

Resourcing the Unanticipated User: Wikipedia as a Cost Model for Scalable Net-Centric Services

Abstract: Wikimedia (www.wikimedia.org) is analyzed as a cost model for operating a large-scale network-based service. The costs associated with operations and maintenance of a network service are shown to be small compared to the development and procurement costs of a typical Department of Defense (DoD) program. This idea is presented to provide a basis for comparison for program managers of the systems that will form the anchors in the net-centric vision of the Department.

Concerns from DoD Programs

Several of the acquisition program offices that are building systems fundamental to Net-Centric Warfare (such as FCS and DCGS¹) have expressed concerns about the Net-Centric tenet requiring support to the 'unanticipated user'. "How can we plan for something," they ask, "that by definition, we didn't anticipate? What if we aren't resourced for that?"

Consider these questions in the context of a dot-com economy: for an Internet-based business, "What if there are more users than we anticipated?" is the same question as "What if our business is too successful?" Realistically, having "more users than anticipated" is a problem that we want programs to have; unanticipated users means more warfighters are having their needs met.

From a program manager's perspective, however, the question is not an unreasonable one, perhaps best rephrased as "What if we've underestimated the number of users by a large amount?" To support the unanticipated user requires (1) a scalable architecture, and (2) the resources to scale it. The purpose of this paper is to show that program managers should focus on (1), and not worry so much about (2). If the architecture is suitably scalable, the actual scaling costs are shown to be low. If a program is architecturally scalable, and warfighters find value in the service, the PPBES processes exist for finding the funding.

Wikimedia as a surrogate for a DoD Net-Centric system

The [Wikimedia Foundation Inc.](http://www.wikimedia.org) is a non-profit corporation that hosts several projects, notably including [Wikipedia](http://en.wikipedia.org), the award-winning online encyclopedia. (Other projects include [Wikibooks](http://en.wikibooks.org), [Wikiquote](http://en.wikiquote.org), [Wiktionary](http://en.wiktionary.org), etc.) These sites are based on the idea of a Wiki: a website that allows users to add content and to edit existing content. Wikimedia presents a useful case study for costs of a net – centric system because the financial records and decision-making process for the foundation are completely open. Also, the idea of Wiki is close to the tenets of net-centricity; that self-synchronized action from many collaborators can accomplish things that would be otherwise impossible.

Scale of Wikimedia and Wikipedia

As of 10 Dec 2004, Wikipedia, the flagship project, is one of the 300 most-visited websites worldwide. It has over 1.1 million articles in over 100 languages. Across all the projects, the foundation's server architecture handles roughly 550 requests per second. In September 2004, the English edition of Wikipedia, served 1.7 terabytes in 228 million page hits. The hosting contract was for 40 Mbs of data bandwidth. There were more than 30,000 registered users, but because most users do not register, the total number of users is unknown. The average number of user-visits was 766 thousand unique sessions per day.

As of 8 Dec 2004, the Wikimedia server architecture consisted of about 30 servers: 5 database servers (MySQL), 2 file servers (NFS), 6 web cache servers (Squid), 3 overseas web cache servers (Squid), 18 web/application servers (Apache/PHP), and 2 new servers (purpose TBD). All computers are mid-to-

¹ Future Combat System and Distributed Common Ground Systems, respectively

low-end Intel or AMD servers with 1 or 2 processors. The primary hosting site is in Florida, with overseas cache servers in France.

Suitability as a Surrogate

Obviously, Wikimedia is not a warfighter system. It is, however, a highly dynamic web-based data repository that serves as an information resource for hundreds of thousands of people.

In terms of scale, Wikimedia is big site, but not big like Amazon, eBay or Google. From a DoD perspective, Wikimedia has far more users than most warfighting or intelligence systems will. (Although the average request size is likely smaller than a system that handles ISR imagery.) A system that is expected to support warfighters may require overseas replication, which increases costs, but Wikimedia has this concern also and has already begun to address it in the current architecture. Systems with universal interest, such as MyPay, will likely have more total users, but less concurrent users.

As a well-known, public-facing service on the Internet, the Information Assurance threat environment of Wikimedia is significantly higher threat than most DoD systems, which typically operate on closed networks such as SIPRNET or JWICS. This higher threat is balanced by the lower sensitivity of the information. While Wikimedia has no analogous requirements for certification and accreditation (C&A) such as the DITSCAP, C&A costs are largely per-architecture, and only moderately proportional to scale. (I.e. Once an accreditation is complete, adding more servers requires only an update of the C&A, which is far less costly than the initial effort.)

While at first glance, Wikimedia is unlike anything used by the DoD, in broad terms of function, scale, and environment, it is not unlike the concepts of net-centric systems that might implement the "shared space" as described in the DoD Net-Centric Data Strategy. The DoD sees significant use of web technology on a much smaller scale. In OIF, for example, many units put up their own web pages on the SIPRNET, in order to share intelligence and operational plans for the day. On a larger scale, the Distributed Common Ground Station programs, example, would be implementing a shared space somewhat like Wikimedia, where, in the broader application, many unanticipated users could be accessing and posting information to that space.

Operation and Maintenance Costs

The Wikimedia Foundation's 2004 Q4 budget was for \$79,200. Of this, \$50K was budgeted on hardware, \$20K on hosting, \$7.6K was for travel, and \$1.6K for other expenses (accountant, domain names, etc.) The hosting provider donates 50% of the bandwidth costs, so there is an implicit additional \$20K expense. Wikimedia also notes that they do not currently have an adequate backup system, which would also increase costs if resourced appropriately. A more realistic estimate of material operations cost would be \$125K. (i.e. costs not including labor)

In addition to the material costs, developers donate their time and expertise to operate the server farm. The value of this manpower contribution to operations is difficult to estimate, since all these people are working on Wikimedia on a part-time, volunteer basis, and the operations cost is not separated from the development cost. A nominal workforce for operating a server farm of a comparable size (the OSD/PA&E servers) is about 3-5 system administrators. Very broadly, these costs could range from \$0 (for programs where the system administrators are already on staff and this task is an additional duty), up to perhaps \$3M/year. For comparison, the current development contract cost ceiling for DCGS is \$500M.

Architectural Scalability

One of the important insights about Wikimedia, in this context, is that it is designed from the very start to be scalable. Success for a Wiki project model is to have many users; more users means more

articles, and more users editing and refining the content. With this in mind, the Wikimedia architecture was built so that as few costs as possible scale with the size of the user population. The number of servers and the amount of bandwidth are unavoidable scaling factors; the number of administrators is not. Similarly, by using only open-source software, Wikimedia's software licensing costs do not scale at all, and contracted software support² (if used at all) is a fixed cost independent of scale. This approach also simplifies the project management costs; there are no contracts to modify or additional negotiation with the software vendors.

Similarly, a business model that, for example, requires human intervention for account creation and password resets will require a number of human system administrators that increases proportionally with the user population. As shown above, humans are an order-of-magnitude more expensive than the computers they operate. The \$3M system-administrator cost estimated above was for general care-and-feeding of high-availability servers, but did not include additional support unique to the business operations of the subject system. If those costs are not avoided architecturally, they can swamp the other operations costs. An alternative architecture that uses PKI certificates for authentication through the use of the security core enterprise services avoids those costs by relying on another provider to manage those administrative costs associated with authentication data and services.

Conclusions

Much of the change incorporated in the vision of Net-Centricity is cultural change. The real success criteria for program managers is to extend the reach of their data and services to all warfighters in the enterprise that find value in these services. To the degree that this success criteria is understood, program managers will also understand that having too many users is an indication of their success, and begin planning for the notion that they might provide services that will help the Department fight wars and conduct business more effectively. Rather than *ignoring* about the unanticipated user, we should be thinking about how to *attract* them by providing services they want to use.

With this attitude shift, net-centric programs will start thinking harder about the scalability of their systems. The Wikimedia example demonstrates that the costs of operating a high-bandwidth, scalable, and popular network service can be fairly low (\$500K/year) under the right conditions. The department has multiple mechanisms through PPBES³ for funding initiatives that turn out to be popular (i.e. useful). The key lesson from Wikimedia is that funding is not so much the issue as architectural limits that prevent it from scaling to meet the demand. Adding more servers and more bandwidth is an easy fix; redesigning a system after the fact is hard.

No matter how hard we try, the unanticipated users will *always* be unanticipated, and we will have to grow the systems we build to accommodate them. If we can't anticipate who, or when, or how many, we need to build systems with enough room to grow when they do arrive.

Reference Links

http://meta.wikimedia.org/wiki/Hardware_capacity_growth_planning
<http://wikimedia.org/stats/en.wikipedia.org/>

Author Contact: Dan Risacher, Daniel.Risacher@osd.mil, 703-602-1098

² Several companies offer technical support for the specific open-source products used by Wikimedia .

³ Planning Programming Budgeting and Execution System