resulted in the site becoming unusable.

In the early morning hours of April 3, the first remediation tactic employed was to redirect the 100hoursofastronomy.org domain to an ESO web server which served up a simple web page with the link to the web cast, which proved effective. This page was migrated to the main 100HA server, and during the web cast was the only content displayed by the site. Since the priority was to provide the link to the web cast, the simple web page remained in place throughout web cast on the 3rd of April. Several times through the day the main site was re-enabled after attempts were made to resolve performance issues, but each attempt was rolled back to the simple HTML page due to the number of visitors rapidly saturating the server. This saturation turned out to be worse than the original loads on the server, since Joomla's built-in caches had expired, and effectively resulted in every access to the site hitting the database server, causing server load to reach extreme levels.

Interestingly, the Open Source software on which the entire 100HA site ran performed extremely well under enormous loads. For example, a measure of utilization of a LINUX server is the Load Average, representing the number of CPUs required to handle the load being imposed on the server. Normally, this measure is less than one (in other words there are no server tasks waiting to be processed at any given time) during the web cast event the load average climbed to a load average of 125 – the web masters expected that the entire system would crash long before this point was reached. However, amazingly, the LINUX server continued operating, the MySQL server continued serving database requests (albeit occasionally causing errors), and the Apache web server software continued to process requests for web pages. At no point, even during the most intense pressure on the server, did any component fail. The technical team was extremely impressed by the robustness and efficiency of the LINUX Operating System (CentOS 5), the MySQL database, Apache web server software and PHP Hypertext Preprocessor. Of particular interest was the ease at which the Apache web server handled the web cast traffic after the main site was disabled - server load averages remained under 1 while the simple page was replacing the Joomla site. This implies that use of content management applications like Joomla require

significantly more resources under high load than a simple HTML page.

A database server was ordered from the hosting provider and was in place by the following day, which proved to be effective in reducing the server load and eliminating database errors, but by that time the web cast had concluded and extreme loads had dissipated. In addition to the database server an additional "mirror" server was put in place, expanding the server infrastructure to two content servers accessing a separate database server. Unfortunately, load balancing would not be available until the last day of the event, which was judged to be too late to prove useful, so a link was placed on the main page of the 100HA site to invite users experiencing performance issues to access the mirror server. This proved effective but given that by that time the load had diminished, the second server saw little activity. As the 100 Hours of Astronomy ended, volumes on the server diminished to the previous levels.

After the Event

The web site remained active with 5000 visitors per day during the weeks immediately following the event, as users accessed recordings of the web cast, entered event reports, and posted photos of their local events. By the end of April 2009 through Mid-May (the time of this writing) visitors had stabilized at approximately 2000 per day as users added reports of their 100HA activities and perused the reports and photos uploaded by others.

Issues and Lessons Learned

The following sections enumerate issues that were experienced during the 100HA program that required action and generated some interesting learning experiences for the Task Group and technical team.

Site Content/Joomla

As the 100HA site got larger, a number of issues arose regarding how much data was being sent to clients logged onto the site with web browsers, particularly through slow dial-up and satellite connections. For example, the front page of the web site grew to approximately 600k in size at one point through extensive use of graphics plus a dynamic map on the main page that contained stickpins in a Google map. Each stickpin data point was sent as an in dependant record in an XML table to the site and, as the number of events around the world grew to thousands, the data being sent to clients became quite voluminous.

For content related issues, the lessons were learned included:

- Limit front page content size as much as possible - this is the first page users see and if it takes prohibitively long to download, users may choose to visit another site. In the 100HA case hindsight shows that a simple HTML page front ending the Joomla web site would have reduced or eliminated subsequent problems.
- Use small graphics and limit the number of graphics - balance the desire to make a very attractive front page against the user experience with the lowest common denominator of connectivity, particularly if use is worldwide and developing countries with lower prevalence of broadband connections are a target audience.
- Optimize HTML content, (such as defining explicit sizes for graphics and tables) so that the HTML content can be rendered and displayed according to the intended design as supporting files (java script, CSS, images) are loading in subsequent steps.
- Test and validate the rendering of the site using as many browser combinations as possible. This can be facilitated by websites such as BrowserShots that can validate the rendering on

a wide variety of web browsers on any of the widely used client operating systems.
- Ensure that the software providing the dynamic content (in our case, Joomla) is producing the minimum amount of data being sent to the browser. For example, Joomla by default loads a large (200+K) Javascript library on every page it serves, requiring the client browser to download this each time the site is accessed. It is important to note, that much of this framework, is a one-time load each time the site is accessed. As the user navigates the site, these one-time load items are left in the browser's cache. However when the page is refreshed, the 200+K is retransmitted.
- Minimization of client bandwidth utilization through the use of on-the-fly compression techniques. While not enabled by default in Apache web server, the mod_deflate plugin takes advantage of the fact that textual content (such as HTML webpages) can be heavily compressed. Whilst requiring a small CPU overhead to compress textual content, the benefits to network consumption are impressive. Using the Joomla framework as an example, which clocks in at 200+K, the plugin compresses the data to roughly 80K before it is transmitted to the client for subsequent decompression when the client visually renders the content.

Planning for extremely high volumes

The huge volume of visits to the site on April 3rd is instructive from a number of points of view and a primary lesson learned is to fund and create a much more resilient technology environment than created for the 100HA event.

Lessons learned:

- Expect the unexpected for volumes (4 times expected might be a good estimate) and configure the site to accommodate those loads
  Provision servers ahead of time – by the time new servers can be brought online, it's too late, at least for short events like 100HA. Costs are fairly minimal (in our case about $400/month per server) compared to the cost of not providing the service when demand is highest.

- Consider using the most simple, fast loading (i.e. simple HTML with few graphics) front page to the site possible, especially if there is content (such as web casts) for which the main site is simply a "gateway".The Apache web server proved to extremely robust and able to serve up pages to almost 100000 visitors during the period where the main page was offline - if the front page of the site had been simple HTML, it is likely that the single server (with a separate database server) could have handled the load.

- Create an "edit enabled" server where content managers can work at maximum performance, even during peak loads.
- Similarly, structure the site (which doesn't change and is normally implemented as simple HTML pages) such that users need only access dynamic content when there is a need for it -in the 100HA example most content was relatively static and only implemented in Joomla to facilitate online content editing by the Task Team. The only content that needed to be dynamic was the Event add/edit/change/display code.

Technology Configuration

On the technology side, there were a number of areas where the 100 Hours of Astronomy experience offered lessons learned:

- Always configure web sites with separate web application server (in the 100HA case Joomla/Apache) and the database server (MySQL) prior to expected peak loads to ensure that high user loads on the web application servers do not cause

- Consider load balancing the web site, particularly when the site is predominately static. Load balancing distributes users across multiple servers for the same domain. This is generally implemented as a hardware appliance that sits in front of the web servers and "divvies" up loads between those servers.

- Optimization of database configuration - although MySQL works 'out-of-the-box', it is not adequately configured for a full-scale production server. Tools such as *sqltuner.pl* should be used to optimize the MySQL database for the available hardware and system resources. Also, see also caching below.

Conclusions

The 100 Hours of Astronomy event was a tremendous success - Sky & Telescope Editor-in-chief Robert Naeye blogged on the 100HA site that "... the actual event exceeded the most optimistic expectations. Reports from around the world indicate that more than a million people looked through a telescope for the first time. And these people were spread over more than 100 nations. This was truly a global event!" (Naeye) While days have been set aside for special events to bring astronomy awareness to the public including Astronomy Day and International Sidewalk Astronomy Night, 100HA was the first united effort with centralized planning and resource management combined with grassroots planning worldwide. Participation in 100HA by astronomy enthusiasts and the public show the value of programs that use the Internet to create truly global programs that are carried out by local planners.

The 100 Hours of Astronomy project demonstrated what a few determined amatuer astronomers can do to enlist the legions of amatuer astronomers worldwide in a project that has a global effect. While not without a few technical hiccups, overall the 100 Hours of Astronomy event's web presence was a crucial underpinning of the whole initiative. Future large scale outreach initiatives (including a potential follow up event in the fall) should consider the 100 Hours of Astronomy initiative as a largely successful prototype, and build upon it's success during this International Year of Astronomy, and beyond.

Works Cited

Naeye, Robert. "100 Hours of Astronomy, A Resounding Success on Several Levels". Retrieved from <http://www.100hoursofastronomy.org/~c100hour/blogs> on May 17 2009.

Web Links

100 Hours of Astronomy Web Site www.100hoursofastronomy.org
Joomla Content Management www.joomla.org
MySQL Database www.mysql.org
The Apache Software Foundation www.apache.org
PHP: Hypertext Preprocessor www.php.net
MySQL Tuner wiki.mysqltuner.com/MySQLTuner
BrowserShots browsershots.org
Apache mod_deflate http://httpd.apache.org/docs/2.0/mod/mod_deflate.html