Dr. Willy Chiu Addresses the World Internet Center on Mega-Site Development

"There is no free lunch. This is tough stuff." — Dr.Willy Chiu

Dr. Willy Chiu is an expert on planning, building and maintaining high-volume Web sites. He advised the audience at the Silicon Valley World Internet Center to follow a development cycle of evaluating infrastructure, determining workload, and then designing, building and testing.

Dr. Chiu stressed the importance of optimizing the system's scalability. Bursty traffic loads are common among the top eBusiness sites. "These traffic spikes can be deadly to a non-scalable system," he explained.

In his work at IBM, Dr. Chiu follows these five best practices for maximum scalability: Thoroughly understand your application environment, categorize the workload, determine the most-impacted components, select the proper scaling techniques and apply them.

For more information, IBM has made four white papers named **Design for Scalability**, **Planning for Growth**, **Design Pages for Performance** and **Web Site Personalization** available at http://www.ibm.com/devloperworks/library/hvws

Planning, workload and scalability key issues for high-volume Web sites

February 28, 2001, Palo Alto, Calif. -- Dr. Willy Chiu, Vice President, High Volume Web Sites for IBM's Software Group, told an audience at the Silicon Valley World Internet Center that, when developing a mega-site (planned for more than 20,000 concurrent users), one must first evaluate its infrastructure, know the workload patterns, and then design, build and optimize for scalability.

Evaluate Infrastructure

A very large eBusiness site typically follows a three-stop evolution. First, information is published on the Web through collaboration. Then, the Web site incorporates some of the business systems for commerce. Finally, business methods are transformed with full integration. Challenges of the last step include infrastructure scaling, seamlessly integrating IT with your business processes, and managing hundreds or thousands of servers across the enterprise with a "cautious sense of fragility."

"Your infrastructure allows you to conduct business in new ways." Dr. Chiu said. "As the process evolves, your value increases."

Know your Workload

A Web-site's development cycle begins with the plan, then design, building and testing, to deployment and service delivery. You must know your workload throughout the entire cycle. Ask yourself what you are trying to do.

How do you categorize your workload pattern: (examples in parenthesis)

? Publish and subscribe
? Customer self-service
? Online trading
? Business-to-business
? Online shopping

Evaluate your infrastructure requirements. What are the requirements based on your workload pattern, and why? How will your site serve and integrate high transaction rates, new and existing systems, rapid application development, security and manageability, open standards, new devices and multi-modality?

Smart Design

At the beginning of the design cycle, plan for growth, structure for scalability and build pages for performance. "You need to ask the business questions of why a particular requirement in necessary." Dr. Chiu advised. "Ninety percent of the time we get answers."

As far as actual page design, Dr. Chiu said, "When you look at Web-page aesthetics, you have to know what business value they bring,."

Prepare for the Burst

As you build and test, closely manage the system's end-to-end performance to optimize your resources. A mega-site requires high transaction rates. Sites in the top-10 traffic statistics receive 20,000-40,000 concurrent users. The Olympics' Web site received 1.2 million hits each minute when Venus Williams was playing in the championship match.

"Bursty site loads are very common. Traffic spikes are deadly… like a tsunami coming to your site." Dr. Chiu said. "You must know how to handle that."

Another question to ask yourself is, "What are my business's core competencies?" If the answer is competitive-edge IT, customers are often more reluctant to outsource. "It depends on who's the CIO, and what kind of power trip he's on," Dr. Chiu said, grinning.

Server Farms

In many cases, it pays to outsource. Online traders plan for 3:1 headroom; their servers are rarely used more than 30%. To combat the wastefulness of idle servers, IBM offers massive server farms, "acres of machines" described by Dr. Chiu. IBM invented capacity-on-demand to help online businesses endure traffic spikes. When you use it, you are billed, offering a quick

demand-fulfilling solution. In this case, if your site peaks while another has a valley, Web servers can be shared. Sometimes application servers can also be shared, with the help of partitioning, clustering and stateless Enterprise Java Beans. Participating in a large server farm that simultaneously hosts many large customers will give you economy of scale, resulting in lowered costs-per-transaction.

A 5% view-into-buy conversion rate is great! One "huge site" handles 4,000 peak transactions per second, with 110,000 concurrent users. Characterize server load by measuring transaction stress in the lab. Maintenance, skills and manageability must also be considered; you want diagonally linear scalability across a throughput graph.

"I learned the hard way: Don't generalize metrics," Dr. Chiu warned.

Optimize for Scalability

Traffic volumes grow quickly, are difficult to predict and are usually underestimated. Scalable architecture is required to cope with unexpected growth. IBM follows a best-practices methodology to optimize scalability:

1. Thoroughly understand your application environment.

2. Categorize the workload: large Web sites are distinguished by workload patterns, which drive scaling-technique selection and are also a factor in server selection and placement.

3. Determine the components most impacted: What percentages of users reach the first tier of machines? The second? The third? Reduce your SQL calls and beware of personalization's impact on performance.

4. Select the scaling techniques to apply: Faster machines, replicated machines, specialized machines, segmented workload, request batching, user-data aggregation, connection management and caching are all options. Implementing some of these techniques might impact your system-management policies.

5. Apply the scaling techniques: Multi-site implementations provide opportunities for applying system-wide scaling techniques.

For More Information

"This is not easy stuff. There is no free lunch." Dr. Chiu told the group. For more information, IBM has made the following white papers available at http://www.ibm.com/devloperworks/library/hvws :

Design for Scalability: Component selection and management techniques to make your site ready to adapt to increasing traffic

Planning for Growth: Methodology for capacity planning, identifying workload patters and configuring site infrastructure

Design Pages for Performance: Design practices that can reduce download time and improve resource utilization.

Web Site Personalization: Current and future techniques for personalizing your Web site.