

Playing Tag In the Dark: Diagnosing Slowness In Library Response Time

In this article the author explores how the systems department at the Auraria Library (which serves more than thirty thousand primarily commuting students at the University of Colorado–Denver, the Metropolitan State College of Denver, and the Community College of Denver) diagnosed and analyzed slow response time when querying proprietary databases. Issues examined include vendor issues, proxy issues, library network hardware, and bandwidth and network traffic.

Why is everything so slow?" This is the question that library systems departments often have the most trouble answering. It is also easy to dismiss because it is often the fault of factors beyond the control of library staff. What usually prompts these questions are the experiences of the reference librarians. When these librarians are trying to help students at the reference desk, it is very frustrating when databases seem to respond to queries slowly, files take forever to load onto the computer screen, and all the while the line in front of the desk get continues to grow. Or the library gets calls from students using databases and the catalog from their homes who complain that searching library resources takes too long, and that they are getting frustrated and using Google instead. This question is so painful because libraries spend so much of their shrinking budgets on high quality information in the form of expensive proprietary databases,

and it is all wasted if users have trouble using them. In this case the problem seemed to be how slow the process of searching for information and downloading documents from databases was. For lack of a better term, the Auraria Library called this the "response time" problem. This article will discuss the various ways the systems (technology) department of the Auraria Library, which serves the University of Colorado–Denver, Metropolitan State College of Denver, and the Community College of Denver, tried to identify problems and improve database response time.

The systems department defined "response time" as the time it took for a person to send a query from a computer at home or in the library to a proprietary information database and receive a response back, or how long it took to load a selected full-text article from a database. When a customer sets out to use a database in the library, the query to the database could be slowed down by many different factors. The first is the proxy, in our case Innovative Interfaces' Inc. Web Access Management (III WAM), a product that authenticates the user via the III API (Application Program Interface) product. To do this the query travels over network hardware, switches, and wires to the III server and back again. Then the query goes to the database's server, which may be almost anywhere in the world. Hardware problems at the database vendor's end can affect this transfer. In the case of Auraria Library this transfer can be influenced by traffic on the library's network, the university's network, and any other place in between. This could also be hampered by the amount of memory in the computer where the query originates, by the amount of tasks being performed by that computer, etc. The bandwidth of the network and its speed can also have an effect. Basically, the bottlenecks needed to be found and fixed. Bottlenecks are described by Webopedia as "the delay in transmission of data through

the circuits of a computer's micro-processor or over a TCP/IP network. The delay typically occurs when a system's bandwidth cannot support the amount of information being relayed at the speed it is being processed. There are, however, many factors that can create a bottleneck in a system."¹

Literature review

There is not a lot on database response slowness in library literature, probably because the issue overlaps with computer science and really is not one problem but a possibility of one of several problems. The issue is figuring out where the problem lies. Gerhan and Mutula examined technical reasons for network slowness, performing bandwidth testing at a library in Botswana and one in the United States using the same computer, and giving several suggestions for testing, fixing technical problems, and issues to examine. Gerhan and Mutula concluded that bandwidth and insufficient network infrastructure were the main culprits in their situation. They studied both bandwidth and bandwidth "squeeze." Looking for the bandwidth "squeeze" means looking along the internet's "journey of many stages through routers and exchange points, each successively farther removed from the user."² Bandwidth bottlenecks could occur at any one or more of those stages in the query's transmission. The following four sections parse that lengthy pathway and examine how each may contribute to delays. Badue et al. in their article "Basic Issues on the Processing of Web Queries," described Web

Margaret Brown-Sica (margaret.brown-sica@ucdenver.edu) is Head of Technology and Distance Education Support, Auraria Library, serving the University of Colorado–Denver, Metropolitan State College of Denver, and the Community College of Denver.

queries, load balancing, and how they function.³ Bertot and McClure's "Assessing Sufficiency and Quality of Bandwidth for Public Libraries" is based on data collected as part of the 2006 *Public Libraries and the Internet* study and provides a very straightforward approach for checking specific areas for problems.⁴ It outlines why basic data such as bandwidth readings may not give the complete picture. It also gives a nice outline of factors involved such as local settings and parameters, ultimate connectivity path, application resource needs, and protocol priority. Azuma, Okamoto, Hasegawa, and Masayuki's "Design, Implementation and Evaluation of Resource Management System for Internet Servers" was very helpful in understanding the role and function of proxy servers and problems they can present.⁵

Vendor issues

This is a very thorny topic because it is out of the library's control, and also because the library has so many databases. The systems department asked the reference staff to send reports of problems listing the type of activity attempted, time and dates, the names of the database, the problem and any error messages encountered. A few that seemed to be the slowest were selected for special examination. One vendor worked extensively with the library and in the end it was believed that there were problems at their end in load balancing, which eventually seemed to be fixed. That company was in the middle of a merger and that may have also been an issue. We also noted that a database that uses very large image files, ARTSTOR, was hard to use because it was so slow. This company sent the library an application that simulated the databases' use and was supposed to test to see if bandwidth at Auraria Library was sufficient for that database. According to the test, it was. Databases that consistently were perceived as the slowest

were those that had the largest documents and pictures, such as those that used primarily PDFs and visual material. This, with the results of the testing, pointed to a problem independent of vendor issues.

Bandwidth and network traffic

The systems department decided to do bandwidth testing on the library's public and staff computers after reading Gerhan and Mutula's article about the University of Botswana. The general perception is that bandwidth is often the primary problem in network slowness, as well as the problems with databases that use larger files. Several of the computers were tested in several successive days during what is usually the busiest time for the network, between noon and 2 p.m. The results were good, averaging about 3000 kilobytes per second (kbps). For this test we used the CNET bandwidth meter, which downloads an image to your computer, measures the time of the download, and compares it to the maximum speeds offered by other Internet service providers.⁶ There are several bandwidth meters available on the Internet. When the network administrator checked the switches for network traffic, they showed low traffic, almost always less than 20 percent of capacity. This was confusing: If the problem was neither with the bandwidth nor the vendors, what was causing the slow network performance?

One of the university network administrators was consulted to see if any factor in their sphere could be having an effect on our network. We knew that the main university network had implemented a bandwidth shaper to regulate bandwidth. "These devices limit bandwidth... by greedy applications, guarantee minimum throughput for users, groups or protocols, and better utilize wide-area connections by smoothing out

bursty traffic."⁷ It was thought that perhaps this might be incorrectly prioritizing some of the library's traffic. This was a dead end, though—the network administrators had stopped using the device.

If the bandwidth was good and the traffic was manageable, then the problem appeared to not be at the library. However, according to Bertot and McClure, the bandwidth question is

complex because typically an arbitrary number describes the number of kbps used to define "broadband." . . . Such arbitrary definitions to describe bandwidth sufficiency are generally not useful. The Federal Communications Commission (FCC), for example, uses the term "high speed" for connections of 200kbps in at least one direction. There are three problematic issues with this definition:

1. It specifies unidirectional bandwidth, meaning that a 200kbps download, but a much slower upload (e.g., 56kbps) would fit this definition;
2. Regardless of direction, bandwidth of 200kbps is neither high speed nor does it allow for a range of Internet-based applications and services. This inadequacy will increase significantly as Internet-based applications continue to demand more bandwidth to operate properly.
3. The definition is in the context of broadband to the single user or household, and does not take into consideration the demands of a high-use multiple-workstation public-access context.⁸

Proxy issues

Auraria Library uses the III WAM proxy server product. There were several things that pointed to the

proxy being an issue. One was that the systems department had been experimenting with invoking the proxy in the library building in order to collect more accurate statistics and found that complaints about speed seemed to have started around the same time as this experiment. But if the bandwidth was not showing inadequacy and the traffic was light, why was this happening? The answer is better explained by Azuma et al.:

Needless to say, busy Web servers must have many simultaneous HTTP sessions, and server throughput is degraded when effective resource management is not considered, even with large network capacity. Web proxy servers must also accommodate a large number of TCP connections, since they are usually prepared by ISPs (Internet Service Providers) for their customers. Furthermore, proxy servers must handle both upward TCP connections (from proxy server to Web servers) and downward TCP connections (from client hosts to proxy server). Hence, the proxy server becomes a likely spot for bottlenecks to occur during Web document transfers, even when the bandwidth of the network and Web server performance are adequate.⁹

Testing was done from on campus and off campus, with and without using the proxy server. The results showed that the connection was faster without the proxy. When testing was done from the health sciences library at the University of Colorado with the same type of server and proxy, the response time was much faster. The difference between Auraria Library and the other library is that the community Auraria Library serves (the Community College of Denver, Metropolitan State College, and the University of Colorado-Denver) has a much larger user population who overwhelmingly use databases from home, therefore taxing the proxy

server. The other library belonged to a smaller campus, but the hardware was the same. The proxy was immediately dropped for on-campus users, and that resulted in some response-time improvements. A conference call was set up with the proxy vendor to determine if improvements in response time might be attained by changing from a proxy server to LDAP (Lightweight Directory Access Protocol) authentication. The response given was that although there might be other benefits, increased response time was not one of them.

Library network hardware

It was evident that the biggest bottleneck was the proxy, so the systems department decided to take a closer look at III's hardware. The switch that regulated traffic between the network and the server that houses our integrated library system, part of which is the proxy server, was discovered to have been set at "half-duplex."

Half-duplex refers to the transmission of data in just one direction at a time. For example, a walkie-talkie is a half-duplex device because only one party can talk at a time. In contrast, a telephone is a *full-duplex* device because both parties can talk simultaneously. Duplex modes often are used in reference to network data transmissions. Some modems contain a switch that lets you select between half-duplex and full-duplex modes. The correct choice depends on which program you are using to transmit data through the modem.¹⁰

When this setting was changed to full duplex response time increased. There was also concern that this switch had not been functioning as well as it could. The switch was replaced, and this also

improved response time. In addition, the old server purchased through III was a generic server that had specifications based on the demands of the ILS software and didn't into consideration the amount of traffic going to the proxy server. Auraria Library, which serves a campus of more than thirty thousand full-time equivalent students, is a library with one of the largest commuter student populations in the country. A new server had been scheduled to be purchased in the near future, so a call was made to the ILS vendor to talk about our hypothesis and requirements. The vendor agreed that the library should change the specification on the new server to make sure it served the library's unique demands. A server will be purchased with increased memory and a second processor to hopefully keep these problems from happening again in the next few years. Also, the cabling between the switch and the server was changed to greater facilitate heavy traffic.

Conclusion

Although it is sometimes a daunting task to try to discover where problems occur in the library's database response time because there are so many contributing factors and because librarians often do not feel that they have enough technical knowledge to analyze such problems, there are certain things that can be examined and analyzed. It is important to look at how each library is unique and may be inadequately served by current bandwidth and hardware configurations. It is also important not to be intimidated by computer science literature and to trust patterns of reported problems. The Auraria Library systems department was fortunate to also be able to compare problems with colleagues at other libraries and test in those libraries, which revealed issues that were unique and therefore most likely due to a problem at the library end. It is important to keep learning about how

your system functions and to try to diagnose the problem by slowly looking at one piece at a time. Though no one ever seems to be completely satisfied with the speed of their network, the employees of Auraria Library, especially those who work with the public, have been pleased with the increased speed they are experiencing when using proprietary databases. Having improved on the response-time speed issue, other problems that are not caused by the proxy hardware have been illuminated, such as browser configuration, which may be hampering certain databases—something that had been attributed to the network.

References

1. Webopedia, s.v. "Bottleneck," www.webopedia.com/TERM/b/bottleneck.html (accessed Oct. 8, 2008).
2. David R. Gerhan and Stephen Mutula, "Bandwidth Bottlenecks at the University of Botswana," *Library Hi Tech* 23, no. 1 (2005): 102–17.
3. Claudine Badue et al., "Basic Issues on the Processing of Web Queries," *SIGIR Forum; 2005 Proceedings* (New York: Association for Computing Machinery, 2005): 577–78.
4. John Carlo Bertot and Charles R. McClure, "Assessing Sufficiency and Quality of Bandwidth for Public Libraries," *Information Technology and Libraries* 26, no. 1 (Mar. 2007): 14–22.
5. Kazuhiro Azuma, Takuya Okamoto, Go Hasegawa, and Murata Masayuki, "Design, Implementation and Evaluation of Resource Management System for Internet Servers," *Journal of High Speed Networks* 14, no. 4 (2005): 301–16.
6. "CNET Bandwidth Meter," <http://reviews.cnet.com/internet-speed-test> (accessed Oct. 8, 2008).
7. Michael J. DeMaria, "Warding off WAN Gridlock," *Network Computing* Nov. 15, 2002, www.networkcomputing.com/showitem.jhtml?docid=1324f3 (accessed Oct. 8, 2008).
8. Bertot and McClure, "Assessing Sufficiency and Quality of Bandwidth for Public Libraries," 14.
9. Azuma, Okamoto, Hasegawa, and Masayuki, "Design, Implementation and Evaluation of Resource Management System for Internet Servers," 302.
10. Webopedia, s.v. "Half-Duplex," www.webopedia.com/TERM/h/half_duplex.html (accessed Oct. 8, 2008).

Index to Advertisers

LITA cover 2, cover 3, cover 4