SYS-CON

WEBSPHERE VS. WEBLOGIC: WHO DO YOU TRUST? PAGE 4

www.SYS-CON.com/WebSphere

WebSphere® JOURNAL

The World's Leading Independent WebSphere Resource

JULY 2004 VOLUME 3 ISSUE 7

CALL FOR PAPERS PAGE 42

web services edge conference expo Web Services Edge 2005 East

SECURITY

Propagating Security Context Across a Distributed Web Services Environment

BY SCOTT MORRISON • PAGE 40

ILLUMINATING WEBSPHERE

J2EE Application Performance Analysis

BY LLOYD HAGEMO & RAVI KALIDINDI

PAGE 28

FINAL THOUGHTS

The Management Options

BY CAMERON PURDY • PAGE 50

DISPLAY UNTIL SEPTEMBER 30, 2004

$9.99US $9.99CAN

0 74470 03422 3 07>

How Low Does BEA Have to Go?

BY JACK MARTIN

How low does BEA have to go before Oracle or, for that matter, anyone makes an offer for BEA? The answer may be lower than you think.

Someone, anyone buying BEA has been a perennial topic of conversation for the past couple of years. I've had people tell me Hewlett-Packard, Sun Microsystems, Oracle, Microsoft, and AOL all have good reason to buy BEA and give IBM a good run for its money. Not one of these companies has stepped up with an offer.

Microsoft could buy BEA for a sliver of its cash and turn the entire BEA portfolio into one giant .NET project, but it hasn't.

It turns out that Oracle did take a look at BEA as part of a list of nine potential acquisition targets in April '03 but passed on BEA and began hunting PeopleSoft. This came out at the Oracle/PeopleSoft antitrust trial in San Francisco, where the government is challenging Oracle's $7.7 billion hostile takeover offer for PeopleSoft. The document is purported to list nine companies and provides a detailed list of the benefits and drawbacks to a deal for BEA Systems. I guess Oracle decided the drawbacks outweighed the benefits, otherwise they would have made an offer.

As Oracle looked at BEA in April of 2003 BEA's stock bounced around $11 a share all month and closed at $10.70 on April 30, 2003. On June 25, 2004, the BEA stock closed at $7.98. That may not seem like a lot but it is. Think about it. A pack of cigarettes in New York City costs the same amount as one share of BEA.

If you were in the market to buy a company, that's great news if you really liked the idea of owning BEA because BEA is currently on sale for about 25% off its April 2003 prices – that's an incredible savings. Even a billionaire would take notice of that figure because it's over $1,000,000,000 off! The places where I shop usually don't have such big savings.

I know a billion dollars doesn't buy what it used to, but if you could buy a really good company and save a billion dollars wouldn't you?

You would think some billionaire or deep-pocketed public company or private equity firm would be foaming at the mouth for this once-in-a-lifetime deal, but they're not. It seems the smart money is either waiting for BEA to go even lower or they just don't care.

I know most of our readers could care less how low BEA stock may go, but if nobody wants to bet on BEA the company, why should anybody bet their business on BEA's software? An enterprise deployment filled with orphanware would be an incredibly painful experience to back out of.

Consider the class action lawsuit against BEA Systems, from the law offices of Marc S. Henzel, that claims, "the WebLogic 8.1 Platform was far from revolutionary and was not selling as claimed."

Or the Stull, Stull & Brody class action lawsuit that claims that BEA, "by issuing a series of material misrepresentations regarding BEA's business and prospects to the market between November 13, 2003, and May 13, 2004. As a result of these false statements, BEA's stock price traded at inflated levels during the Class Period, increasing to as high as $14 in early 2004, whereby the Company's top officers and directors sold more than $13 million worth of their own shares. Then on May 13, 2004, BEA reported disappointing first quarter results, citing the difficult selling environment and sales execution issues as the primary reasons. On this news, the Company's shares fell 30% to $8 per share." Yikes!! Attention billionaire shoppers, this is a 30% off sale.

Then Wechsler Harwood LLP announced that it has filed a federal securities fraud class action law suit on June 25, 2004. "The action, entitled Stroh v. BEA Systems, Inc., et al., Case No. 04-CV-2562 (SC), is pending in the United States District Court for the Northern District of California as defendants, the company, its Chairman, President and Chief Executive Officer, Alfred S. Chuang, its Executive Vice President of Worldwide Sales, Charles L. Ill, and its President of Worldwide Services, Thomas M. Ashburn." Maybe Charlie Ill should have stayed at IBM?

With nobody making a move on BEA no matter how low their stock goes and considering the seriousness of the class action lawsuits you have to ask, is BEA still a viable alternative to WebSphere.

I wonder what Scott Dietzen, BEA's chief technology officer, talked about when he delivered his keynote at JavaOne? Did he have anything revolutionary to talk about?

The author does not own any securities in any of the companies mentioned in this editorial.

• • •

For more on this issue, see the article on page 46.

Jack Martin, editor-in-chief of ***WebSphere Journal***, is cofounder and CEO of Simplex Knowledge Company, an Internet software boutique specializing in WebSphere development. Simplex developed the first remote video transmission system designed specifically for childcare centers, which received worldwide media attention; and the world's first diagnostic-quality ultrasound broadcast system. Jack is coauthor of ***Understanding WebSphere***, from Prentice Hall. jack@sys-con.com

Two years without a vacation.
The application's up. It's down.
It's up. It's down.

I'm to blame. Steve's to blame.
Someone's always to blame.

Not any more.

Get Wily.™

Enterprise Java
Application Management
1 888 GET WILY
www.wilytech.com

wily technology

Why WebSphere?

WebSphere vs WebLogic: Making a choice

TOM INMAN
VICE PRESIDENT OF PRODUCT MANAGEMENT AND MARKETING, IBM WEBSPHERE SOFTWARE

Jack Martin, editor-in-chief of ***WebSphere Journal***, recently spent some time chatting with Tom Inman, vice president of product management and marketing, IBM WebSphere Software, about WebSphere's market share relative to BEA's WebLogic, how IBM handles WebSphere sales, its customer and partner programs, and its plans for the future. This month, ***WJ*** offers part 1 of this interview, on market share and BEA.

WJ: A LOT OF PEOPLE HAVE BEEN TALKING ABOUT YOUR MARKET SHARE, HOW IT'S GROWING AND HOW BEA'S IS CONTRACTING. WHAT'S YOUR VIEW OF WHAT'S GOING ON IN THE MARKET SHARE WORLD RIGHT NOW?

TI: First of all, we are obviously happy that we are gaining market share. We think market share is a measure of client satisfaction. So that being the case, we believe that the more you can delight a client, the more they will buy from you. Word of mouth will then spread and others will buy as well. I believe what we are doing is a better job, better than our competitors, at satisfying the needs of our clients.

Having said that, I think that there are multiple variables that are behind our considerable market share gains. One of them is that across the board we have very good products. Our application server products, our portal products, our integration suites and business integration software products, our software development tools, in their own way, are all best of breed and are market share–leading. They've got strong momentum behind them. They each possess high quality and have won a lot of awards. And so our individual products are very, very competitive. We would not succeed with overall market share gain without that.

The other thing we're seeing is that the sort of problems that our clients are trying to solve today have grown more sophisticated in many ways, and the requirements they have tend to span a broader set of needs. They need an application server, they need a portal, they need integrated development tools, they need integration software, they need systems management tools, and they need business modeling tools to model the business process. They need the ability to connect a wide variety of existing ISV applications or custom-built applications. We have a broader, deeper, and more integrated portfolio than any of our competitors. So that is another one of the reasons that we are gaining such market share.

The third very important reason is the continuous growth in the network effect of WebSphere and the ecosystem of partners that we have. In a network effect what happens is the more market share you gain, the more it attracts the business partners, and the business partners sell more, which, in turn, generates more market share. Then, other business partners – who are in the business to make money – see more market share around WebSphere, so they start building their business around it. It's a network effect. It continues to build on itself. A great example of a network effect is why eBay has gotten so big, so quickly. What eBay does is to bring together buyers and sellers and enable very efficient markets for goods and services. The more buyers they attract, the more sellers they attract. It builds on itself.

So, to summarize, I believe our market share gains are a combination of the industry's best products; the broadest, deepest, most integrated portfolio; and tremendous momentum in the business partner ecosystem with an enormous base of highly satisfied clients.

WJ: THAT'S INTERESTING TO HEAR BECAUSE I BROUGHT A COUPLE OF INTERVIEWS WITH ALFRED CHUANG FROM BEA SYSTEMS WHERE HE SAYS THAT HE THINKS THAT COMPANIES LIKE IBM ARE TOO BIG NOW, THAT YOUR SALES PROCESS IS TOO COMPLICATED AND UPDATING LEGACY IS TOO HARD, THAT BEA CAN OUT-INNOVATE YOU GUYS. IT DOESN'T SOUND LIKE THAT'S WHAT'S HAPPENING THOUGH. IT SOUNDS LIKE YOU'RE OUT-INNOVATING THEM.

TI: Fundamentally, we are out-innovating them, you know, whether it is the patents that we have, or how

quickly our engineers turn new innovations into new product features. Or if you'd like to compare and contrast, we can take any series of BEA announcements and look at how all they've done is to try to mirror what IBM has already announced.

For example, in April of this year IBM had a significant set of announcements around service-oriented architecture, or SOA. In this announcement, we introduced some new offerings and laid out prescriptive approaches for clients and partners to realize value from implementing service-oriented architectures. In late May, BEA does the same.

IBM has been in the integration software market for years and we're the market leader. And now that IBM has pulled away as the clear market share leader in the application server market – where BEA has traditionally been focused and has been a "one-trick pony" – BEA tries to reposition themselves as an integration company. Another example is in the area of software to enable access to information and processes via pervasive devices. IBM has been delivering pervasive software for several years now and has significant assets and momentum around the pervasive space. In the areas of enabling enterprise clients to extend their middleware and application assets; to enable access from pervasive devices, to enable connectivity and integration to business processes and systems with RFID devices, or for embedded solution providers to enable mobile computing for an automobile company to embed infrastructure in their automobile to enable various mobile services. IBM has won numerous design patterns and tremendous share in these categories of pervasive computing. Now, years later, BEA is announcing themselves as a mobile software provider.

WJ: ANOTHER THING THAT I'VE BEEN HEARING COMING OUT OF THE BEA CAMP IS THAT THEY'VE BEEN TALKING ABOUT THEIR BUSINESS PARTNER PROGRAM AND IF MEMORY SERVES ME RIGHT, IBM, I BELIEVE, HAS THE WORLD'S LARGEST BUSINESS PARTNER PROGRAM. HOW DO YOU SEE THEIR BUSINESS PARTNER PROGRAM IMPACTING YOU IN THE NEXT YEAR?

TI: I don't see it impacting us.

WJ: WHY IS THAT?

TI: Well, first of all, I just read the cover story of a recent *VAR Business Magazine* that asks, "Can You Trust BEA This Time?"

WJ: SHOULD YOU?

TI: Well, I don't know.

WJ: WHAT DO THEY SAY?

TI: Based on the past, if history is any indication of the future, I wouldn't trust BEA. What I mean by that is, they've been extremely inconsistent in their partnering programs. In the past, BEA has decreed that the channel partner is strategically important, and then, shortly after this decree, has diverted their attention and resources away from the channel partners. They said they built an organization structure to cater to partners, and then they dismantled it because they were seeing declining license revenues and, you know, were not able to weather the storm. They have not been seeing any revenue gain. For most of their last 12–15 quarters or so, BEA has had declining software license revenues. When you're not growing, then you look at areas to cut back, and a number of times they've cut back on the partnering programs. License revenue growth is critical for a software company and is the fuel that feeds investments.

On the other hand, with all of IBM's growth, we have done nothing but focus on helping our partners to profit – that's what it means to be partner friendly. We have invested considerably in the infrastructure technology that partners need to support the creation of their own profitability. We've brought a broad and deep software platform to market with strong support for partners to build on, whether they are an ISV creating software for sale or they are a custom solution provider building and integrating solutions for their clients.

We've invested in the technology and we have invested in the technical support infrastructure needed to get partners enabled. For large partners that require direct support, or for greater volumes of smaller partners, we have invested in what we call innovation centers, or virtual innovation centers, either run by IBM itself or a select number of our large VADs to support the enablement of our partner ecosystem I discussed earlier. We've got a very impressive infrastructure in place to do the technical support of the partners needed either to enable their solution assets before they bring them to market or to assist them to succeed in their daily client engagements with their offerings.

Further, to help our partners generate revenues, we have invested considerably in co-marketing and co-selling programs. In February of this year, IBM announced an enhancement to our Partner World partner program designed to enable more and more vertical solutions for our clients across the 17 industries in which IBM goes to market – we call this Partner World Industry Networks. Partner World Industry Networks include business partners who have a strong capability that is relevant to one or more of the 17 industries. With this we have formed a deeper form of comarketing and coselling relationship. So, I think that the bar from IBM, as it relates to profiting from partnering with a vendor, is extremely high. In the case of BEA and partnering, I think the ball has been dropped a number of times. So, yeah, I think *VAR Business* has the right question – "Can you Really Trust BEA This Time?" Only history will tell, but their track record has not been impressive.

WJ: I'M READING AN INTERVIEW WITH ALFRED CHUANG FROM BEA ON CNET WHERE THEY ASKED HIM HOW BEA IS GOING TO ULTIMATELY BECOME A $3 BILLION COMPANY BY 2005 AND AS PART

OF HIS ANSWER HE SAYS, "WE HAVE TO LOOK AT WHAT TECHNOLOGY WE HAVE THAT ISN'T WORTH SELLING ON OUR OWN ANYMORE" AND THAT, THOSE ARE THE TECHNOLOGIES THAT THE DISTRIBUTION CHANNEL CAN REALLY HELP BEA IN SELLING. HOW ARE YOU DOING THE SAME THING? WHAT DO YOU SAY TO THAT?

TI: Well, basically what he said is, "It isn't worth it for us to sell, so let's give it to our partners."

WJ: THAT'S EXACTLY WHAT HE SAID.

TI: Again, basically he's saying, "I'll try to get our business partners to flog it. See if they can figure out to how sell something I can't." That's not IBM's strategy at all.

WJ: WHAT'S YOUR STRATEGY?

TI: Our strategy is one of market coverage. How do we get the right skills in the appropriate markets around the world, whether those are geographic markets or size-of-customer markets or targeting distinct decision-making units, such as IT versus calling on lines of business. With these, we need to be clear on the partner's value to our clients and what is our value to the partners. You know, how do we help the partners to profit? By the way, that makes it clear that our strategy isn't one of asking our partners to sell the junk we can't.

So, again, we start by gaining an understanding of the value of partners to the markets we intend to serve and then our value to the partners, in helping them to profit in the markets that we would like their help addressing. Then we recruit and enable them. We enable them technically and enable them with comarketing and coselling, for the right partnerships, to go after those markets that generate that value for the business partner and for IBM. So, we think of the world through the eyes of the end market we're trying to serve – the solutions they need and the skills and value they need to solve the problem. And we align with partners that can profit through their ability to solve those problems. They think about the end customer and they know what value they have to add out of our technology to solve that customer's problems. And, they too can profit.

WJ: SO YOUR BUSINESS PARTNERS ARE GETTING IT FIRST. YOUR BUSINESS PARTNERS ARE SELLING THE SAME PRODUCTS THE IBM SALES FORCE IS SELLING?

TI: The way that the partners sell differs across various markets. That's an example of what I mean by saying we start with the market, understand what it is the market needs from partners, and then seek partners that can help solve those problems and can profit off the usage of our technologies. So, we don't ask them to sell something we can't or won't.

"we start by gaining an understanding of the value of partners to the markets we intend to serve"

WJ: I WANT TO ASK YOU ONE MORE CHUANG/BEA QUESTION. ALFRED WAS ALSO ASKED IN THIS INTERVIEW, "YOU'VE TALKED A LOT ABOUT SERVICE-ORIENTED ARCHITECTURE OR SOA. OTHER COMPANIES, INCLUDING IBM ARE TALKING ABOUT IT AS WELL. IS THERE SOMETHING THAT YOU CAN SAY THAT WOULD DIFFERENTIATE YOU FROM THE OTHERS?" CHUANG'S ANSWER WAS, "WE ACTUALLY HAVE THE REAL THING. WE ACTUALLY HAVE A PRODUCT THAT IS AN SOA PLATFORM PRODUCT REALLY BUILT FROM SCRATCH. THEY'RE ALL PRE-INTEGRATED. YOU OPEN THE BOX AND YOU'RE READY TO GO AND IMPLEMENT A SERVICE-ORIENTED ARCHITECTURE APPLICATION. EVERYBODY ELSE HAS 400 PRODUCTS UNDER ONE NAME. THEY'RE NOT THE REAL THING. THAT IS A DIFFERENTIATOR ON ITS OWN." DO YOU HAVE ANYTHING TO SAY TO THAT?.

TI: I guess the only comment I'd make on that is that what BEA is trying to do is to take what they're doing around base application servers and add some function to it, call it a software platform, and then attempt to articulate that somehow that is the best platform for delivering an SOA. IBM, on the other hand, has been helping clients implement service-oriented architectures, applications, and systems for years. As the leader in message-oriented middleware, with WebSphere MQ and our message broker products, and as the leader in the definition of and shipment of Web services–enabled products, IBM has, literally, thousands of clients that have implemented service-oriented architectures. We've taken these experiences and mapped the patterns and best practices and are scaling out thousands of consultants armed with helping customers realize business value from SOA implementations or helping them to be self-sufficient in this endeavor. Again, I think this is another example of BEA following IBM's lead. Their behavior is flattering.

WJ: THANK YOU.

TI: BEA has to be the last of the major vendors to introduce anything around SOA, but it's not clear what they've done other than once again follow the leader.

WJ: THEY ARE GOOD AT SENDING OUT PRESS RELEASES.

TI: They study what we do then they try to mimic it. It's flattering, but I don't see how it's going to do anything to gain them market share.

In part 2 of this interview, WJ *will look at how IBM handles its sales and partner programs, and what it has in store for WebSphere in the next year.*

An interview with Devi Gupta

A Successful Ingredient Offers Choice

After talking with Tom Inman (see page 6), Jack Martin also spoke to Devi Gupta, director of Strategic Marketing at Prolifics, about IBM WebSphere and BEA WebLogic.

Devi Gupta is the director of Strategic Marketing and IBM Alliance Manager for Prolifics. She is a keystone to Prolifics and has fulfilled a variety of principal functions since joining the company in 1991. In her position, Devi directs the strategic positioning of the company, leads the effort to implement Prolifics' vision, and manages the company's key partner alliances. Under her guidance, Prolifics has made the critical transition from a product and services company to become one of only three WebSphere Service Providers retained by IBM and the 2003 winner of an IBM Business Partner Leadership Award.
dgupta@prolifics.com

WJ: DEVI, AS DIRECTOR OF STRATEGIC MARKETING FOR PROLIFICS, WHAT'S A TYPICAL DAY LIKE FOR YOU?

DG: Well, have you seen the IBM commercials that are on TV? I think a lot of companies that hear those messages feel that vision – the On Demand vision – is perhaps beyond them. I am Prolifics' messenger. One of my key goals, and the goals of my team, is to spread the word – let all the medium-size businesses know that Prolifics can make their On Demand solutions a reality – let all the larger enterprises know that Prolifics is *"the"* WebSphere expert. Our focus as a business and as one of IBM's top three core WebSphere Service Providers is not to sell customers WebSphere and then leave. Our focus is to stay and ensure that the company is successful with the software and that it has realized all its potential. I commend IBM for its dedication to customer satisfaction – putting together this specialist team solely dedicated to enabling customer success with WebSphere.

Another key goal for me is to help define Prolifics' business strategy. This involves market research, business analysis, and developing alliances. I dedicate a large portion of my week to working with our key strategic alliance, IBM. I work with my IBM counterparts to create joint campaigns that offer our software and services to companies around the world. In a good partnership it is essential to communicate and work together toward a common goal where both parties – as well as the customers they service – win. We have that good partnership with IBM and our customers.

WJ: WHO IS YOUR IBM COUNTERPART?

DG: On a day-to-day basis, I work with our Business Partner representatives, Deanna Landivar and Mike Lowry. Our executive sponsor at IBM is Tom Inman, who has been incredibly supportive of us for five years now.

WJ: WHEN YOU'RE LOOKING TO DRIVE AWARENESS, WHAT TYPES OF COMPANIES ARE YOU WORKING WITH? WHAT INDUSTRIES ARE THEY IN? WHAT SIZE ARE THEY?

DG: Our focus has always been the highly transactional industries, so really we have a strong presence across several industries. In fact, we located our headquarters in the Wall Street area of New York City because of our strong finance and insurance backgrounds. However, we do service customers across several other industries – telco, health care, government, manufacturing. Regarding size, they tend to range from the medium-size business all the way up to the large organizations. At least 60% or more of our business is dedicated to the medium-size business and the rest is dedicated to servicing enterprise-size customers like AT&T and UPS.

WJ: IS YOUR COMPANY SPECIFIC IN ITS WEBSPHERE VALUE?

DG: Well, I mentioned earlier that we are one of three core WebSphere Service Provider partners. Let me explain. Because we are ranked a Level 4 & 5 WebSphere partner – IBM's highest for technical skill – we are part of a select group of highly

specialized experts that IBM brings in to service its most challenging WebSphere customer requirements. We are experts in the field – experts in WebSphere Portal and Business Integration solutions. We have serviced over 300 WebSphere customers on the Eastern Seaboard alone within the past three years. So, what we really bring to the table is our years of expertise that go into understanding customers' requirements, solving their difficult problems, and understanding how to make their software projects a success. Having serviced customers for 26 years now, we keep our IT solutions practical – understanding and scoping to exactly what the customer needs to do, not beyond what they need to do.

WJ: HOW BIG A HEAD COUNT DOES YOUR COMPANY HAVE?

DG: We have around 120 employees.

WJ: OF THOSE, HOW MANY ARE JAVA DEVELOPERS?

DG: Over 70.

WJ: SO OVER 50% OF YOUR COMPANY IS A JAVA SHOP. THAT'S INTERESTING. DO YOU WORK WITH ECLIPSE? WHAT IDE DO YOU USE?

DG: We typically use WebSphere Studio Application Developer. But of course that is augmented by each consultant's previous years of expertise with native programming and other IDEs.

WJ: HOW DO YOU FIND USING WEBSPHERE STUDIO ON ACTUAL ENGAGEMENTS? IS IT FLEXIBLE? DOES IT WORK FOR YOU?

DG: Yes, very flexible – it works well for us. We use it on all of our WebSphere and portal customer engagements, and have even used it to build some of our own internal run-the-business applications. I think the beauty of WebSphere Studio is that it hooks into the Eclipse Framework, allowing it to extend its functionality, and that it is so open. Obviously, it is ideal when developing WebSphere applications; however, I believe that we've been able to utilize the software when a customer is in a non-WebSphere environment or when migrating a customer to WebSphere from another application server.

WJ: WHICH ONES HAVE YOU USED IT ON?

DG: Tomcat and BEA WebLogic.

WJ: WHEN A CUSTOMER COMES TO YOU LOOKING FOR A WEBSPHERE DEPLOYMENT, DO THEY HAVE A SPECIFIC SET OF BUSINESS CHALLENGES IN FRONT OF THEM OR IS SOMETHING BROKEN?

DG: A little bit of both. Our customers might come to us during a critical moment when they need to seek out highly specialized experts. Or often, particularly with our portal or business integration customers, they come to us with a specific business problem, such as the need to create a business-to-employee portal, or to eliminate redundant systems and processes within their organization.

WJ: IT SEEMS THAT YOUR COMPANY HAS BET THE RANCH ON WEBSPHERE FOR WHAT YOU SPEND THE BULK OF YOUR WORK DOING. CORRECT?

DG: That's true.

WJ: WHY?

DG: Well, I think we are unique in that way. We've been in business for 26 years and for most of our history we kept a neutral stance. However, with technology being as complex as it is, it is implausible to say that we are the experts at everything. We can provide better service to our customers by being the best at one thing rather than good at everything. So five years ago we evaluated and spoke with some of the major players in the J2EE space, and decided to strategically align ourselves with IBM.

WJ: WHAT DID IBM WALK IN WITH THAT WAS BETTER THAN ANYBODY ELSE?

DG: Two primary things. One reason we selected IBM, which was evident right away, was the level of importance, energy, and investment that they place on business partners as a channel, particularly as compared to other vendors. The second is that what IBM had to offer to our customer base was of great value. They're not just a middleware vendor. They're not predominantly a database vendor. What we found them to be was truly a total solutions provider – they had software technology that satisfied pretty much all of the requirements toward delivering a transactional system for a customer, not to mention their hardware offerings, services arm, and business partner community solutions. So, we felt that we could offer our customers a solution to do anything they needed to do today and support anything they may need to do in the future.

WJ: HAVE YOU EVER DONE ANY WEBLOGIC WORK?

DG: We have. Yes.

WJ: HOW DID YOU FIND WORKING WITH BEA'S PRODUCTS AS OPPOSED TO IBM'S?

DG: BEA's software is good. From a technology perspective, if we compare the WebLogic app server to the WebSphere app server we are comparing two products that are quite similar. I don't think a customer should base their decisions simply on a feature-by-feature product comparison. Particularly when you know that the two competitors will catch up with each other shortly. With the level of investment that IBM puts into WebSphere, you have to know they will keep advancing and complying to the latest J2EE specs. From

a portfolio perspective, I would say that WebSphere wins hands down – they have collaborative portal solutions, complex business integration adapters, support for voice activation and pervasive devices – whatever a business needs for its future. It's a comprehensive solution.

WJ: SO, IN YOUR VIEW, FROM A TECHNICAL PERSPECTIVE, BEA WEBLOGIC AND WEBSPHERE PRETTY MUCH ARE ON A PAR IN THE COURSE OF ANY 12-MONTH PERIOD AS THEY MOVE BACK AND FORTH WITH EACH OTHER.

HAVE YOU CONVERTED ANY CUSTOMERS FROM WEBLOGIC TO WEBSPHERE?

DG: We have. We've converted more than 20 customers from WebLogic to WebSphere. In fact, we've also converted several customers from Sun's iPlanet and OracleAS.

WJ: WHY DO YOU THINK THOSE CUSTOMERS CONVERTED?

DG: For a number of reasons. They can reduce their overall maintenance costs by moving to WebSphere – that's been a tremendous factor in the decision making. Partly because IBM's pricing model is very competitive and partly because IBM provides all the productivity tools to speed development and decrease costs.

WJ: I'VE HEARD THAT SOME PEOPLE WHO DO WEBLOGIC INSTALLATIONS ACTUALLY USE WEBSPHERE'S APPLICATION DEVELOPER TOOLS.

DG: Yes, that's right. Even though customers may be using WebLogic as an application server they are still turning to IBM to provide the comprehensive development platform. Another reason I can give you for why a customer chooses to migrate from BEA WebLogic to IBM WebSphere is because they are making a vendor decision. They're looking for a company that is going to be around in a few years – one with credibility and long-term viability.

WJ: ARE YOU SAYING THAT SOME CUSTOMERS YOU HAVE WORKED WITH HAVE ACTUALLY CHOSEN WEBSPHERE OVER WEBLOGIC BASED ON WHAT COULD HAPPEN TO BEA AS A COMPANY?

DG: Well, there certainly have been rumors and suspicions in the marketplace as to whether or not BEA would be acquired. If that were the case, what is the long-term strategy for the technology?

WJ: YOU MEAN THERE'S A CONCERN THAT IF MICROSOFT BOUGHT IT, MAYBE THEY'D TURN THE WHOLE THING INTO A .NET SHOP AS OPPOSED TO ORACLE BUYING IT AND TURNING IT INTO A DATABASE?

DG: Well, I don't think there is any synergy between Microsoft and BEA. If Oracle or HP acquired BEA, the question would of course be how WebLogic would be integrated with Oracle's and HP's own respective application server and what their customers could expect as the product's future.

WJ: INTERESTING. SO YOU HAVE A LOT OF DEEP-THINKING CUSTOMERS THAT LOOK FORWARD IN TIME BEFORE THEY MAKE THEIR ACTUAL DECISION.

DG: It's a very important decision to make. Our customers are choosing a corporate standard. The technology will be supporting them for at least the next five years or more and they need it to grow with the unexpected changes in their business. Many customers are going through mergers and acquisitions, having to comply with government regulations, and will create many other business requirements. So, it's an extremely important question for these customers to answer, and I think they take it very seriously, both with evaluating the technology and also beyond that to evaluating the vendor.

WJ: LET ME ASK YOU THIS. IN THE SPIRIT OF BEING FAIR TO OUR FRIENDS OVER AT BEA, HAVE YOU LOST ANY OF YOUR CUSTOMERS FROM WEBSPHERE TO WEBLOGIC?

DG: We haven't necessarily lost customers from IBM to BEA. What we have seen are a few customers that might already have BEA tech-

> **"What we found [IBM] to be was truly a total solutions provider"**

nology embedded, and although they may be interested in moving to WebSphere, they make the decision to not move. They may be forced, because of internal timeframes, to stay put.

WJ: SO YOU'VE SEEN CUSTOMERS HOLD PAT IF THEY ALREADY HAVE IT, BUT YOU HAVEN'T HAD ANY OF YOUR CUSTOMERS RIP WEBSPHERE OUT BY THE ROOTS AND REPLACE IT WITH WEBLOGIC?
DG: No. We haven't.

WJ: YOUR COMPANY BELONGS TO A VERY, VERY SPECIAL CLUB. IBM COULD HAVE PICKED ANY THREE COMPANIES AND ALMOST ANY THREE TECHNOLOGY COMPANIES IN THE WORLD WOULD HAVE HAPPILY TAKEN THE ASSIGNMENT; WHY DO YOU THINK IBM PICKED PROLIFICS?
DG: That's easy. We're proven and we've been proven for 26 years.

WJ: PROVEN AT WHAT?
DG: We have a 96% success rate measured by an independent analyst organization of delivering solutions to our customers on time, on budget, and within specifications.

WJ: DO YOU REALIZE THAT TWO-THIRDS OF THE IT PROJECTS, ALL IT PROJECTS THAT HAPPEN IN THE WORLD, FAIL EITHER BECAUSE THEY MISSED THE BUDGET, THEY MISSED THE TIMEFRAME, OR THE THING JUST DIDN'T WORK?
DG: I do. In fact the industry standard success rate is somewhere around 23%. It's shocking really – and because of that we make every effort to ensure that a customer's application successfully deploys. We are incredibly proud of our work.

WJ: HOW DO YOU HAVE SUCH SUCCESS? NOW I UNDERSTAND WHY IBM PICKED YOU. NOW I WANT TO KNOW THE SECRET SAUCE. WHY ARE YOU GUYS SO GOOD?
DG: The secret is our people, our methodology, our practical experiences. The secret is how we learn from our practical experiences and apply it back into the success of other projects. The rest of our secret is a secret.

WJ: YOU SHOULD BE VERY PROUD TO WORK THERE AND SO SHOULD EVERYBODY ELSE.

Tom Inman (see page 6), also spoke to us about Prolifics' relationship with IBM.

WJ: I UNDERSTAND THAT YOU ARE PROLIFICS' EXECUTIVE SPONSOR OVER AT IBM AND THEY ARE ONE OF IBM'S THREE CORE WEBSPHERE SERVICE PROVIDERS. CAN YOU TELL ME, WHAT MAKES PROLIFICS SO SPECIAL?
TI: Prolifics is one of three companies; Devi is referring to one of the business relationships that they have with us. I say this because we have several relationships with each other. They are what we call a "retained services partner," which means essentially we have them on a contractual relationship with us whereby when we deliver services and we are the provider in the eyes of the customer quite often we will have a Prolifics person or team on that project. That's a testament to the quality of the skills and reputation of the individuals that are employed by Prolifics. They are truly world-class professionals in terms of their skills and the professionalism that they deliver. We are willing to put an IBM brand name on their work.

WJ: THAT'S UNUSUAL.
TI: So, that is one of the relationships. There are others. They do client integration projects on IBM contracts and they do integration business on their own. They also resell. Further, they have a software products business that is very complementary to that of WebSphere. But, I think what makes them particularly capable is that they have been in the business of solving fairly complex customer problems for years. They have been a custom integration solution provider for years and they have a track record of delivery. They have been in business for over 20 years, so probably one of the things that make, them unique is their longevity. You know they've been in the market for a while and that gives their customers a lot of confidence.

They have very, very satisfied client relationships. To the IBM sales team they have a reputation for delivery. When you call on them, they bring world-class experts to the table who know exactly what they are talking about. They are very impressive in front of clients. When you bring them in on a project, they simply deliver. It's that level of successful delivery that is their reputation. So, when our field team thinks about a trusted partner, they think Prolifics. When a client engages Prolifics, they are delighted with the results.

WJ: SO THEY FIT THE BILL?
TI: Clients and IBM can count on Prolifics. Since Prolifics was once an important business partner of BEA's, they are living proof of what we discussed earlier. That BEA is not a company that you may want to trust as a partner. It is BEA's lack of real partnering abilities that helped fuel the mutually profitable relationship between Prolifics, IBM, and our clients.

What//: are your Web applications saying to your customers?

diagnosys

Are your Web applications open 24 hours? Get the tips and best practices you need to keep your apps running at peak efficiency 24 hours a day with WebApp News, the free newsletter from the performance experts at H&W.

Sign up today at:
www.hwcs.com/wsj10.asp

You might as well be closed for business if your Web applications aren't running at top speed 24 hours a day. With poor Web performance costing businesses $25 billion a year, you need to find problems, and you need to find them fast.

You need DiagnoSys.

DiagnoSys is the first intelligent performance management solution that analyzes your J2EE and .NET applications from end to end. DiagnoSys starts analyzing quickly, gathers data directly from all application resources, and proactively finds and helps fix the real cause of problems – before they cost your business big time.

So when it's absolutely essential that your Web applications say "Open 24 hours," trust DiagnoSys.

H&W
www.hwcs.com | 1.800.338.6692

Bridging the Enterprise Computing Gap For 25 Years

An interview with George Kosmides

A Leader with New Customers

Jack Martin also spoke to George Kosmides, president of Noospherics, about their IBM relationship.

WJ: WHEN WAS NOOSPHERICS FOUNDED?
GK: Noospherics was started in 1997.

WJ: WHY DID YOU START, OTHER THAN THE OBVIOUS, NOOSPHERICS?
GK: We had a lot of experience with object technology from the mid-1980s until the mid-1990s, primarily in Smalltalk, and had a lot of object design capability. We applied this in the Java and the Web worlds and formed Noospherics specifically to work on an approach-level, object-based project. The technology primarily translates to J2EE.

WJ: HOW MANY PEOPLE WORK AT NOOSPHERICS?
GK: About 25.

WJ: WHAT IS YOUR CURRENT SPECIALTY?
GK: Right now we work primarily on project development services for small to medium-sized businesses doing all of the enterprise integration that's required to apply WebSphere, MQSeries, and WebSphere Portal to enterprise architecture application design and development, and also training and mentoring.

WJ: DO YOU FOCUS ON ANY SPECIFIC INDUSTRY?
GK: No. We cover a broad range of industries including banking, insurance, retail, financial services, federal government, state government, universities.

WJ: SO YOUR SPECIALTY IS DOMAINS SPECIFIC TO THE TECHNOLOGY AND TRANSFERABLE TO WHEREVER THE CUSTOMER NEEDS TO GO?
GK: Exactly.

WJ: DO YOU DO ANY TROUBLESHOOTING, FOR EXAMPLE IF A CUSTOMER HAS A REAL PROBLEM ON THEIR HANDS. DO YOU COME INTO THE JOB AT THAT POINT?
GK: We can work with the entire project so we come in at different points. Sometimes we start at the beginning with a project through the architecture, the training and mentoring, the analysis, and the design used; or we take the project all the way through. Other times we come in at the middle when they need some help or they need to supplement their work. Other times we'll come in like a field team, just dropping in during tuning or troubleshooting or solving a number of problems; we really cover the whole spectrum.

WJ: ARE YOU SEEING A LOT OF ACTIVITY IN THE BUSINESS INTEGRATION SPACE?
GK: Right now it seems pretty clear that there are a number of different technology solutions that have been working for a while. Existing systems and several generations of what some call "legacy systems" need to be integrated and updated. People need a common middleware solution whether it's an ERP system or a CRM or a supply chain system. What we're seeing a lot is that people have solutions, but they need them updated and integrated, and that's where we come in. Typically at the center of these solutions will be WebSphere, WebSphere Portal, and MQ.

WJ: FROM WHAT I'VE HEARD YOU LEAD WITH WEBSPHERE WHEN YOU MEET A NEW CUSTOMER. WHY IS THAT?
GK: The WebSphere family is comprehensive enough now that it really has all of the pieces that you need for an enterprise solution. Sometimes we'll lead with WebSphere Application Server, and sometimes we'll lead with WebSphere Portal Server as an employee portal, for instance. We often use WebSphere Business Integrator to tie into existing business systems.

WJ: DO YOU WORK WITH ANY OTHER APPLICATION SERVERS? IF A CUSTOMER ASKS, WILL YOU?
GK: Yes. We strongly believe that the underlying principle, the underlying foundation for a solution, is what matters and that's the object model, the business model, the business process. The most important thing in any project is to get that resolved.

George Kosmides is an enterprise architect and cofounder of Noospherics Technologies who has been working with object technology for the past 15 years. He has extensive experience with object-based architectures in the banking, insurance, mortgage, medical information systems, health care, process control, retail, and government sectors.

To get that business process captured in your model using model-driven architecture is the "crown jewel." That's the core thing. Then how do you apply that to a technology solution? How do you actualize the model into productive solutions? One way to map this model and architecture is to use a J2EE enterprise solution. That being said, once we have that established we could use BEA's WebLogic. We could also use IBM WebSphere. A huge portion of the time we find that we go to WebSphere.

WJ: WHY IS THAT?

GK: First I'm going to say that BEA has a great solution. WebLogic is what I first used for J2EE application servers and I like their approach. It performs well and we've gone through the various generations from WebLogic 4 up to the current WebLogic 8.1. We've worked with WebLogic and it's a good product. WebSphere matches every feature and every capability that WebLogic has, but WebSphere and the WebSphere family in most situations that we run into offer more. We've developed a complete integration capability with MQ. We've got a complete portal solution, not just one piece of a portal solution, but we've got the collaborations that we did through Domino. On the back end we've got the ability to tie into the mainframe and the rest of the enterprise. Again, the WebSphere family has the fullest set of solutions. WebSphere at the core of the enterprise is a remarkable solution.

WJ: DO ANY OF YOUR CUSTOMERS ASK YOU FOR JBOSS?

GK: They hadn't until the last six months and now we have customers asking us very tough questions. "I have 500 stores, or I have an entire division and why should I pay for 10 licenses when I could use JBoss instead?" That's a very important question and now with JBoss we are having a harder time answering that. I think that's something that the industry needs to address. This question addresses the real business value of the vendors' offerings.

WJ: DO YOU THINK THE APPLICATION SERVER MARKET WILL GO TO OPEN SOURCE?

GK: We've had a number of debates within the company and there are two sides to it. You know there are people who say, absolutely, JBoss has been big and for that matter three-quarters of typical customer needs are going to be addressed with some sort of open source app server solution. The other side of the discussion is that open source core servers are not built on a sustainable business model. Where is the R&D going to come from? Who is going to make the crucial investments? IBM has pushed a lot of the J2EE technology so far because there was money to be made by it. They are pushing new frontiers with the portals because there is money to be found there. I don't know what the business model is on that end, so I think that the jury is still out. We're watching it closely and it's remarkable how many of our customers in the last six months have asked us about it as a possible solution.

WJ: OF ALL THOSE WHO ASKED ABOUT IT, HOW MANY HAVE ACTUALLY DEPLOYED?

GK: Zero.

WJ: WHY DO YOU THINK THAT IS?

GK: They're waiting, and I'm talking about WebLogic and WebSphere customers.

WJ: OUT OF EVERYONE YOU SPEAK TO ABOUT JBOSS, YOU'RE SAYING THAT AFTER THEY GET DONE TALKING ABOUT IT, ZERO MOVE FORWARD?

GK: Exactly.

WJ: WHAT DO YOU THINK THEY'RE ACTUALLY WAITING FOR?

GK: They're waiting for investment and support. They've delved into open source and a lot of them have limited success. A lot of them are running with Struts. Insurance companies that told us three years ago they did not want our Struts training or Struts mentoring with WSAD are using Struts, and you know, if the JBoss solution is robust enough, there's going to be a very interesting, very compelling argument for considering it. The reason that they are not yet is because they are waiting. They are curious to see the robustness, the clustering. You know you get a lot of great things from WebSphere and the clustering and performance and support; will the JBoss solution be able to provide all of that? The jury is, as far as the customers are concerned, still out, but there is a lot of interest there.

WJ: I HEAR YOU'VE GOT A VERY INTERESTING BACKGROUND THAT NOT ONLY WORKING IN INFORMATION TECHNOLOGY. YOU'VE GOTTEN INVOLVED WITH SOME VERY EXOTIC ARCHITECTURE. CAN YOU TELL ME JUST A LITTLE BIT ABOUT THAT?

GK: Twenty-five years ago, as an electrical engineer, I was heavily involved in solar energy research and I've been involved with the project in Arizona started by an architect named Paolo Soleri. Soleri started building a prototype of a city, a three-dimensional city that doesn't sprawl across the land and doesn"t use cars. He named that prototype city Arcosanti and I was involved with it then as an engineer and I remained involved doing pro bono work for the project.

WJ: DID THEY EVER BUILD ARCOSANTI?

GK: It's about 5% complete. It's located south of Sedona and north of Phoenix. There are 60,000 visitors each year who observe the construction.

WJ: SO NO ONE IS LIVING THERE YET?

GK: People do live there. There are 50–80 staff people living there and building it. And I'm actually headed there next week. I take my family every year and we participate and work on several projects there.

Retrieving, assembling, and presenting information

J2EE Caching

BY HARI KANANGI

Hari Kanangi works as a consultant for Stratus Solutions Inc., helping clients design, develop, and deploy J2EE-based solutions. He is a Sun Certified Enterprise Archect, J2EE architect, and a WebSphere Certified specialist. In the past five years, Hari has been concentrating on WebSphere, J2EE, and Java-based technologies.
hari.kanangi@ieee.org

Most Web applications are typically based on the presentation of information, meaning that functional operations pertaining to retrieving, assembling, and presenting information in the form of content and data largely outnumber functional operations that actually modify the information.

For example, browsing detailed product information in an online store occurs exponentially more often than updating the product information. Caching offers critical benefits to the application that include, but are not limited to:

- ***Faster response times and increased application responsiveness:*** Requests are serviced faster, as pre-cached information is served up without having to query the data store and apply complex business rules and calculations to the result set.
- ***Increased throughput:*** Applications designed to take advantage of caching have much higher throughput rates because it is not necessary to retrieve data and perform complex calculations on the data for every request running on the same physical hardware infrastructure.
- ***High availability:*** If back-end components go offline temporarily, previously cached information can still be served up to the client.
- ***Scalability:*** Caching in the presentation and business tiers of the J2EE architecture greatly helps to reduce the consumption of system resources, allowing the applications to be scaled horizontally more easily when the need to support a higher client load arises. The application itself will need to be well designed to ensure horizontal scalability.

Caching and Web Application Architecture

Figure 1 depicts a widely used robust Web architecture. There are several opportunities in this architecture where information can be cached efficiently and served up.

Following is a brief description of the various points at which caching can be accomplished:

- ***WebSphere Edge Server in a reverse proxy configuration:*** WebSphere Edge Server is the initial point of contact for every request within an organization's DMZ. The Edge Server Web Traffic Express (WTE) component caches responses of dynamic content like JSP pages and servlets. An Edge Server adapter module is installed in WebSphere Application Server (WAS), which results in all dynamic content responses being modified by the adapter before returning to the Web server. The WebSphere Edge Server and WAS work together in invalidating caches to keep stale information from being served up in the cache.
- ***Web server caching:*** Most Web servers allow static files being served to be cached. Caching dynamic content can be done either by the Web server itself or by the WAS Web Server plug-in component. Both of these scenarios are typically limited in terms of cache invalidations and can be used as simple caching mechanisms with time-based invalidations.
- ***Presentation and business tier caching:*** These two logical tiers can be physically located in the same virtual machine, different virtual machines, or even on different nodes. Caching here offers the most flexibility of control over the cache in terms of security, configuration, and maintenance of the cache at run time.
- ***EIS caches:*** WAS caches prepared statements and data itself can be configured to be cached in most enterprise information stores.

The general rule of thumb is to cache information as close to the source of the request (the client) as possible to get the best caching result. Several other factors need to be considered carefully along with response times, including security (caching of sensitive information in the DMZ will have enterprise security architects up in arms), hardware and software limitations, caching requirements and level of control required on the caching engine in terms of configuration, monitoring, etc.

Presentation and Business Tier Caching

Content being cached in these two tiers of the J2EE Web application architecture varies. For example, the presentation tier's dynamic components (servlet and JSP output) are usually prime candidates to be cached in part, like portlet output that forms part of a bigger page. Alternatively, in the business tier, objects that result from a combination of complex queries and business process calculations on result sets are typical candidates for caching.

CACHING CONSIDERATIONS

One of the key elements to designing a successful application caching mechanism is to understand the caching requirements. Some important questions that will help determine the design of caching mechanism include:

- What information is to be cached?
- Will the information be cached in the presentation tier, business tier, or both tiers?
- How do you invalidate the information in the cache to ensure stale information does not exist in the cache?

Along with these questions, it is also important to understand:

- What is the size of an object/dynamic content being cached? What will be the approximate size of the entire cache? Will the objects be cached entirely in memory or be shared between memory and disk?
- What algorithm will be used, if the cache is exhausted, to move content/data out of the cache?
- Will the cache need to be clustered between various nodes in a clustered environment? If it does, what mechanism will be used to accomplish this? When clustered, will there be a single remote cache or will each node maintain its own local copy of the cache to synch?
- How does clustering impact network traffic – will the clustering mechanism used result in a noticeable increase in network traffic just to maintain the consistency of the cache creating network bottlenecks, or will the cache implement an efficient mechanism that results in minimal network traffic to maintain the consistency of the caches in a cluster?

SAMPLE APPLICATION

We will use an example to describe asynchronous caching solutions of which most Web applications can take advantage. The sample application is to build part of the ACME online store. (the sample EAR file with source and configuration instructions is available as a download). We will tackle one simple use case as shown in Figure 2.

The architecture of the sample application (as shown in Figure 3) is made up of one WebSphere cell containing two nodes: Node A and Node B. Node A runs the WAS instance hosting the application that services the end clients of ACME, allowing clients to browse, view, and buy products from ACME using their Web site. Node B runs a JMS Server component (either the Embedded Messaging Server running under its own dedicated JVM or WebSphere MQ) and WAS hosts an application that lets ACME application administrators update product details.

Open Symphony's OSCACHE is used as the cache engine. Alternatively, other caching products, such as WebSphere's built-in dynamic cache, can be effectively plugged to achieve the same result. The OSCACHE design allows caches to be created and manipulated using JSP tags and/or an API. The tag library is used to cache parts or entire dynamic Web page contents. The API can be used to manipulate the caches in both the presentation and the business tiers that are built after complex processes. The example will use the API to cache product value objects in the business tier.

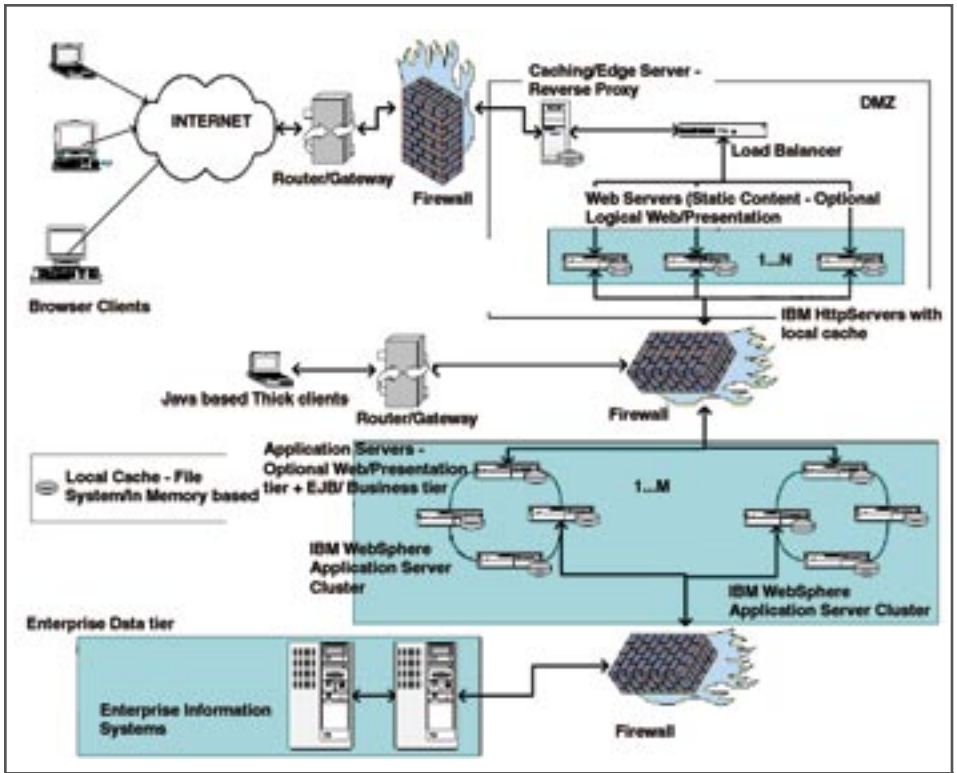


FIG 1: A ROBUST WEB ARCHITECTURE

BUSINESS TIER CACHING

Figure 4 demonstrates how asynchronous caching using JMS and message beans can be employed in the business tier of an enterprise application. In the figure, logical boundaries (the logical components can physically be in the same virtual machine, different virtual machines, or even different servers) are indicated by rounded rectangles. A solid line connecting two shapes represents a request while a dotted line connecting two shapes represents an optional request that may or may not be required.

The most notable steps to caching in the business tier are:

- ***Steps 1–3:*** An HTTP request comes from the browser requesting the product details handled by a controller servlet. The request is forwarded to the business tier (this is a remote request if the Web container and EJB container reside in different virtual machines) through a business delegate.
- ***Step 4:*** In well-designed applications, a session EJB facade handles requests to the EJB container and a value object is returned back to the Web container as requested data. This allows a framework to be built that could allow the facade to check the cache on every request in the EJB container. If the requested value object is already in the cache, the data is retrieved from the cache and returned immediately. Effectively, neither the data is retrieved from the EIS tier nor are complex operations performed on the retrieved data, enhancing the health of the various servers and the application itself.
- ***Step 5:*** The product details are not found in the cache, so they have to be retrieved from the EIS store. All operations are performed on this data. The resulting product details value object is stored in the cache immediately. Future requests for details will be served up from the cache.
- ***Step 6:*** The product administrator updates the information using the product update component.
- ***Step 7:*** The product update component updates the data store and publishes an asynchronous JMS message to the product update topic. It is important to ensure that the JMS message is sent on product updates; otherwise, stale data will be pulled from the cache. To make sure this does not happen, the update transaction is processed using a two-phase transaction (XA enabled).
- ***Steps 8–9:*** The update cache message bean is configured to be a subscriber to the product update topic and is triggered when the product update JMS message is published successfully. The message bean clears the product entry from the cache and optionally reloads the cache with the updated information. If the optional step is not done, the product detail is cached on the next request.

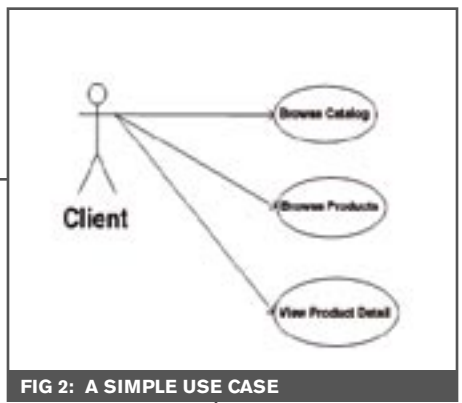


FIG 2: A SIMPLE USE CASE

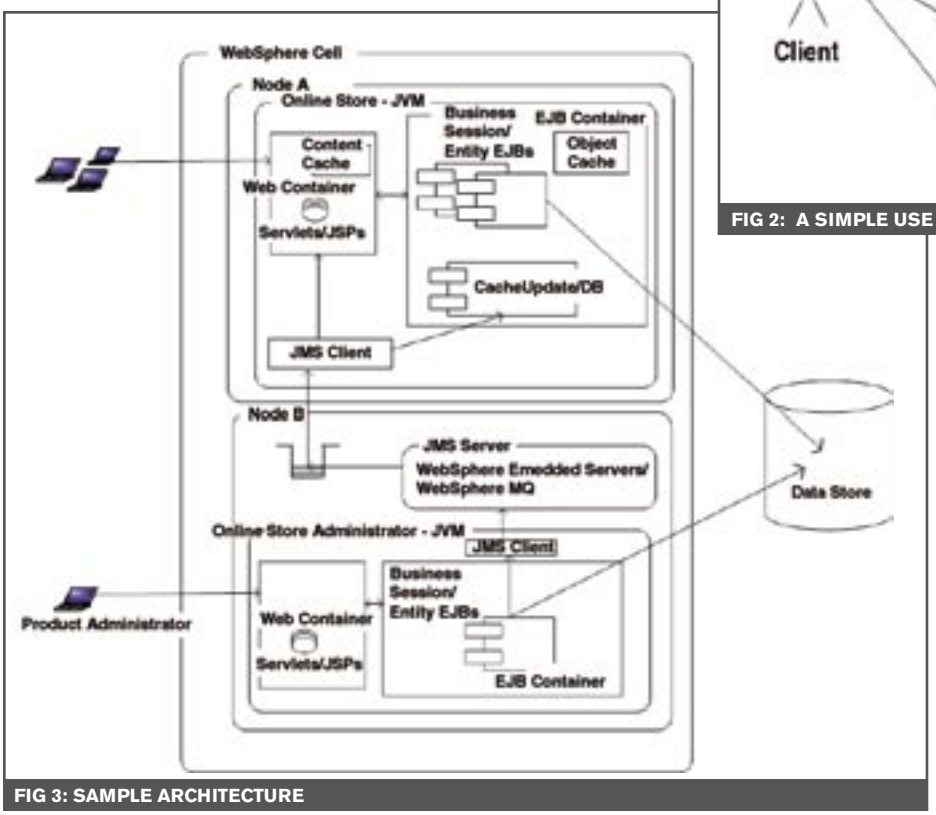


FIG 3: SAMPLE ARCHITECTURE

The caching API is similar to any data structure supporting key-value pairs (hashtable, properties, maps, etc.). The OSCACHE API throws an exception if a particular entry is not found, forcing clients to handle exceptions in all retrievals from the cache, which could lead to cumbersome operations. Utility wrapper classes that provide a client-friendly cache API will alleviate this issue.

Web Tier Caching

CACHING DYNAMIC CONTENT

The OSCACHE caching tag libraries provide custom tags that allow applications to cache the contents of entire (or selective parts of) JSP pages (and servlets). In this example, we will use JSP pages, but the API

can be used to simulate the same behavior for servlets. The OSCACHE custom tag "cache" is specified as:

```
<cache:cache key = ' <%= request.
getParameter ( "productId" ) %>'
refreshpolicyclass = "com.stratus.
sample.web.cache.ProductCacheRefr
eshPolicy" refreshpolicyparam ='
<%=request.getParameter("productI
d")%>' >
<!--RETRIEVE PRODUCT DETAILS AND
DISPLAY DETAILS HERE -->
</cache>
```

The product ID is used as a unique key to store the rendered content for each product. The refreshpolicyclass attribute specifies a class that implements the needsRefresh() method of com.opensymphony.oscache.web.WebEntryRefreshPolicy interface. This method will be called on every request by the OSCACHE framework to verify whether or not the JSP code between the tags needs to be processed. If the needsRefresh() method returns true, the product details will be retrieved from the business tier. If the needsRefresh() method returns false, the rendered content will be delivered out of the cache. The value of refreshpolicyparam will be made available to the ProductCacheRefreshPolicy class to help needsRefresh() method in making the decision on refreshing the cache.

INVALIDATING DYNAMIC CONTENT

We looked at how the custom cache tag can be used to cache dynamic content; we will now look at how the needsRefresh() method decides whether or not the cache needs to be invalidated for a product servicing a request.

In the Web tier, the message bean could have been substituted by registering a javax.jms.MessageListener to a TopicSession, but one thing to note here is that, although the JMS API allows registration of message listeners for particular topics using TopicSession.setMessageListener(), this cannot be used, as most application servers do not support the registration of message listeners in the Web tier due to connection pooling of JMS connections. In the upcoming J2EE v1.5 specification, this rule is mandated. Therefore, we will create a JMS message topic subscriber thread that is started when the Web application is started. The three will block on the TopicSubscriber.receive () method until "product updated" messages are published by the administrative application. The thread maintains and asynchronously updates a static list of updated product IDs based on the incoming messages. The needsRefresh() method checks this list of updated products, which is updated asynchronously by the subscriber thread.

Figure 5 demonstrates how asynchronous caching can be employed effectively in the Web tier of an enterprise application. The business tier is left untouched and is the same as before. The most notable steps with respect to caching in the business tier are:

- ***Step 1:*** An HTTP request comes from the browser requesting the product details handled by a controller servlet and forwarded to the JSP page handling the product details.
- ***Steps 2–3:*** The cache tag executes the refresh policy and invokes the needRefresh() method, which checks to see if the requested product ID is in the list of updated products. This list is updated by the JMS topic subscriber thread. Once the check is done, the product is removed from the updated list as the product is cached after the check.
- ***Steps 3a–3b:*** These are optional steps that are called if the rendered content for the requested product is not available in the cache or the product has been updated.
- ***Step 4:*** If the rendered output for the requested product is available in the cache and the needRefresh() method of the WebRefreshPolicy returns false, the content is retrieved from the cache and returned to the browser.
- ***Steps 5–8:*** This is the same as explained previously in the business tier section.
- ***Steps 9–10:*** When the product administrator updates the product, the application publishes a product updated message to the topic. Both the message bean in the business tier and the subscriber thread in the Web tier are active subscribers to the product update topic and receive the product updated messages and update their respective caches.

Common Challenges

Having a well-defined application framework is critical, as most or all requests coming into the Web and EJB container should be handled by controllers and facades respectively. This enables the framework to provide automatic caching functionality to business applications that use these frameworks.

Caching large objects requires a significant amount of memory. Deciding on the size of memory, the size of the disk allocated for caching, the number of object instances that can be cached, and the algorithm to use when removing entries from the cache due to maximum limits all have a tremendous effect on the cache performance and the overall application and have to be well thought out and defined based on the application.

The OSCACHE cache engine provides a robust mechanism for updating clustered caches. This is disparate from the WebSphere or any other appserver clustering mechanism. OSCACHE uses either JMS or JavaGroups to cluster cache components. In a clustered environment, each node maintains its own local copy of the cache; and when you flush an entry from the cache, a request to clear it is sent to all nodes in the network. This results in lower

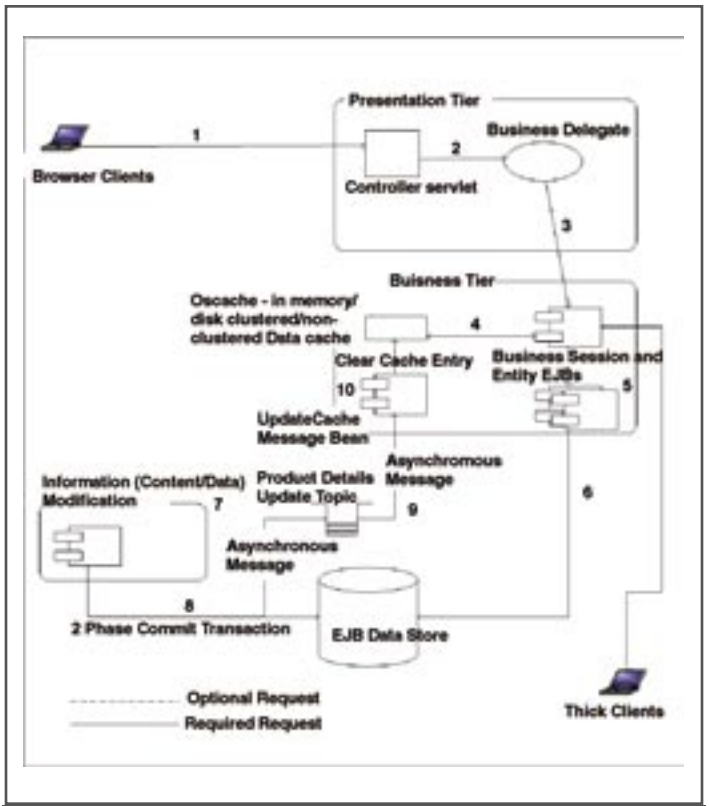


FIG 4: ASYNCHRONOUS CACHING – BUSINESS TEIR

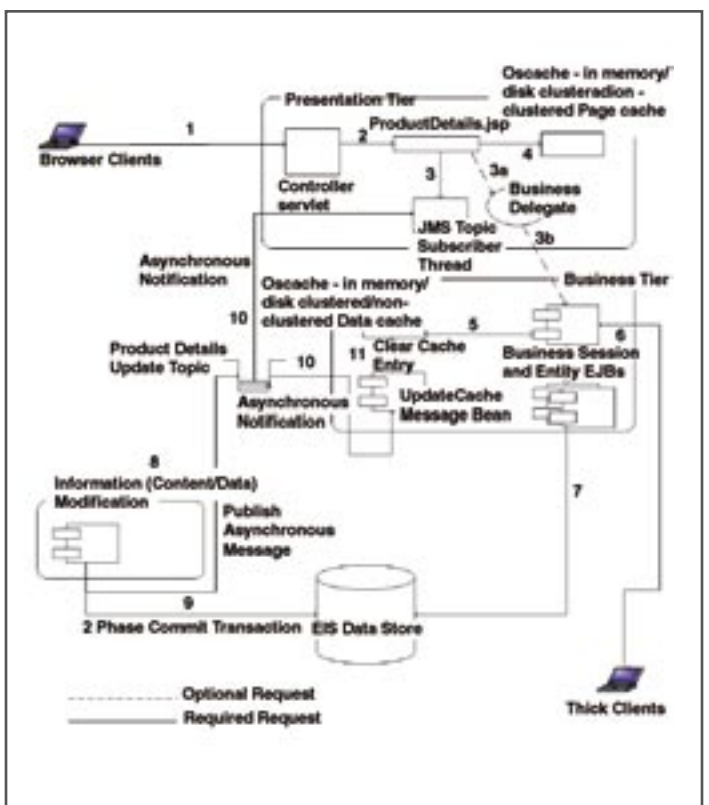


FIG 5: ASYCHRONOUS CASHIN – WEB TIER

network traffic as opposed to a single cache with each node querying the cache on every request.

OSCACHE throws an exception if an entry is not present in the cache. This might be a hassle for clients to deal with in extending the appropriate OSCACHE cache administrator and cache classes. It will also block threads depending on cache settings when the cache entry is being updated. Extra care should be taken to make sure this does not cause any problems.

Two-phase commits should be used to ensure that appropriate update messages are published when the EIS Tier is updated. Most EJB containers will handle two-phase commits, although it should be noted that two-phase commits have inherent overhead associated with them.

Choosing a Caching Solution

Here are some considerations to think of before choosing a caching product or solution:

- ***Does the caching solution impose requirements on the architecture?*** Caching solutions should not define any sort of special requirement or change the way a request is handled in the J2EE architecture. The solution should be introducible into an existing architecture without modifications to the Web architecture itself.
- ***Is it easy to configure and use?*** Does it provide a graphical interface to configure the cache?
- ***How many mechanisms are available to invalidate the cache?*** A caching solution should at the least include time, group, and rules as well as provide an extensible programming mechanism for invalidating the cache.
- ***Does the caching solution provide clustering support?*** Is the clustering solution effective? Are there multiple mechanisms for clustering caches? Is the clustering mechanism efficient? A caching solution should at the least provide one form of clustering mechanism and it should ideally result in minimal network traffic.
- ***Does the caching solution require additional software to provide the capabilities specified above?*** A caching solution should not impose any other software requirements.
- ***Does the caching solution have a monitoring and reporting mechanism?*** A caching solution should ideally have monitoring and reporting mechanisms that allow the cache configurations to be further refined and tuned.

Caching engines include:

- OpenSymphony OSCache
- WebSphere Dynamic Cache (built into WebSphere Application Server)
- WebSphere Edge Server
- JCache (JSR-107)

Conclusion

One of the most important responsibilities for application architects and designers is to provide robust, scalable, and highly available Web applications while imposing minimal hardware requirements. An available and powerful option that goes a long way in accomplishing these goals is the

–continued on page 49

Forget something?

Post-launch is NOT the time to be verifying web applications.

The wild blue yonder of operational monitoring and management is extremely unforgiving. Which means that going live with the monitoring software you used in development is a great way to go dead—quickly! You simply can't support operations if your staff is drowning in details provided by development profiling tools and debuggers. **Let NetIQ cover your apps...with AppManager.**

AppManager—the industry's easiest-to-use Systems Management suite—is a proven management system for monitoring J2EE application servers, databases, operating systems and even end-user response time. NetIQ's AppManager monitors ALL application components—not just your server. **NetIQ. Nobody does UNIX better. Nobody.**

Visit us at www.netiq.com/solutions/web to learn how we can help you address the challenges of your operational monitoring and management.

netiq Work Smarter.

Care for the Caregivers

Four solutions that cut costs without losing TLC

Today's hospitals, clinics, doctor's offices, insurers, and government agencies have more sophisticated technology and information to use in providing quality health care than would have seemed possible even a few years ago. However, tapping this gold mine of data and technical capability to deliver the most effective, efficient patient care remains a challenge for this unique and critical industry.

The two key players in the health care industry – the providers and the payers – share some issues. Both groups face tremendous pressure to improve patient care while mitigating the effects of skyrocketing costs. Both must meet increasing regulatory requirements related to privacy, security, and more.

Using IT to address these critical issues is challenging because key systems and information are often isolated. The result is a limited ability to share and synchronize data, particularly on the fly when critical patient care decisions are made. Administrative reporting is complicated and can be labor intensive, with much of it manual (particularly in provider settings) and thus slow, error prone, and duplicative. A lack of comprehensive, quality data adversely affects patients daily.

The efficient sharing of valuable data is difficult, whether between providers and payers or groups of clinicians within a single health care organization. For example, a hospital's diagnosis, patient information, lab testing, and pharmacy systems are typically not tightly linked. This lack of integration and reliance on paper, fax, mail, and EDI is very costly and allows errors to go undetected, often until it is too late.

To help address these financial, treatment, and regulatory challenges, health care entities need to integrate, within and beyond their own organizations. For some institutions this may be a matter of survival. They need to become on-demand enterprises – organizations in which processes are integrated across all groups and partner institutions, allowing them to respond with speed to patient's and professional's demands, research needs, and regulatory requirements.

IBM works closely with all types of health care organizations to address their critical challenges and help them become on-demand organizations. This experience with health care customers guided IBM's creation of middleware solutions designed specifically to address some of the industry's most pressing challenges. The solutions are part of IBM Software Group's effort to deliver industry middleware solutions – a strategy based on customer buying behavior that indicates they prefer to buy solutions designed for their industry. Each industry middleware solution contains software from IBM's WebSphere, Lotus, Tivoli, DB2, and Rational middleware brands, combined with industry-specific middleware, applications from independent software vendors (ISVs), and industry-expert services.

The health care solutions are for providers and payers. IBM's middleware solutions for providers are "designed to address cost and regulatory issues individually and to help build an information management structure," said Tony Bosselait, IBM's Worldwide Software Sales Business Unit executive – Public Sector, Health care. "A customer can start with one solution addressing a given challenge and build out to others as they need to in an open, integrated infrastructure."

For academic medical centers, public health care agencies, and hospital networks, the IBM solutions are coordinated within the Aligned Clinical Environment. It has been "architected so that you can easily and incrementally add functionality that responds to the issues the clients want to tackle, while retaining their existing infrastructure," said David Epstein, director, Solution

Development, Public Sector, IBM.

For academic medical centers, public health care agencies, and hospital networks the IBM solutions are coordinated within the Aligned Clinical Environment. It was "designed to add functionality but not add to the customer's infrastructure," said Epstein. "It requires no radical shift in technology and there is a growth path. Our solutions can be integrated into the existing infrastructure starting with any issue the customer wants to tackle."

"The environment guides how IBM designs, builds, and operates an integrated clinical and research information management infrastructure, allowing these clients to access and/or aggregate data that already exists to provide better patient care and create sustainable competitive differentiation," said Lynn Everitt, IBM Software Group Market Segment manager, Health care. "It addresses clinical and research integration, clinical performance improvement, disease monitoring and public health surveillance with four middleware solutions."

One of those solutions, the IBM Middleware Solution for Healthcare Collaborative Network (HCN), provides accelerated integration of disparate applications to improve the availability of integrated clinical information. This secure solution provides a user-friendly interface for specifying what kinds of clinical patient data should be monitored across an institution's various clinical systems. The results are improved quality and safety of care, improved staff productivity because they can focus on event exceptions to care protocols, application/information integration from a customer view, and increased clinician satisfaction and utilization of IT systems.

IBM's goal in developing the HCN was to provide a unified infrastructure for the transmission of clinical information from a local to a national scale, according to Epstein. To determine the appropriate architecture, the developers worked with three U.S. government agencies – the Food and Drug Administration, the Centers for Disease Control, and the Centers for Medicaid/Medicare Services – that needed similar data for different purposes.

"We also worked with hospitals to make sure we could automatically get this data," said Epstein. "Despite the challenge of their very heterogeneous environments, we were able to build a neutral integration point with the required security features."

New York Presbyterian Hospital used the solution to move from paper to a network for data exchange. The new system, which includes pre-existing applications and infrastructure, has reduced the complexity, time, and cost of data sharing, and is flexible enough to adapt quickly to changing requirements. As participants in the Healthcare Collaborative Network increase, the solution will allow hospitals, insurers and the U.S. government to respond rapidly and decisively to routine or unexpected events and identify patterns and trends in their rapidly changing industry.

"With all the sophisticated technology found in a modern hospital, the lack of coordination among hospital systems seems almost primitive," said Herbert Pardes, M.D., President and CEO, New York-Presbyterian. "A seamless, integrated network of information could do as much to protect patient safety and improve care as many other medical breakthroughs."

Many hospitals have disparate clinical and hospital information systems. This means they are often unable to provide clinicians with real time actionable data or correlate data from multiple applications. Furthermore, there is a tremendous amount of redundancy in patient and clinical information and managing paper-based processes is costly.

The IBM Middleware Solution for Healthcare Clinical Decision Intelligence can help improve quality of care and safety by decreasing the time to clinical diagnosis and treatment and reducing safety risks posed by a lack of information. By aggregating data from multiple, a disparate database, this solution improves decision-making by providing physicians to tools and the data necessary to practice evidence-based medicine.

The Clinical Decision Intelligence solution helps clinicians be more responsive and improve medical outcomes by providing real-time secure access to consolidated patient and research data. It also enables the data for analysis and profiling on such issues as management and cost containment. The solution also integrates people, processes, and information regardless of device used or location.

H. Lee Moffitt Cancer Center & Research Institute, a leading cancer center, plans to use the Collaborative Clinical Portal solution to unify its legacy systems, provide collaborative systems worldwide, and provide access to its enterprise data warehouse and key applications. The solution, consisting of IBM WebSphere Portal software and hardware, with Lotus Workplace Team Collaboration, has integrated business processes and information. It has helped improve patient care, helped physicians collaborate around the world, unified access control and unified the system look.

The IBM Middleware Solution for Healthcare Clinical Decision Intelligence can help improve quality of care and safety by decreasing the time to clinical diagnosis and treatment and reducing safety risks posed by a lack of information. By aggregating data from multiple, disparate applications, the solution improves workflows, straight-through processing, process monitoring and reduces processing costs.

The Clinical Decision Intelligence solution helps clinicians be more responsive and improve medical outcomes by providing real-time secure access to patient and research data. It also provides in-depth prospective and retrospective data for analysis and profiling on such issues as management and cost containment. The solution also integrates people, processes, and information regardless of device used or location.

The Mayo Clinic in Rochester, Minnesota, used the solution when teaming with IBM on its Computation Biology Collaboration Project. The Mayo Clinic had vast

stores of critical patient and research data, including detailed genetic, genomic, and proteomic data that was useful to its physicians and scientists on a daily basis. Unfortunately, the data existed in various places and in various forms throughout the clinic, making it difficult for employees to access. The Mayo Clinic wanted to consolidate the information into a centralized data warehouse.

The initial phase of the project was the creation of a data warehouse using a DB2 Universal Database and WebSphere technology running on IBM eServer pSeries 650 and 690 servers running AIX. The complete data warehouse is housed on an IBM TotalStorage Enterprise Storage Server. To enhance usage of the data warehouse, which stores information for more than 4.3 million patients, IBM developed a Data Query Abstraction tool.

The Mayo Clinic now has a foundation for using genomic data to further study illnesses and their treatments. Specifically, the solution provides the ability for nontechnical personnel to run complex queries, to express queries in end-user terms, and to define queries without regard for physical data representation. It also provides support for durable queries that can adapt to change in physical data schema and the clinic can manage query and result sharing.

The Mayo Clinic also used parts of the IBM Middleware Solution for Healthcare and Life Sciences Clinical Genomics, which makes it easier to access and analyze increasing amounts of data. The solution improves collaboration and provides an environment for tapping into and analyzing genomic data for use in research. It also improves application and information integration by linking multiple data sources into a single accessible data warehouse. For example, the solution can decrease the time it takes to develop new drugs by providing an environment for capturing clinical patient data for reuse as the basis for directed drug development research.

Overall, the Clinical Genomics solution helps accelerate targeted, innovative, clinical, and basic research by optimizing the capture and integration of phenotypic and genotypic data and enabling secure, cross-institutional sharing of data for collaborative research. The Clinical Genomics solution provides an environment for capturing and integrating clinical patient data with high throughput research data in order to identify and validate novel therapeutic targets, conduct more focused clinical research, and ultimately revolutionize the way diseases are diagnosed and treated.

The payers in the health care industry are also facing a critical need to improve service while keeping rapidly rising costs as low as possible. In the face of these budget pressures, they need to improve organizational performance with existing resources. In addition, there is a scarcity of skilled workers and manpower is costly, existing applications are difficult to learn and application consolidation requires major deployment and retraining costs.

The IBM Middleware Solution for Healthcare Payer Services Portal is designed to improve business and IT operations efficiency. It improves organizational performance by making it easier for customer services personnel to learn and use legacy and new applications through a consistent role-based interface. It helps reduce administration costs with the secure framework where employees can do their work from anywhere, at anytime, faster (providing single sign-on and secure application access.) It also provides collaborative and workflow capabilities to streamline and speed up business processes. The result is improved workflow, ease of doing business, integrated customer view and potential for increased market share.

WellChoice, the largest and oldest health insurance company in the State of New York, used the solution to consolidate disparate membership, billing and claims processing operations into a single enterprise system. Their goal was to improve customer relationships and provide a self-service model. Empire used IBM software and hardware to launch four portals for members, physicians, brokers and employers.

The result was reduced operational costs (from six legacy systems to two), centralized membership, billing and claims processing activities combined into a single set of business operations, and enhanced delivery of services to different customer groups.

IBM's Middleware Solution for Payer Plan Administration also addresses the need to cut the cost of administering customer interactions. Payers have difficulty cutting these costs because of a lack of usable data to manage them. Insurance risk and pricing can't be modeled because data is incomplete, not detailed, or not available until too late. In addition, customer service is not adequate because the representatives don't have the right information at the right time.

The Payer Plan Administration solution automates processes and creates an efficient business environment to reduce costs. It draws out transaction information to improve administrative and medical expense management. It also allows payers to respond to sponsor and member demands for excellent self-service and improved access to decision-making information.

It was used by BlueCross BlueShield of South Carolina to improve the efficiency of their claims processing and provide customers with self-service capabilities. IBM's solution allowed BlueCross to furnish its clients with data concisely via the Web and MQSeries enabled them to bridge the gap between different computing platforms to provide transparent customer facing applications. The result was the creation of a sophisticated data processing center that can adjudicate all claims in real-time.

"It's an example of the type of solution many health care organizations need, said Stephen K. Wiggins, CIO, BlueCross. It's been through this working partnership that our companies have developed a solution in health care that anticipates and exceeds marketplace demands."

Best practices for developing a robust architecture

J2EE Application Performance Analysis

BY LLOYD HAGEMO & RAVI KALIDINDI

Lloyd Hagemo is a senior director for the Application Infrastructure Management (AIM) group at Candle Corp. He is responsible for IBM WebSphere tools development. He has led the successful development of more than 20 products for the WebSphere environment that include operating system utilities, network performance and tuning products, WebSphere MQ configuration and management tools, and application integration solutions.

How well does your application perform? It is probably one of the toughest questions to answer accurately. It is not only a question of how many requests your application serves per second or per minute, but also how your application scales with respect to other performance metrics. It remains challenging to quantify application health quickly because there are a number of variables to consider. This article outlines the key terminologies associated with measuring Java 2 Enterprise Edition (J2EE) application performance. Importantly, this article also outlines best practices for creating a robust architecture and optimizing J2EE application performance.

When we discuss performance analysis terminology, it is not uncommon to find a variety of terms. The scope of application performance also varies depending on the application. Key performance terms follow.

RESPONSE TIME

Response time refers to the length of time required to complete a transaction. This is a primary factor for consideration from a user perspective. It is good practice to measure response time for each application component. This approach enables you to identify if a component's response time is longer than average and to look for coding or configuration problems that are causing the performance delays.

THROUGHPUT

Throughput measures application capacity, rather than user response time. It can be measured by the number of requests per second or number of pages per second that an application serves. Throughput is also measured by transactions or users per second.

SCALABILITY

Scalability denotes the ability to increase application capacity by expanding hardware and related resources that support the applications. There are two types of scalability: vertical and horizontal. Vertical scaling refers to the ability to install multiple instances on a single machine and thereby achieve maximum utilization of one machine. Horizontal scaling denotes the deployment of an application to additional computers.

These factors should be considered for current and future application uses to ensure successful deployments and expansions of new applications or services.

J2EE Application Analysis

J2EE applications should be examined from the macro and micro perspectives to gauge application performance accurately. The micro view analyzes metrics for system-level hardware and software components to provide a granular measurement of each component. The micro perspective focuses on the application modules, classes, and methods deployed on application servers. The micro view typically includes performance analysis of J2EE components, such as Enterprise JavaBeans (EJBs), servlets, portlets, etc. The micro view enables you to calibrate each application component to ensure a high level of application performance.

Macro analysis, by contrast, takes a collective, global view of all components associated with an application. The macro view of an enterprise includes the analysis of proxy, application, and other servers. The macro analysis reflects the end user's perspective of an application's performance. A combined micro and macro analysis provides the granular component-level and broader enterprisewide perspectives required to calibrate J2EE application performance accurately to maintain a positive user experience.

J2EE Workload Management

The J2EE environment provides Java application programmers with a set of services used to solve business problems. There are several methods, classes, and interfaces used to leverage these services.

A workload is the set of methods invoked to satisfy a user's request. An application method calling sequence, for example, can be defined as a workload. These methods include the J2EE service classes. All workloads have well-defined start and end points.

One popular workload monitoring technique is to use logging to indicate when methods are invoked. The application using this technique would wrap any Java or J2EE method call. This simple solution, however, will not provide all of the information required to find a workload bottleneck because the J2EE services that also call other services would not be tracked.

To optimize application performance, you should tune applications based on the response times of external J2EE call behaviors in your enterprise environment. External J2EE calls to track include Java Database Connectivity (JDBC), Java Message Service (JMS), and J2EE Connector Architecture (JCA).

A standard way to instrument application performance is to deploy an agent that instruments the application byte code when the Java Virtual Machine (JVM) loads the application. Because application and J2EE classes are instrumented at load time, you need to identify specific system and application classes to instrument. Most instrumentation products use an Extensible Markup Language (XML) configuration file to identify the classes. The key with this approach is to minimize the overhead and not change the operation of the application. Therefore, care must be taken to capture only the data needed to identify bottlenecks. Most J2EE application performance monitoring vendors provide this type of instrumentation.

Advanced J2EE application performance monitoring products typically include an editor that reads Enterprise Archive (EAR), Java Archive (JAR), and Web Archive (WAR) files to simplify this task. You identify what you want to instrument, and the editor creates the XML file for you. This approach is optimal because most application projects involve teams of Java developers. These teams build their own classes,

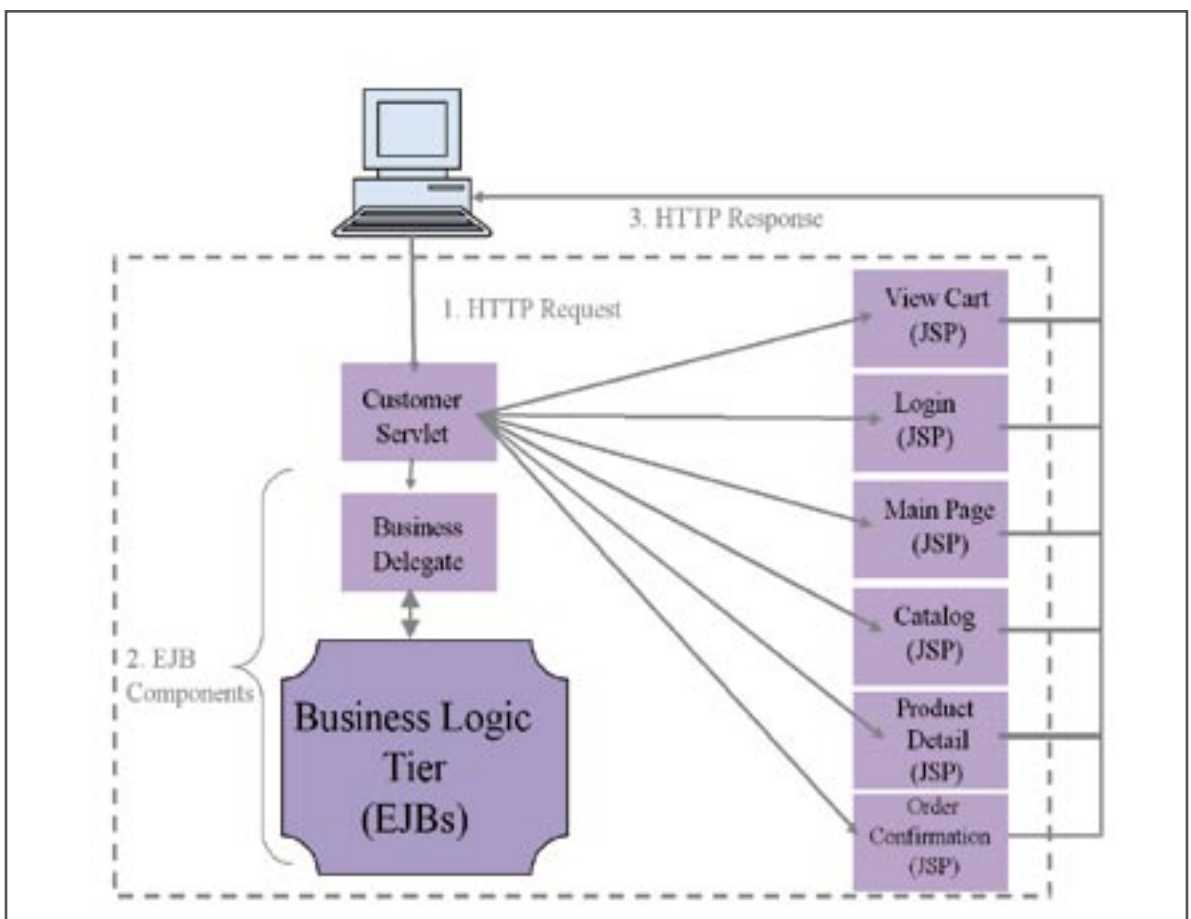


FIG 1: CUSTOMER WEB INTERFACE

Ravi Kalidindi is a senior software engineer for Candle's AIM group. He has worked closely with Java and J2EE since their inception, on a wide variety of J2EE projects. He has also published several articles that focus on Java and J2EE best practices.

which may not be visible to the person doing the system performance testing. Naturally, J2EE resource classes are instrumented automatically for reporting.

You must also be able to turn this feature on and off at will. If an application was deployed and tuned to eliminate performance problems six months ago, this level of application performance monitoring is not currently necessary. However, you must be able to reload instrumentation tools onto applications that require workload monitoring to pinpoint the cause of application performance problems in the future.

Workload analysis is a powerful technique that can be used to quickly identify J2EE application bottlenecks. Key response-time measurements you should track include servlets, JavaServer Pages (JSPs), portlets, EJBs, and user-defined workloads. Further, J2EE application server components to track include CICS Transaction Gateway (CTG), Java Transaction APIs (JTA), JCA, JMS, Java Naming Directory Interface (JNDI), and Open Database Connectivity (ODBC) services. Allowing an application programmer to drill deeper into specific workloads to identify specific delays will help accelerate the identification and elimination of performance bottlenecks.

Case Study

Organizations are continually working to streamline their procurement processes to increase the level of service to customers and partners. The example highlights an organization's Web site that provides customers with online access to an automotive parts catalog system. The organization used multiple design patterns to implement a consistent, well-understood architecture that streamlines operations and provides consistent J2EE application performance. The company deployed a Model View Controller (MVC) architecture to process order requests. Responses were processed using MVC JSPs that provide the appropriate data in Web browser responses (see Figure 1).

This system is supported by a strong EJB back-end system. That system managed the shopping cart, catalog, credit check, and pricing. Once customers have purchased their items, the system would send requests to an EJB workflow architecture component of the J2EE order fulfillment system.

Figure 2 illustrates the back-end system used to process customer orders. This represents a common application used today by Internet shoppers. Multiple design patterns were used to provide a consistent architecture. Design patterns include:

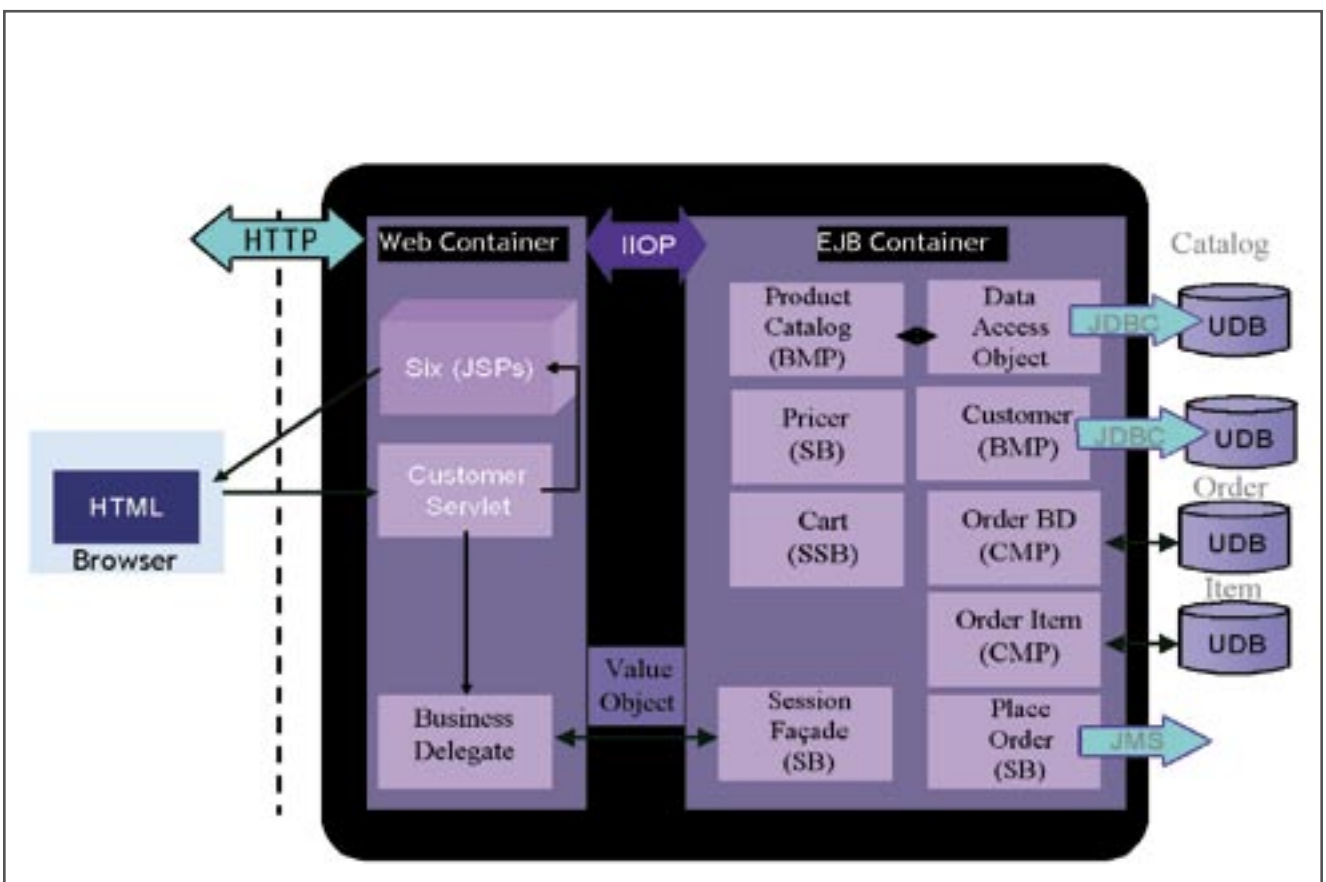


FIG 2: ORDER PROCESSING APPLICATION

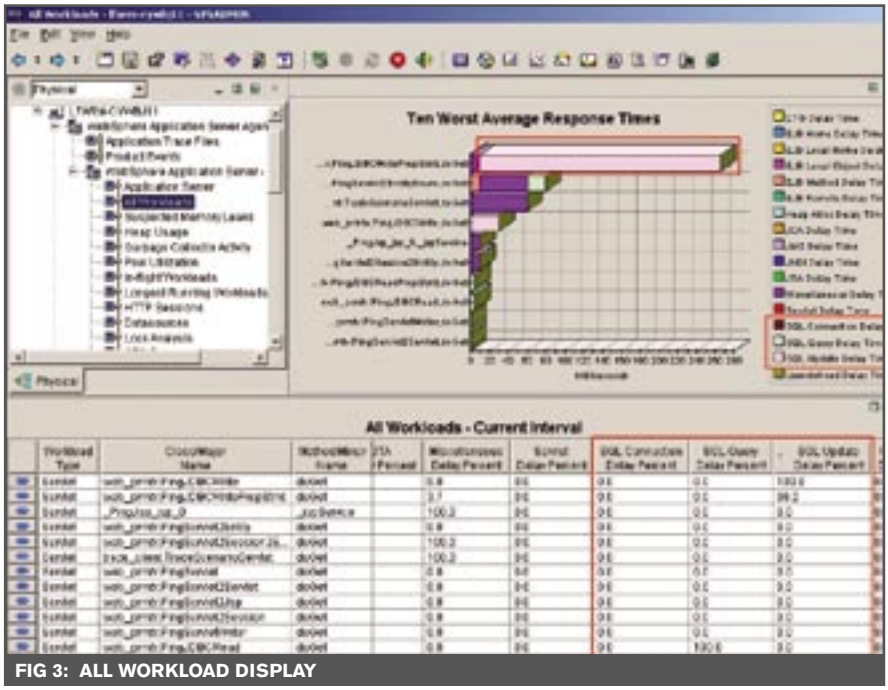


FIG 3: ALL WORKLOAD DISPLAY

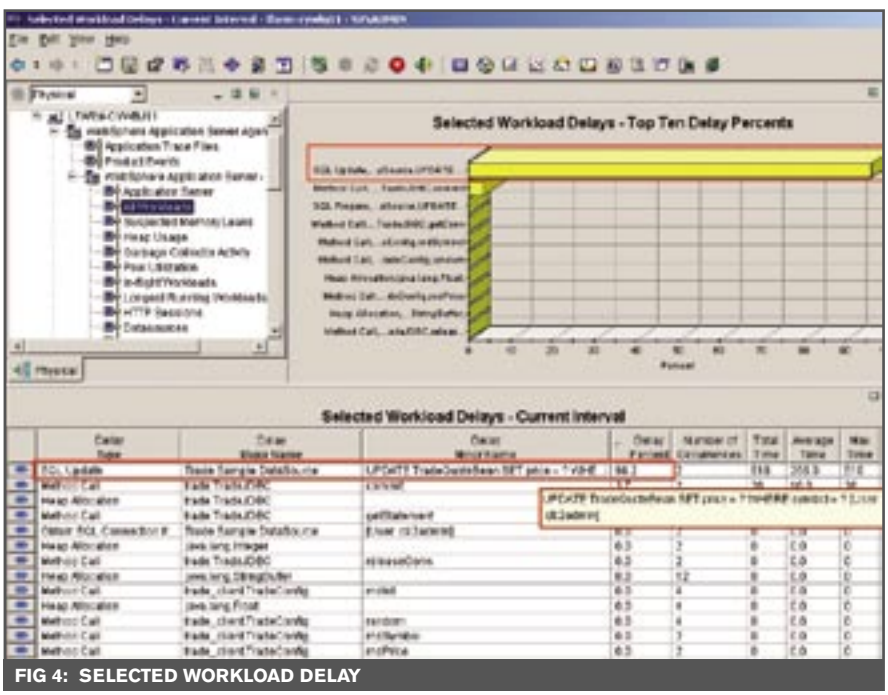


FIG 4: SELECTED WORKLOAD DELAY

- Session facade for routing EJB requests
- Value object to pass data between the Web and EJB container
- Data access object to manage the catalog

Workload analysis instruments the JDBC drivers that have been loaded as system classes. This instrumentation captures dynamic and static Standard Query Language (SQL) commands. Several types of information can be made available using byte-code instrumentation. For example, an application developer can identify the longest-running workloads (see Figure 3). An SQL update is responsible for most of the time spent in this workload. Through more detailed analysis of the workload, the application developer would be able to identify the specific method making this SQL call. Developers that focus on workload delays in system classes are well positioned to identify and eliminate complex performance delays.

The "Ten Worst Average Response Times" graph quickly identifies the longest-running workload. The color-coded bar segments the graph into distinct system and user delays. The graph at the top of the screen and the table below clearly define the cause of the performance problem. In this case, SQL created the performance bottleneck. Further analysis is required to identify the source of the SQL delay. By drilling into the specific workload, you can see the specific method and SQL statement that caused the performance slowdown.

In the drill-down display, you see the specific method within the database update EJB that caused the delay. It identifies the specific method name, along with the corresponding SQL statement. You can use this information to optimize the SQL statement or change the database table to increase performance. Once the change is made, you can reload the application to instrument the changed method and verify the desired results. The same technique can quickly identify any performance bottleneck within a J2EE application.

Conclusion

Today's J2EE application servers and corresponding testing tools provide robust functionality that simplifies application development and performance management. Design patterns used for architecture and detail implementation can minimize design and deployment missteps that can cause applications to perform poorly or not have the flexibility to address emerging business requirements. To tune J2EE applications properly, you should strongly consider tools that show specific workload delays and also enable you to drill down to the code level and eliminate the source of the bottleneck.

LinuxWorld CONFERENCE & EXPO

CONFERENCE: August 2 – 5, 2004 EXPO: August 3 – 5, 2004

MOSCONE CENTER • SAN FRANCISCO, CA

LinuxWorld Conference & Expo

is the #1 venue for decision-makers and influencers to discover real business solutions for real business problems, learn how Linux is accelerating as a total enterprise solution, and understand how it can be applied effectively to save their companies time and money.

Where OPEN MINDS Meet

CORNERSTONE SPONSOR

ORACLE

PLATINUM SPONSORS

Computer Associates · DELL · hp invent · IBM · intel · Novell · Sun microsystems

Keynote Speakers

Matthew Szulik,
Red Hat Chairman, Chief Executive Officer and President
Tuesday, August 3
9:15 am – 10:00 am

Nick Donofrio,
Senior Vice President, Technology and Manufacturing, IBM Corp.
Wednesday, August 4
11:00 am – 11:45 am

Martin Fink,
Vice President, Linux Business, Enterprise Storage and Services Business Unit, Hewlett-Packard Company
Tuesday, August 3
11:45 am – 12:30 pm

Alfred Chuang,
BEA Systems Founder, Chairman and CEO
Wednesday, August 4
2:30 pm – 3:15 pm

Michael Steven Rocha,
Executive Vice President, Global Support Services & Platform Technologies, Oracle Corporation
Tuesday, August 3
2:45 pm – 3:30 pm

Attend LinuxWorld and...

- Realize Linux as more than an operating system, but as **a world of applications**
- Evaluate the **latest applications and innovations** from leading open source companies
- Explore interoperability issues and opportunities in **open source and proprietary environments**
- Stay on the cusp of **emerging technologies** and the acceleration of open source adoption in enterprise computing
- Review the latest open source initiatives, their deployment and successes to help **make informed decisions** for your company

Register Online With Priority Code: D2801

WWW.LINUXWORLDEXPO.COM

IDG WORLD EXPO

At the Heart of an Enterprise

The benefits of hub-and-spoke and WebSphere Business Integration Server

BY DAVID LAWRENCE EPSTEIN & CAMERON MAJIDI

David Lawrence Epstein is a team lead and senior member of a team of WebSphere Business Integration experts at Prolifics. For over 15 years he has been designing and developing production software. Having authored numerous Java programming courseware and curricula, educational videos, and CDs, David has extensive experience in training and mentoring developers, architects, and administrators of large corporations. depstein@prolifics.com

There are a handful of enterprise application integration (EAI) products aimed at helping small, medium, and large businesses integrate enterprise assets to reduce costs and improve communications. This article is the first of a two-part series introducing IBM's EAI solution, WebSphere Business Integration Server, and its benefits. It is first essential to understand the advantages of a hub-and-spoke architecture, particularly in comparison to point-to-point integrations.

With years of expertise conducting EAI training sessions and implementing customer integration solutions, we have seen a gradual increase in the awareness of the financial and technical benefits of replacing point-to-point programs with a hub-and-spoke architecture. While many readers do not need to be convinced of the benefits of EAI, more often than not we find ourselves drawing a hub-and-spoke architecture on whiteboards for company decision makers. This tells us that the benefits of using a hub-and-spoke solution are not yet well understood throughout corporations worldwide.

Eventually, an understanding of hub-and-spoke will be widespread and its advantages over point-to-point interfaces will be obvious to all serious interface engineers. This article is an attempt to bring us closer to that day.

This first article focuses on the benefits of a hub-and-spoke architecture. The second article focuses on WebSphere Business Integration Server and discusses how the principles of a hub-and-spoke architecture are consistently reflected in several different aspects of the product. It is surprising how frequently hub-and-spoke is the right answer to common integration problems.

The overall goal of this series is to describe how WebSphere Business Integration Server is a simple, clean, modular, and easy-to-understand tool for obtaining a hub-and-spoke integration solution.

Hub-and-Spoke vs. Point-to-Point

In this section, we will detail some of the benefits of a hub-and-spoke architecture.

FEWER INTEGRATION PROGRAMS

A visual comparison of hub-and-spoke and point-to-point should make the most obvious benefit of hub-and-spoke apparent: there are fewer dependencies to maintain.

In Figure 1, there are five applications at CompanyX: appA, appB, appC, appD, appE. These applications could be specified as actual products such as SAP for manufacturing, Siebel for sales, PeopleSoft for HR, Clarify for support, and an internal application, developed in-house for CompanyX's niche market. For the purposes of this article, however, we will name the five applications appA, appB, appC, appD, and appE.

QUESTION: If every application needed to communicate with every other application using a specialized point-to-point program, how many total point-to-point programs are needed? For example, one program is needed to move data from appA to appB, and a second program is needed to move data from appA to appC.

ANSWER: All 5 need to talk to the other 4, so 5x4=20.

Figure 2 demonstrates this model. Note that only 10 lines are shown, but each line can actually be thought of as being two programs. The program that transfers data from appA into appB could be vastly different from the program that transfers data from appB into appA.

In general, if n is the number of applications, $n*(n-1)$ is the total number of programs required for a complete integration. Depending on the company and its objectives, it is not likely that every application must talk with every other application, yet some require multiple inputs. For example, the inventory received and inventory shipped input might come from the manufacturing system. For this reason, the upper bound of $n*(n-1)$ remains

a good estimator of the approximate number of interfaces for a well-integrated company that is leveraging its data.

A hub-and-spoke design for integrating these five applications will look much simpler. As diagrammed in Figure 3, each of the five applications, appA, appB, appC, appD, and appE, is a spoke and each connects with only one other application, the hub.

Instead of 20 programs to integrate all of the applications, there are 5. To be consistent, because each line was counted twice in the point-to-point diagram, the hub-and-spoke solution should be thought of as having 10 connections rather than 5. For the purpose of comparison, moving data from appA into the hub and moving data from the hub into appA should be thought of as two separate programs.

Perhaps comparing 20 programs with 10 programs does not look like a huge savings, only cutting the total number of integration programs in half? More to the point, this ratio might not look attractive enough for a company to replace a "working" integration solution with new software that implements a hub-and-spoke solution.

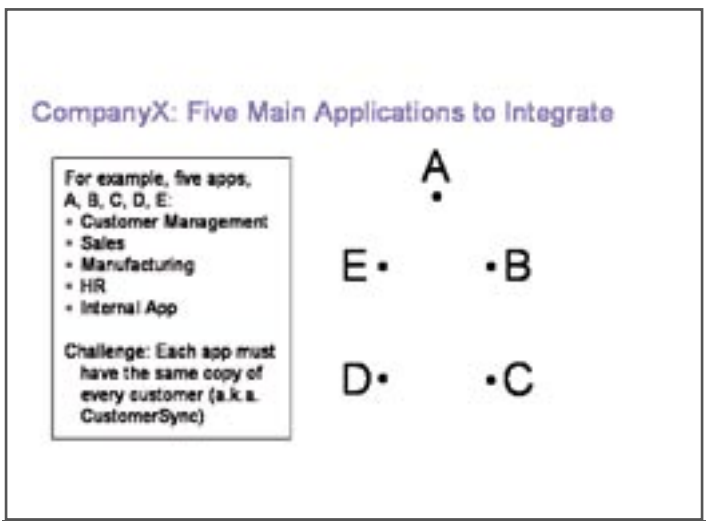


FIG 1: COMPANYX HAS FIVE DISPARATE APPLICATIONS THAT REQUIRE INTEGRATION

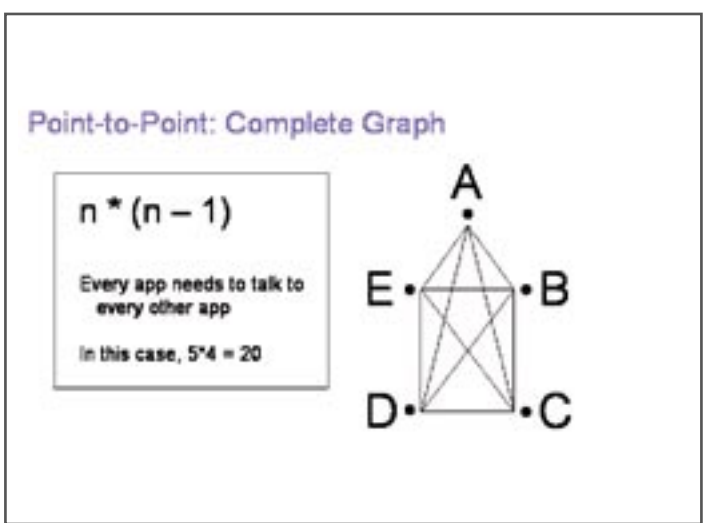


FIG 2: A POINT-TO-POINT INTEGRATION SOLUTION

However, as enterprises grow, it is important to consider the incremental costs of future enhancements to your integrated environment.

LOWER COSTS OF MAINTENANCE

The ratio of 2 to 1 changes quickly in favor of a hub-and-spoke solution when adding a sixth application. With the point-to-point solution, the total number of programs grows from 20 to 30 (6x5 = 30). For the hub-and-spoke solution, another connection into the hub counts as 2, making a total of 12 (10+2 = 12). Adding just one application, the ratio has gone from 20:10 to 30:12 (that is, from 2 : 1 to 2 _ : 1). Adding a seventh application should complete the justification from the mathematical viewpoint, as the point-to-point solution requires 3 times as many programs as the hub-and-spoke solution. Consider how often your organization adds new applications or replaces existing systems.

DECOUPLING OF APPLICATIONS

One general reason to adopt a hub-and-spoke design instead of numerous point-to-point interfaces is the decoupling of applications. Instead of applications relying on communication with the other applications, the only dependency is the connection with the hub. The freedom provided by such an anonymous system results in applications maintaining their independence.

Besides providing anonymity, the decoupling of applications also allows for asynchronous communication. From the point of view of an application, once the data is sent to the hub the effort is complete; it's the hub's job to make sure integration is successful. This enables each application to continue executing what it was deployed to do instead of spending cycles verifying the success of the data integration. A simple example of asynchronous communication is the process of sending an e-mail. The e-mail is sent without knowing or needing to know if the receivers are currently available. Without this feature, it would be practically impossible to use e-mail to correspond with a group on a project.

ACCOMPLISHING MORE THAN DATA INTEGRATION

In the hub-and-spoke design, all data flows through one place – the hub. It provides an opportunity to do more than just integration, for example automating a business process. Consider attempting to determine how many customers – who have paid for unlimited support – are contacting a company early Monday morning compared to late Friday afternoon. Depending on the applications, this could be quite a challenge. In fact, corporations are likely to limit their ideas to match the boundaries of their software design. With the hub-and-spoke solution, due to its centralized core of information gathering, this request is implemented with little effort, easily accomplished by

Cameron Majidi has been a veteran consultant to Prolifics for nearly 10 years. As a senior enterprise solutions architect, Cameron is part of Prolifics' specialized team of WebSphere consultants that IBM calls on to service its most challenging customer requirements. Cameron has designed and implemented WebSphere solutions for customers around the world.
cmaj@prolifics.com

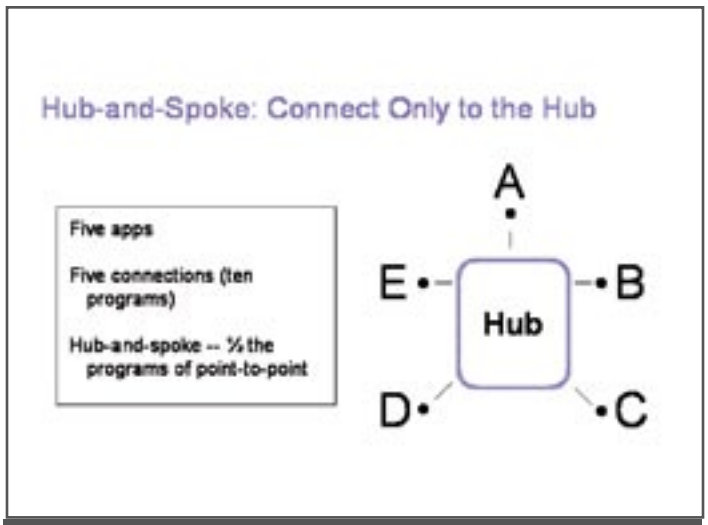


FIG 3: A HUB-AND-SPOKE SOLUTION

a nondeveloper in a single afternoon.

To assist in such tasks, the WebSphere Business Integration family of products includes tools to model, monitor, and manage these business processes. The tools enable you to design, simulate, and justify changes to your current business processes, as well as measure and improve the new processes once they are in a production environment.

EASE OF GENERAL ADMINISTRATION OF INTEGRATED APPLICATIONS

Depending on the environment, the costs of handling system failures could outweigh the costs of maintaining the integration software. It might be quite difficult to detect and recover from errors that occur while extracting or inserting data to and from software applications that run on various computers.

For example, consider the situation if the computer that hosts appC goes down at some point during the execution of the point-to-point integration programs. It could be before, after, or during the integration. This disastrous situation results in either corrupt data or in the difficult chore of attempting to figure out exactly which data from the other applications (appA, appB, appD, and appE) made it completely or partially into appC, as well as from appC into the other applications.

A hub-based environment centralizes all of the data and processes concentrating on the handling of the hardware and software failures.

FLEXIBILITY TO UPGRADE OR REPLACE APPLICATIONS

Considering the number of interdependencies that exist for any one of the five applications in our example, simply upgrading to a new version of the application could require modifications to four different point-to-point interfaces. With the hub-and-spoke design, the task of updating or replacing any application involves only the interface between it and the hub. As was detailed earlier, even more evident is the ease and flexibility gained when adding and integrating a new application into the environment.

Having explained some of the benefits of hub-and-spoke over point-to-point, you might wonder if using point-to-point interfaces is actually a working solution. Obviously, point-to-point works, and has for dozens of years at many companies. The difference between today and previous decades (when many enterprise applications such as PeopleSoft were introduced), is the reality of the coexistence of numerous enterprise applications along with the "latest-and-greatest" applications such as Internet-savvy, custom e-Business software.

One scenario in which it is beneficial to use a point-to-point program is if there are only two applications. In that case, point-to-point requires fewer integration steps than hub-and-spoke.

Point-to-Point: A Closer Look

Figure 2 is actually a simplistic view of point-to-point programs – a single line is drawn from appA to appB. The program to move data from appA to appB is in itself quite complex. As shown in Figure 4, from the highest level the program to integrate data from appA to appB involves three major pieces: extracting data, transporting data, and inserting data.

This is once again quite oversimplified, as the task of extracting and inserting data could be challenging. Issues such as security and navigating firewalls could make the transporting task far from trivial. Finally, overlooked in the diagram is the task of converting the data from appA into the format expected by appB.

To explain the task of converting data, consider simply one common piece of data, a customer's name. One application might store the first name and last name in two separate fields, while another application might store the full name in one field, using a space to separate the first name from the last name. Yet another application might store full name in one field as last name, space, first name. This provides us with enough background for a discussion of converting data, but it is worth pointing

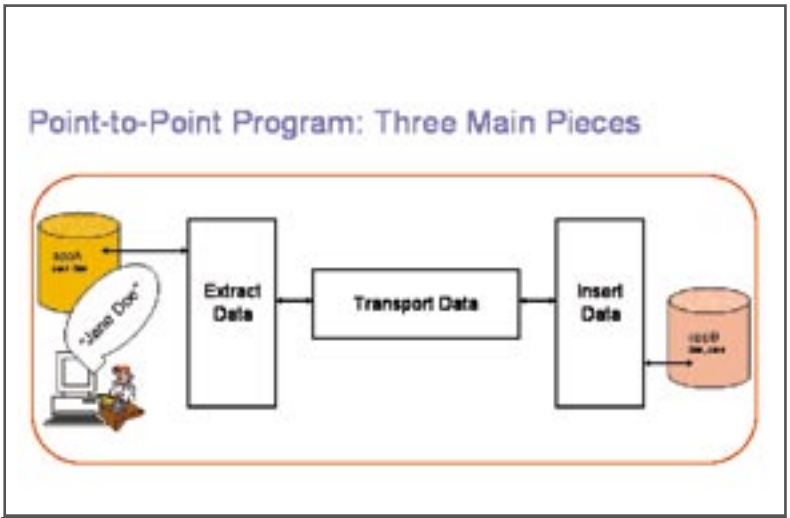


FIG 4: A POINT-TO-POINT SOLUTION

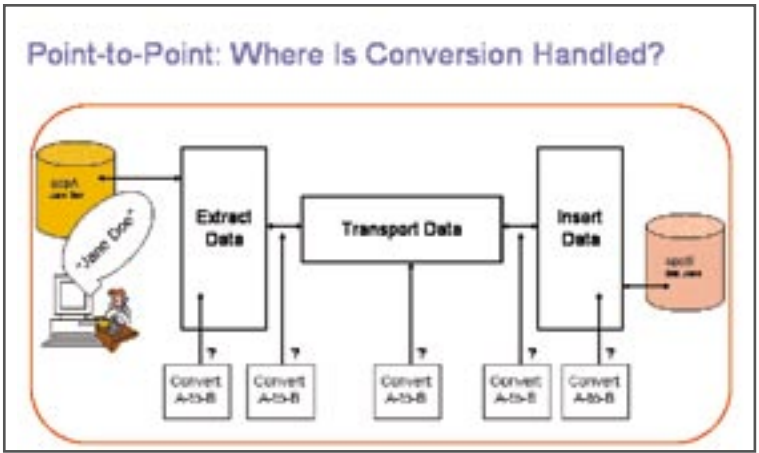


FIG 5: WHERE IS THE DATA CONVERSION CONDUCTED?

out that "Dr. Jane J. Doe III" and similar names provide an opportunity for a far more involved discussion than would fit in this article.

Data Conversion

Returning now to the task of converting data from appA's format to appB's format, let's assume appA stores first names and last names in two separate fields and appB stores the full names in one field as last name, first name. As questioned in Figure 5, between which of the three major steps does it make the most sense to do this conversion?

One might think the answer is obvious, but major EAI vendors do not agree on exactly where and when to do the conversion(s). For this reason, they have different architectures and different abilities.

Remembering that beyond appA and appB are appC, appD, and appE, the complexities of using point-to-point to integrate applications again become completely evident. An appA-to-appB conversion program is needed as well as numerous similar-yet-different conversion programs from appA into appC, appD, and appE.

In a hub-and-spoke solution, the first realization is that conversion will never be directly from one application to another: appA data is not converted to appB data. Instead of converting data directly from appA to appB, appA's data is converted into the hub's format. The next step is to convert from the hub's format to appB's format. It might appear wasteful to write two programs instead of one to convert data from appA to appB. However, it is important to consider the reduced effort when appC, appD, and appE are integrated into the picture.

Data Representation in the Hub

The final decision to make is to figure out what format to use for the hub. Returning to the discussion of representing first and last names as separate fields, or as one field using either a space or a comma and a space to separate first names from last names, Table 1 lists example formats for appA, appB, appC, appD, and appE.

Given this setup, which would be the ideal representation for storing the customer name in the hub? Because there is a conversion of data to the hub format, the hub format must have a definition, so the question is which format should be chosen for this example?

When implementing a hub-and-spoke solution, it is essential to focus on the overall design of the hub. Often, too much attention is placed on the integration of the data, reducing the role and effectiveness the hub can offer in the hub-and-spoke design.

The hub-and-spoke design is valuable not only to reduce the total number of applications that must be maintained, but also to provide a single, centralized place to which all data flows. The benefits of this could be enormous. Companies become better equipped to gain access to corporate data to analyze and improve their current business processes. Organizations can also add business processes between the applications, not just pass data, such that the applications work together as if they were indeed one unified application. For example, sales from early Monday morning can be compared with sales from late Friday afternoon. Even better, if sales on Fridays are above projections, the process can take action immediately and allocate appropriate resources (for example, e-mailing an executive or manager, or automatically sending purchase orders to appropriate trading partners).

Keeping an eye on the bigger picture reveals an opportunity to greatly improve and evolve yesterday's business processes. Returning now to the question of which data representation to use in the hub, appA, B, C, D, or E – but keeping an eye on the opportunity to use the hub as the owner of business processes – it becomes clear who drives the definition of the hub's data. The designers of the business processes use the hub's data definition, keeping the business process application independent.

In this example, the business processes are owned by CompanyX, so the representation of a customer's name should be determined by CompanyX's business processes, completely independent of the appA, B, C, D, E specifics.

Data Formats

APPA	APPB	APPC	APPD	APPE
first_name: Jane last_name: Doe	fullName: Doe, Jane	Name: Jane Doe	first_name: Jane middle_initial: null last_name: Doe	name1: Jane name2: Doe name3: null name4: null name5: null

TABLE 1: EXAMPLE DATA FORMATS

"Instead of applications relying on communication with the other applications, the only dependency is the connection with the hub"

It would seem that we are asking a lot from CompanyX. Not only will they phase out numerous point-to-point programs and replace them with a new hub-and-spoke system, they must also define the company's representation of a customer as well as specify their business processes that use this new data representation. This is where the WebSphere Business Integration family of products provides an exceptional offering. Together with the software and tools required to obtain a hub-and-spoke design, it provides out-of-the-box business processes and the corresponding data representations.

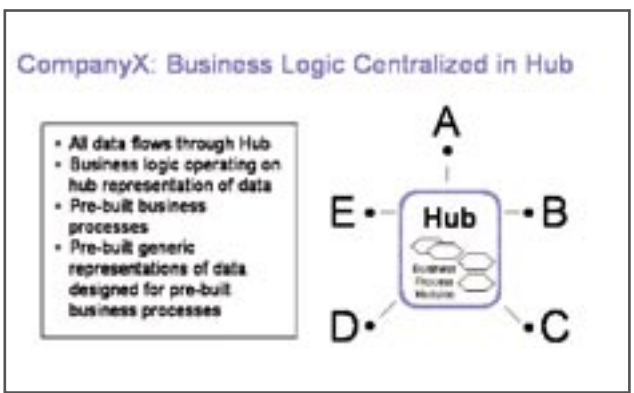


FIG 6: HUB-AND-SPOKE DIAGRAM WITH BUSINESS PROCESSES IN THE CENTER

WebSphere Business Integration Family

The WebSphere Business Integration family of products enables you to optimize your current business processes and integrate your disparate applications.

MODEL, MONITOR, AND MANAGE

The WebSphere Business Integration Modeler and Monitor provide you with the tools required to design and simulate business functions and processes, and then measure and improve your performance dynamically against your business objectives.

The modeler provides a graphical interface to detail current processes and associated costs, and design new processes, simulating various optimization choices in order to project which provide the best business benefits. A great benefit of modeler is its automatic flow from the business analysts to the integration programmers, generating code that is imported into IBM's workflow engine.

The monitor provides customizable dashboards and reports that reflect how the modeled business processes are working in the production environment. These processes can then be changed dynamically to improve effectiveness and provide for real-time business resource reallocation.

Integrate and Connect

WebSphere Business Integration Connect is an ideal solution for B2B requirements, enabling you to connect with your business partners to reduce costs and improve productivity. Internal business processes can be reused with your trading partners and customers.

At the heart of the enterprise and at the heart of this article is WebSphere Business Integration Server. The design of WebSphere Business Integration Server revolves around prebuilt out-of-the-box business processes, referred to as "collaborations." These collaborations are based on years of experience helping customers solve real problems and noticing commonality and patterns among various solutions. As shown in Figure 6, the WebSphere Business Integration Server hub becomes a container for a company's business logic.

IBM also recently released WebSphere Business Integration Server Express, which is designed to integrate and automate business processes for small and medium businesses.

Conclusion

Replacing expensive point-to-point programs with a hub-and-spoke architecture enables a company to compete using the power of today's software technology. A hub-and-spoke architecture provides much more than reduced long-term software maintenance expenses; it is an opportunity to benefit from an overall application design that introduces a centerpiece for managing all business processes.

The foundation has been set for a better understanding of the benefits of implementing a hub-and-spoke solution using WebSphere Business Integration Server. The prebuilt collaborations, designed to automate common business processes, are at the root of the overall design. The second part of this series of articles explains collaborations in greater detail and introduces the other major components of WebSphere Business Integration Server.

For more information about WebSphere Business Integration solutions, attend our Prolifics WebSphere Business Integration Server Webinar and Workshop series at www.prolifics.com/new/events.html.

Why, when, and how

Propagating Security Context Across a Distributed Web Services Environment

SCOTT MORRISON

It's a problem as old as networked computing. Consider two applications. They negotiate a level of trust. How can that trust – or security context – be transferred to a third application, one that may exist in an entirely different security domain from the first?

This problem has been solved before, but is limited by proprietary solutions that resist integration. The challenge now, which is a significant one, is to solve it again, but this time for Web services – a task complicated by the need to accommodate a broad range of established security procedures and legacy technologies.

Context in Context

Security context is an ambiguous term. Take, for example, the SSL protocol. Here, security context is largely cryptographic metadata – the master key, derived session keys, ciphers and hashes, etc. – which are associated by a public SSL session ID. The session ID exists precisely to allow reuse of these across independent connections and thus avoid the expensive public key-based handshakes that would be necessary to re-establish them. Authentication might not even be involved; such is the case with the Diffie-Hellman cipher suite.

In this article, we will explore the more fundamental problem of transferring the security context established by an act of authentication – that is, a sufficiently substantiated claim of principle identity – between applications in a Web services environment. In doing so, we will use two important OASIS Web services standards, WS-Security (WSS) and the Security Assertions Markup Language (SAML).

WSS and the Security Token Mechanism

Back when I was still in high school, my parents gathered up the family and spent a summer traveling in China. During a few days in Beijing, I had a chop, or signature seal, carved with my name rendered phonetically in Chinese. In China, chops have been used as a means of signature and identity since the period of the Warring States, nearly 2500 years ago.

The chop implements a security model called proof of possession. It is something physical you have, something you need to protect, and something you can use to create a security token that binds another object – a contract, a painting, etc. – to yourself. The binding consists of a stamp, most commonly rendered in red ink, of the name carved into the chop. The artistry of the carving establishes uniqueness and is a simple guard against forgery.

Shortly after we returned home, thieves broke into our house. Along with the usual targets for theft – items like TVs and stereos – they took, oddly enough, my chop. I've always thought that this was a strange thing to steal: were they drawn to it because it was shiny and elegant; or was it an early example of identity theft? Perhaps there are checking accounts open in my name somewhere in Fujian.

The real problem with my chop is that it really wasn't bound to my identity. It was fine for creating security tokens while I possessed it, but once lost, the thieves could create unlimited identity tokens with no real means for me to stop them.

Security tokens, of course, come in many different forms and with varied purposes. A token could transport credentials; it might describe an authorization decision; it may encapsulate a key for a cryptography session. This diversity is one of the challenges faced by the technical committee developing WS Security. To this end, their specification does not attempt to mandate one form of security token over another; instead, it defines a simple encapsulation mechanism that should be able to accommodate most existing methods and technologies. Thus, in WSS, applications can make claims to identity, supported by tokens. The details of how to support a particular token mechanism is defined outside the main specification, in a separate document called a token profile.

Security tokens appear as elements subordinate to the <Security> header, the block in a SOAP message under which all WS-Security related

parameters appear. Most of the currently existing profiles are concerned with establishing a security context around identity. Consider, for example, username and password, probably the most familiar of all security token schemes. WSS defines a profile called – not surprisingly – the Web Services Security Username Token Profile. It defines a very simple and logical organization for usernames, passwords, and nonces; the latter enabling digest authentication schemes to provide credential validation without direct password transmission. WSS calls this type of token a proof-of-possession claim. Listing 1, taken directly from the specification, is an example. (the code is online at www.sys-con.com/websphere/sourcec.cfm)

But what about transferring an existing context? You could argue against the need for this – after all, if you have a means of authentication, why not simply re-authenticate continuously with every independent transaction? HTTP basic authentication works in this way. When a browser successfully meets an authentication challenge, it will proactively insert credentials into the HTTP Authorize header for every subsequent request in the same realm.

For some Web services applications, this is sufficient. For others, it can be tremendously expensive – the overhead of continuous credential validation can bring a directory to it knees. Furthermore, it may be unrealistic to believe that every server is capable of performing this operation. Often, this is because of access restrictions placed on central directories, perhaps due to topology, but often due to politics.

Transferring a previously established identity context, then, is a valuable thing. But it's also difficult to carry out securely. WSS provides a means to do this within its abstract token profile mechanism. Under this use case, the tokens don't establish initial identity, but describe an existing security context. These tokens have to be authoritative, so that if a token is stolen – like my chop – it can't be used to hijack or destroy an existing application or cryptography session. This challenge is addressed by the WSS SAML profile.

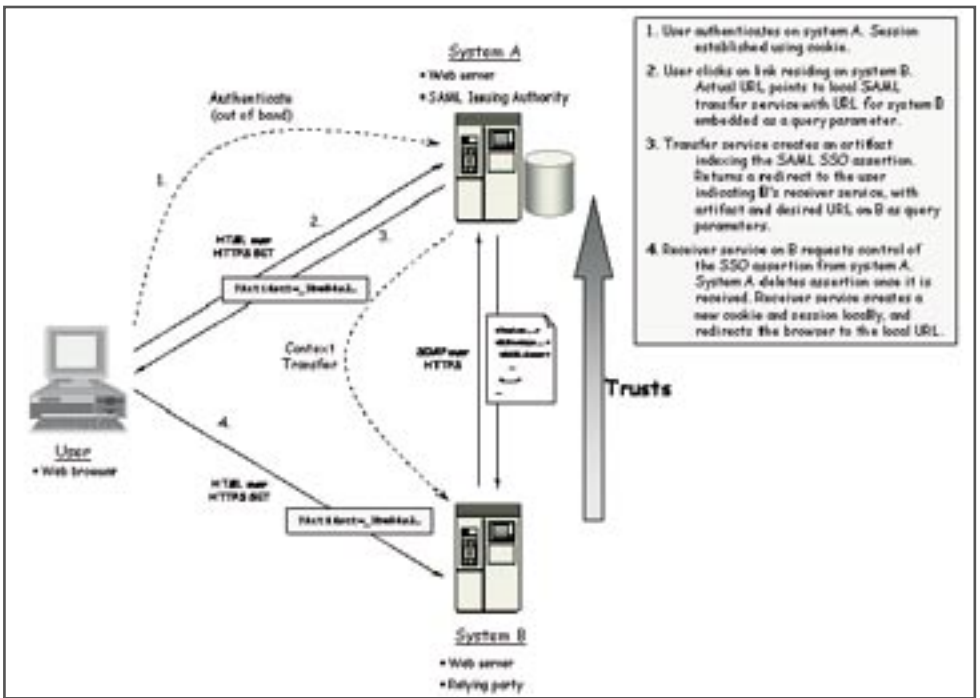


FIG 1: SAML BROWSER/ARTIFACT PULL PROFILE

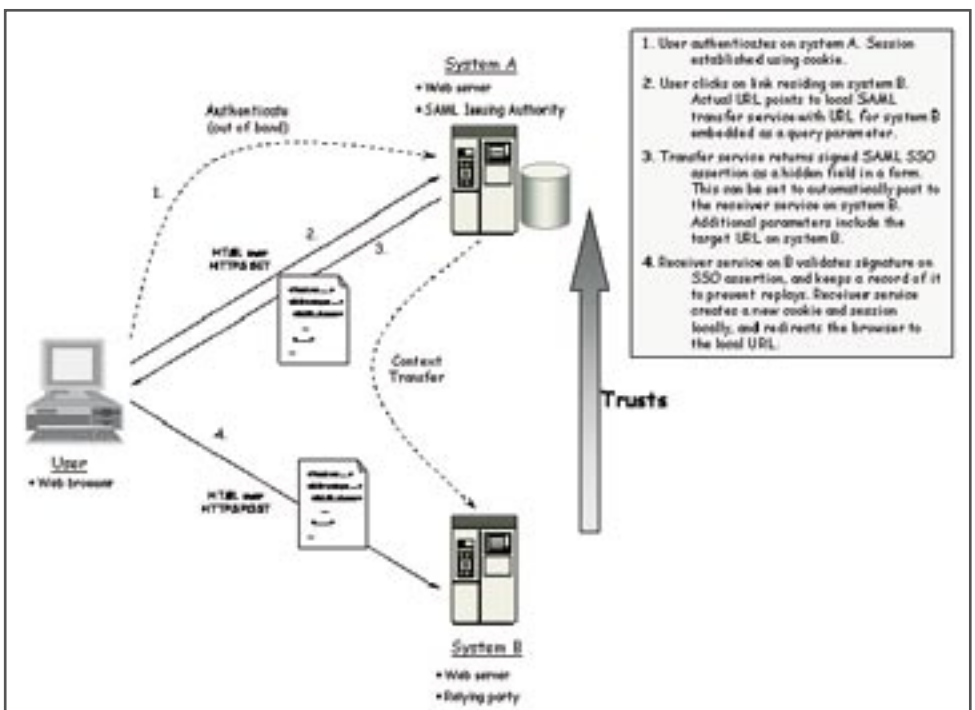


FIG 2: SAML BROWSER/ARTIFACT PUSH PROFILE

SAML and Context Transfer

SAML is designed to pass security information between systems. The basis of SAML is a markup language for declaring assertions. Assertions

are declarations of facts about a subject. You can think of a subject as the binding of an entity, such as a person or a computer, to an identity in a security domain. Assertions are generated by an issuing authority, which may front an existing identity server such as an LDAP directory. In SAML, there are three distinct kinds of assertions:

- ***Authentication assertions:*** These statements describe acts of authentication that have already taken place. An authentication assertion does not describe another method to perform authentication, such as using an X509 certificate; it simply affirms that a subject S was authenticated by means M at time T. In listing 2, the authentication assertion declares that subject smorrison authenticated against the Layer 7 Technologies corporate directory using a password.
- ***Authorization assertions:*** An SAML issuing authority can make an authorization decision to allow or deny access for a subject to a particular resource.
- ***Attribute assertions:*** These assert that a subject is associated with a collection of attributes, represented as simple name/value pairs. For example, an SAML authority might declare that subject Scott is associated with group=developers and company=Layer 7 Technologies.

By providing a generalized attribute mechanism, SAML makes an important point: that security context is more than just authentication and authorization, but also includes associated metadata that might be important in a security decision, such as a subject holding gold status in a frequent flyer system.

In addition to assertions, SAML defines a request/response protocol for obtaining assertions from SAML authorities, bindings to protocols such as SOAP for transporting assertions and queries, and profiles, which take a more holistic approach to integrating SAML within an existing framework, such as SOAP messaging or conventional, HTML-oriented HTTP.

While the vision behind SAML has been to produce a general-purpose language for communicating security context between distributed systems, its initial focus, growing out of a widespread and immediate need, has been on browser-based communications – in particular, single sign on (SSO) for the Web. SAML defines two additional profiles to address this, and in these, we can find a model for how SAML will ultimately support Web services (see Figures 1 and 2).

Both scenarios are functionally similar. The user, authenticated on system A, clicks on a URL addressing content that resides on system B. The user should not have to re-authenticate on system B (thus establishing a separate, independent security context), but instead should transfer the existing context completely to B. To complicate matters further, B may reside in a different security domain from A, so B literally may not be able to validate the subject's credentials even if they are made available. Therefore, a trust relationship must be established between A and B, so that B relies on A's word that a subject has been necessarily and sufficiently identified. Virtually every large organization attempting to integrate their internal Web servers has encountered this problem.

The difference between these profiles appears in implementation. Figure 1 depicts a pull scenario, in which a security token, called an SAML artifact, is passed to system B as a query parameter affixed to the URL. System B uses the artifact as a handle to take complete ownership of the security context from A; this is illustrated in the figure as a SOAP call from B to A, requesting control of the context and taking delivery as a collection of SAML assertions. SAML ensures that the server-side half of the context can only exist in a single place at any given instant.

In contrast is the push scenario, which transports the context entirely within a message – in this instance, the assertions reside as a hidden field that's POSTed in a form. This eliminates the need for system B to retrieve the context from A, but requires that the assertions be signed to prevent tampering. This is actually closer to a typical Web services scenario, where context is a security token rendered into a SOAP message, but more on that later.

In practice, this process usually involves a centralized issuing authority and clever use of HTTP redirects. But what is noteworthy here is the security model. These browser profiles rely on SSL and HTTP authentication mechanisms as a means to protect the confidentiality, integrity, and trust of assertions (or artifacts). It uses existing Web security to ensure that assertions are relayed only through the subject they describe. This eliminates the threat of replay attacks and session hijacking. It's a crucial point: an assertion, even signed by an issuing authority, needs to be bound to the subject presenting it. Otherwise, what's to stop an intruder from simply copying a signed authentication assertion and using this to stake claim to that assertion's correlated security context? Unbound from identity, an assertion is like my stolen chop.

In the browser profiles, secure, authenticated channels are necessary to ensure that security tokens only pass between trusted entities. In Web services, where security is implemented on a message-by-message basis and no secure channel exists, there needs to be a different approach.

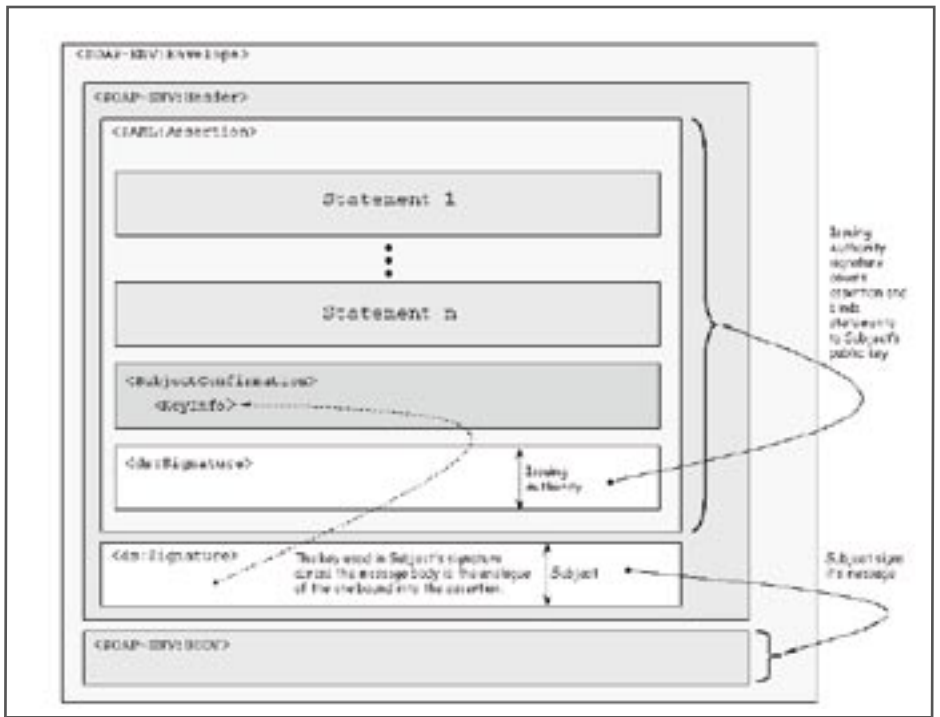


FIG 3: A SOAP MESSAGE

WSS SAML Profile

SAML, of course, fits cleanly into the WSS Security token structure. The real challenge, though, is more subtle than syntactic contracts. WSS is about providing security on a message-by-message basis. Furthermore, it is concerned with absorbing security into the message itself and decoupling it from channel strategies like SSL to be able to provide continuity in encryption, integrity, authentication, and reliability across a diverse set of transports and intermediates – from SMTP to MOM to plain text files, in as many hops as the application demands.

The challenge, therefore, is binding a subject's identity to an assertion so that it is verifiable by the ultimate receiver of a SOAP message. SAML addresses this with an assertion element we have not encountered yet, <SubjectConfirmation>.

An SAML issuing authority uses SubjectConfirmation to bind a particular subject's iden-

tity to an assertion. There are various strategies for this, such as including Kerberos service tickets, and these are left for specification in the relevant profile. The WSS SAML Token Profile adopts an interesting approach. Within this element, the issuing authority can insert the subject's public key. Remember, the issuing authority is making a definitive statement about an act of authentication that has already taken place, so it's likely to hold the subject's public key. If the subject authenticated using its certificate under the WSS x509 Token Profile, the key is there. Alternatively, it should be able to retrieve the key from a trusted certificate server after firmly establishing the subject's identity under a different authentication scheme, such as username/password. The key resides within the SubjectConfirmation element, inside a standard <ds:KeyInfo> block, a rich structure already described in the W3C XML Digital Signature specification.

The issuing authority then signs the entire assertion, thus authoritatively binding assertion and key. In this way, it's not unlike a certificate, which uses the signature of a trusted party to bind a public key to an identity (represented as a DN). Consider also, what often makes a certificate useful is the additional information residing within it. An SSL certificate binds a DNS name to the CN, thus allowing clients to verify that the TCP socket they've opened is indeed connected to the Internet entity described in the certificate. E-mail addresses were added to support similar trust validation, and v3.0 extension fields in the x509 specification promote still richer models of trust.

Listing 3 shows an authentication assertion, generated and signed by an SAML issuing authority. This signature binds the subject of the assertion to its DSA public key. It also includes its X509 certificate against which a receiver can compare a pre-existing trust relationship.

So why is this useful? Precisely because now the subject can create an undeniable association between any SOAP message it authors and this assertion. Our problem up until now has been that, even though it is signed by the issuing authority, a plain assertion could be intercepted and used by anyone. In this way, it's like the thieves who stole my chop, and could forge unlimited new messages claiming to come from me. Suppose we are a SOAP receiver – say some arbitrary service downstream – and we take delivery of a SOAP message containing an assertion claiming

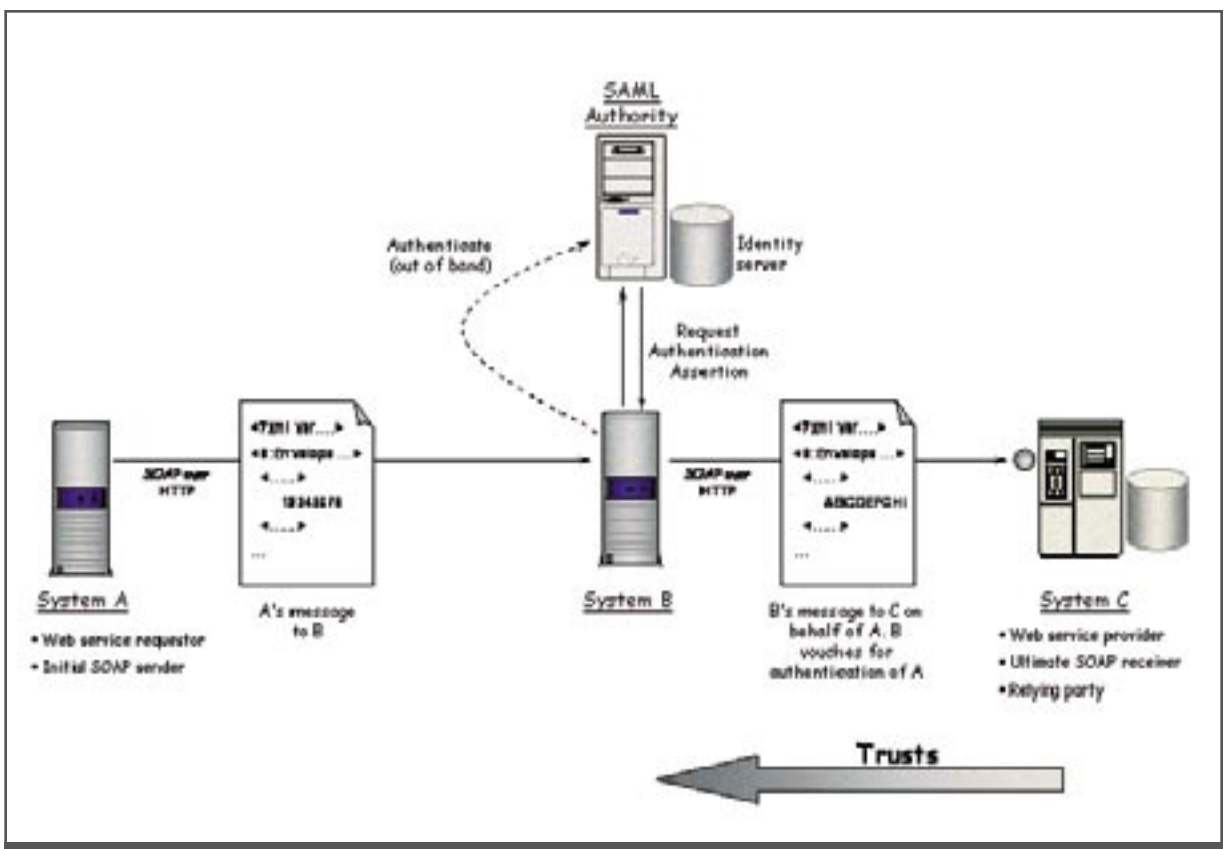


FIG 4: A TYPICAL WEB SERVICES INTERACTION IN THE SENDER VOUCES SCENARIO

that Scott was authenticated at noon on Tuesday. How can we be confident that the sender of the SOAP message really is subject Scott, described in the assertion? SubjectConfirmation is the key – or perhaps better, holds the key. WS-Security calls this an endorsed claim, as it's been sanctioned by a trusted third party.

Figure 3 is a block diagram of a SOAP message that shows how it all comes together. It maps a typical message that a Web services consumer would compose while participating in an SSO scenario. Fundamentally, this is the same as the browser SSO model illustrated in Figure 2, with SOAP service invocations substituted for HTTP/HTML. In this use case, System A might be a centralized authentication service that consumes username and password credentials (under the WSS token profile described previously), and returns a signed SAML SSO assertion (an aggregate of an authentication assertion, time of validity, and other optional attribute fields). Bound in this assertion is the public key of the authenticated subject.

To compose the message in Figure 3, the consumer copies the signed assertion into the SOAP message unchanged. To prove rightful ownership of this assertion, the subject signs the message body. Remember, only the actual can do this, as only the subject possesses the private key paired with the public key in the assertion. This establishes an irrefutable connection between the author of the SOAP message and the assertion describing an authentication event.

It is the receiver's responsibility to process this message appropriately and take action on it based on its predetermined trust relationship with the SAML issuing authority. Under SAML, the ultimate receiver of the message is called the relying party – a logical piece of nomenclature, as the receiver relies on the trust it has with an issuing authority.

Listing 4 shows what the SOAP message looks like, as it might be delivered to the receiver. It's becoming complex, because we now have signatures from two different parties: the SAML issuing authority (over the assertion only); and the SOAP sender (over the message body). The sender, could, of course, extend its signature across the entire envelope if that is the level of integrity that the application required. In the subject's signature block, the SecurityTokenReference element contains a reference back to the assertion, where we can retrieve the public key for signature validation. Ultimately, what we have created in a chain to a trust root, not unlike a certificate chain. It might help to refer back to the block diagram in Figure 3 to help navigate through this complexity.

There's one important detail we have yet to cover. Find the <SubjectConfirmation> element in the authentication assertion in Listing 4. It has a subordinate element called <ConfirmationMethod>. In the example, ConfirmationMethod takes the value of the SAML-defined identifier urn:oasis:names:tc:SAML:1.0:cm:holder-of-key. This informs a receiver that, when processing any SOAP message containing this assertion, the attesting entity must prove their association with the assertion using a signature.

An alternative processing model is called sender-vouches, indicated by the constant urn:oasis:names:tc:SAML:1.0:cm:sender-vouches. This addresses an issue found in another common Web services scenario. In sender-vouches, the attesting entity is not the subject described in the SAML assertion. However, it is acting on behalf of that subject. The receiver, therefore, must trust that this is indeed the case, that the sender has validated the true subject in some way and is working on its behalf. To make this work, the attesting entity must protect both the relevant parts of the SOAP message and the assertion itself to prove that it has made that association (after all, with nothing to conjoin these data, the aggregation could have come from anyone).

Figure 4 depicts a typical scenario where this might take place. It's very similar to the classic three tier browser-based application – just substitute SOAP for HTTP/HTML, RMI or IIOP, and JDBC. System A is a Web services client. System B consumes and validates its credentials against an issuing authority. System C trusts that B validated A accurately, and processes messages from B with confidence that they are a consequence of an initial request of A.

Conclusion

Inside the WSS SAML token profile, we find the basic mechanism necessary to transfer one type of security context between applications using Web services. But don't lose sight of its limitations in scope and maturity. SAML – and by extension, WSS – does not deal with larger issues like cryptography or application sessions, global sign-out, or account linking. Some of these are more appropriately addressed in federation specifications like WS-Federation and Liberty. Some are addressed in other emerging standards efforts like WS-Secure Conversation and WS-Trust. Others will see light in SAML v2.0. Nevertheless, there is some very valuable work here by people who deeply understand the issues in distributed computing security, and elements of the specifications are relevant today. Which is good, because we've needed this for a long time.

How Long Can BEA Survive, Industry Asks

The May issue of *Java Developer's Journal* (Vol. 9, issue 5), scooped the annual Gartner/Dataquest Report six weeks before it was published and asked the question, "How Long Can BEA Survive Against IBM?" The day after the report came out, BEA announced its quarterly earnings and its stock had dropped 23%.

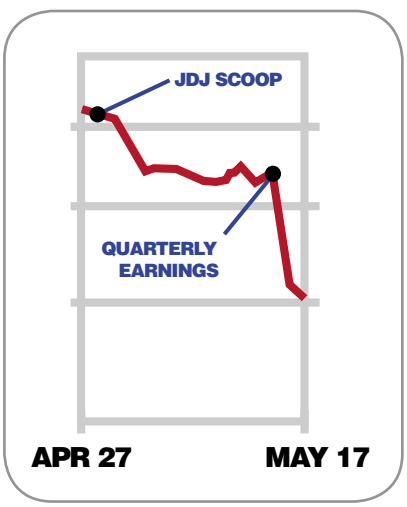
As predicted in the May issue of *JDJ*, the Gartner/Dataquest report showed that the proprietary application server market is dominated more than ever before by IBM, with BEA's share slipping for the second successive year.

IBM's share in 2003, according to the annual study, hit 41.3% ($429.7 million) "up from 36.4% in 2002," while BEA's slumped to 27.5% ($286.2 million), down from 29% in 2002.

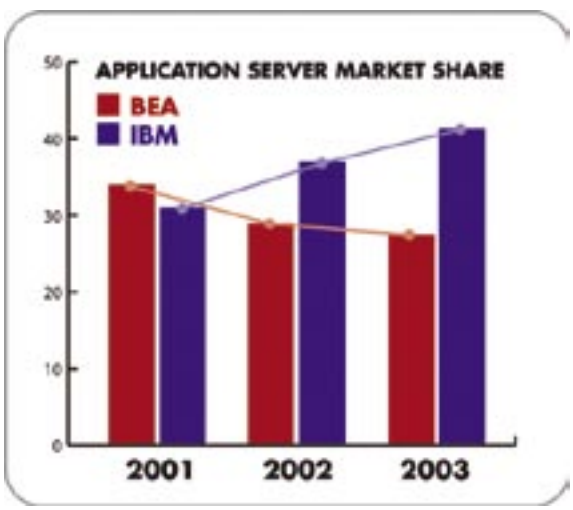
The figures are a vindication of Armonk, NY-based IBM's approach to the Java app server market, especially bearing in mind that the total market fell in 2003 by 8.8% compared to 2002

It was in 2002 that BEA first slipped from the top spot, dropping from 2001's 34% to 29%. In a race that began in 2000, it took Big Blue just two years to catch up. Oracle came in third to BEA in 2003, Gartner's study reveals, with a 19% market share.

Joanne Correia, vice president of the Gartner/Dataquest Software Team and the analyst who has produced this particular study each year, confirmed, "IBM is gaining share in every market, whereas most vendors were flat or negative."



BEA SYSTEMS STOCK MARKET ACTIVITY
APRIL 27–MAY 17, 2004



SOURCE: GARTNER DATAQUEST

Just two days after its stock dropped 23% – its biggest drop in more than five years – BEA came out fighting on May 19th, with the news (revealed by *WebSphere Journal*'s sister publication *WLDJ* before the official announcement) that it was donating – to what CTO Scott Dietzen referred to as "Open Source Land" – the first open source application framework targeted at Java-based Web applications: "Project Beehive."

"Project Beehive" is the name BEA has given to its release of the runtime application development framework from its BEA WebLogic Workshop tool and includes the controls in Workshop.

Announcing the release to the world, Dietzen emphasized that this was no sudden shift in BEA strategy just to resist stock market pressures. "We had planned long-term to announce Beehive before eWorld," he said. "This is a foundational piece [of BEA's future strategy] and is the motivation for this announcement now, so that the Java community has a chance to assimilate and see how it fits into everything before our user conference next week in San Francisco."

The response to the news was also strong, as opinions changed from firing the CEO Alfred Chuang to suggestions that HP, Microsoft, or Oracle may bid to buy the company.

Rumors and suggestions appeared on the message boards that *JDJ*'s publisher had warned Chuang back in his March editorial, "Success, Arrogance, Rise and Fall," that the direction in which he was taking his company did not look good.

Following is just some of the discussion on the BEA news, which includes the Gartner/Dataquest report, BEA's latest quarterly earning results, and Beehive.

Feedback

WEBLOGIC TRANSITION WILL CONTINUE IN Q2

Ouch...conf call mentioned that the WebLogic transition will continue in Q2

Translation: license revenue could be lower again.

– Laid Off Last Summer

DOUBLE SPEAK BY ALFRED

Double speak, poor sales execution – excuses, excuses – Alfred blamed merger and acquisition for poor results. Meanwhile, Oracle and IBM are gaining license revenue – what the hell?

– redhotchillpepperlover

BEA SUPPORTERS ARE MISSING THE BIG PICTURE

BEA supporters are missing the big picture. BEA got here basically selling app server to techies, first to market with the best product in a boom economy.

To get to the next step, they need to sell their platform product against the big boys with more mature products and much bigger install bases to farm.

The biggest hurdle is that they will use price as their sword. They can discount and give their software away because alternate lucrative sources of revenue exist for them. BEA cannot and that will be their Achilles heel. BEA does not have alternate revenue streams to compete to what amounts to "free." If they try the net, net is zero or negative license revenue growth.

This is why this stock is getting bashed, not for their results today, but for expected results during the next quarter and beyond.

– mpraps

BEA SHOULD JOIN ECLIPSE

We do not understand why BEA is not joining the Eclipse Board? Eclipse is becoming the leading IDE. WebSphere Application Server is getting more and more market shares because of its integration inside WSAD and Eclipse. BEA and WebSphere are both great application servers!!!

– Vlad VARNICA

EXPECT ALFRED TO GET FIRED IN THE NEXT 6 MONTHS OR SO

Folks, if you can't beat a dinosaur like IBM in the app server space, you have to ask yourself, how low have you sunk? Since when did IBM ever have a lead in any software space (apart from mainframe)?

Alfred the magician's answer was to hire the crappy dinosaur execs from the laggards (IBM, HP) and bring them to BEA so that BEA could finally end up like a small IBM – formal, but really useless.

Expect Alfred to get fired in the next 6 months or so and for BEA to have a sliver of a final chance of recovering.

Expect to see this stock hover around the 6–7 range in the next few weeks. Could even go back to test support at 5, if the markets tank. Get out now longs. Shorts will make a killing on this in the next few weeks. IMO.

– stockmeister88

WHAT ABOUT JBOSS?

Aside from BEA and IBM, the report attempts to imply 31% in the "others" category. Independent industry analysis shows that JBOSS owns 24% to 27% of that. How is that not important and why was it not mentioned? Is the goal of the article to inform or to concoct some artificial metric called "commercial" and report on "commercial" application servers? What they mean by "commercial" is not that the products are used for industrial commercial purposes but rather that you paid a large sum for the product. Highly misleading and a real disservice to readers, in my opinion.

– Robert Kashmir

ALFRED'S MOTIVATIONAL SPEECH

I will never forget his motivational speech to the sales force a few years ago. He said something like, "All of you must kick some ass and have fun." He must have been referring to management kicking the sales force's ass...and working there was definitely no fun! They turn over the sales force every year. Not a good use of money in my opinion.

– foo_fighter007

I'M NOT SURE WHO OR WHAT IS BRINGING BEA DOWN

I am not sure who or what is bringing BEA down, but it seems unlikely to be journalism: 4 out of the 5 research firms listed here by Yahoo Finance recommended a downgrade on May 14 after their quarterly earnings report.

The graph for the last 3 months tells its own story.

– barthrh213

THEY WILL NEVER SEE IBM IN THEIR REAR VIEW MIRROR AGAIN

Can BEA survive against IBM? Give me a break – they will never see IBM in their rear view mirror again. BUT, now they see Oracle in their rear view (despite their lame assertions that they don't). The real question is "can they survive against Oracle" – which is undoubtedly the database underneath the large majority of their installs. Microsoft is still operating on a largely obstacle-free non-Java parallel path, so no stopping them. But Oracle...that spells big trouble for the BEA gang. The fight with Big Blue is over, and if they aren't careful, the fight with Big Red could be their undoing.

– John

ROFLMAO!!!!

The fact is that Alfred has been a better CEO than 99% of all CEO's in existence.

ROFLMAO!!!!

In fact, let's see how many of you can name another company that went from 0 to a billion dollars in such a short period of time....

The fact is that Bill Coleman was in charge when BEA was formed, and Bill Coleman was still in charge when BEA became a billion dollar company!

Alfred took over, decided to put ALL BEA's eggs in the Java/J2EE basket, and that strategy has failed!!!

WebLogic Platform 8 (Alfred's baby) is basically dead on arrival – its sales have actually DECLINED the last three quarters. That is a failure my friend, nothing else.

I've worked under both Coleman and Alfred – Alfred doesn't come close to Coleman. He's an engineer and has NO business running a company. He should be replaced, and I predict he will be replaced shortly!!

Oh, and by the way....

TOLD YA!!!!

–kikuyo4me

BEA WILL NOT BE ABLE TO SURVIVE AT ALL IN THE LONG TERM

I am of the opinion that BEA will not be able to survive at all in the long term. My prime reason for this opinion is that I believe money cannot be made by just selling middleware. In most large accounts, IBM/Oracle/Sun are giving away middleware (app/web/portal/dir servers) to get more hardware and service revenue. It will be hard for BEA to compete with, for example, Sun's $100/employee middleware pricing model.

–AM

THE PART OF THE TAG LINE IS VERY REVEALING

Not good at all IMHO!

The part of the tag line that is very revealing is this: "Protection from Vendor Lock-in."

That is a very real problem for BEA sales right now. WebLogic 8.1 platform, as it stands now, is a very strong "lock-in" for BEA. There's all this "stuff" in the form of callable APIs, but if you develop to those APIs, you have to pay very big bucks to BEA and ONLY BEA to deploy whatever system you develop. You are "locked-in" to BEA for deployment licenses and maintenance fees.

I guarantee you that the tag line was created in direct response to field sales feedback about why they aren't able to move the 8.1 platform in the quantities BEA had hoped.

But how will Beehive help BEA?

It smacks of a desperation move, a quick fix to stem the tide of the open source movement. It will also have a chilling effect on current 8.1 Platform deals, as customers hold off to see if they can get a cheaper or free open source implementation of what is currently available in Platform 8.1.

Sorry longs, I don't think this bodes well for the long term, watch for pops and exit as painlessly as you can!

IMHO of course!

–kikuyo4me

WebSphere® JOURNAL

ADVISORY BOARD
Richard Arone, David Caddis, Mike Curwen, Devi Gupta, Lloyd Hagemo, Barbara Johnson, Shannon Lynd, James Martin, Doug Stephen, Victor Valle

EDITOR-IN-CHIEF JACK MARTIN
EXECUTIVE EDITOR GAIL SCHULTZ
CONTRIBUTING EDITOR PATTI MARTIN
PRODUCT REVIEW EDITOR JAY JOHNSON
SPECIAL PROJECTS EDITOR RICHARD ARONE
EDITOR NANCY VALENTINE
ASSOCIATE EDITORS TORREY GAVER, JAMIE MATUSOW, JENNIFER VAN WINCKEL

WRITERS IN THIS ISSUE
David Lawrence Epstein, Lloyd Hagemo, Ravi Kalidindi, Hari Kanangi, Cameron Majidi, Jack Martin, Scott Morrison, Cameron Purdy

SUBSCRIPTIONS
For subscriptions and requests for bulk orders, please send your letters to Subscription Department.
SUBSCRIBE@SYS-CON.COM
Cover Price: $8.99/Issue Domestic: $149/YR (12 Issues)
Canada/Mexico: $169/YR Overseas: $179/YR
(U.S. Banks or Money Orders)
Back issues: $15 U.S. $20 All others

PRESIDENT AND CEO FUAT KIRCAALI
PRESIDENT, SYS-CON EVENTS GRISHA DAVIDA
GROUP PUBLISHER JEREMY GEELAN
SENIOR VP, SALES & MARKETING CARMEN GONZALEZ
VP, INFORMATION SYSTEMS ROBERT DIAMOND
VP, SALES & MARKETING MILES SILVERMAN
PRODUCTION CONSULTANT JIM MORGAN
FINANCIAL ANALYST JOAN LAROSE
ACCOUNTS PAYABLE BETTY WHITE
ADVERTISING DIRECTOR ROBYN FORMA
ADVERTISING MANAGER MEGAN MUSSA
ASSOCIATE SALES MANAGERS KRISTIN KUHNLE, BETH JONES, DOROTHY GIL
LEAD DESIGNER LOUIS F. CUFFARI
ART DIRECTOR ALEX BOTERO
ASSOCIATE ART DIRECTORS RICHARD SILVERBERG, TAMI BEATTY, ANDREA BODEN
WEB DESIGNERS STEPHEN KILMURRAY, MATTHEW POLLOTTA
CIRCULATION SERVICE COORDINATORS EDNA EARLE RUSSELL, LINDA LIPTON

EDITORIAL OFFICES
SYS-CON Publications, Inc.
135 Chestnut Ridge Road, Montvale, NJ 07645
Telephone: 201 802-3000 Fax: 201 782-9637
SUBSCRIBE@SYS-CON.COM
WebSphere® Journal (ISSN# 1535-6914) is published monthly (12 times a year).
Postmaster send address changes to:
WebSphere Journal, SYS-CON Publications, Inc.
135 Chestnut Ridge Road, Montvale, NJ 07645

WORLDWIDE NEWSSTAND DISTRIBUTION
CURTIS CIRCULATION COMPANY, NEW MILFORD, NJ

FOR LIST RENTAL INFORMATION:
KEVIN COLLOPY: 845 731-2684, KEVIN.COLLOPY@EDITHROMAN.COM
FRANK CIPOLLA: 845 731-3832, FRANK.CIPOLLA@EPOSTDIRECT.COM

J2EE Caching

–continued from page 22

caching of data and/or content at various tiers of the J2EE architecture. To manage caches in real time, asynchronous cache invalidation maintains the health of the cache. It accurately eliminates stale data without redundantly invalidating caches based on fixed-time intervals or periodically checking the state of the cache and invalidating stale entries. Providing application libraries in the initial phases of the design and development process takes advantage of asynchronous caching mechanisms and lay the foundation for well-designed, robust applications.

The Management Option

–continued from page 50

Ask yourself if you truly need session failover capability, because it is the most expensive option in terms of computing resources, and will require either the memory-to-memory or the session persistence option. If you need session persistence, ask yourself if you can tolerate the potential loss of data on failover that the TIME_BASED_WRITES option allows to occur. If you are fortunate, you don't need any of those options, and can simply use nonmobile nondistributed nonreplicated, nonpersisted sessions.

Last, there are a few downsides to watch out for, particularly with respect to the size of the HTTP session object. Even with new options such as memory-to-memory replication, WebSphere staff are careful to point out the importance of keeping your sessions small (from http://tinyurl.com/24cve).

So, make sure you follow the engineering rules for supporting session, keep your sessions small, and know for certain what availability options you actually require.

WebSphere Journal Advertiser Index...

WebSphere Journal Coming Next Month...

WHAT'S HAPPENING IN WEBSPHERE, PART 2
A continuation of the interview with Tom Inman, as he discusses what IBM has in store for WebSphere and what can be expected in the coming year.

WEBSPHERE MIGRATION
Calming the configuration beast

UNDERSTANDING JAVASERVER FACES
Assemble your Web applications effectively

UPDATING KEY BUSINESS PRACTICES AND IT SYSTEMS
Software solutions extend the functionality of the insurance industry

SURFACING A SIMPLE DOMINO APPLICATION IN WEBSPHERE PORTAL
Solutions By Design II

Subscribe Today!

– INCLUDES –
FREE DIGITAL EDITION!
(WITH PAID SUBSCRIPTION)
GET YOUR ACCESS CODE INSTANTLY!

*The major infosecurity issues of the day... **identity theft, cyber-terrorism, encryption, perimeter defense, and more** come to the forefront in* **ISSJ** *the storage and security magazine targeted at IT professionals, managers, and decision makers*

Editorial Mission

Greater Collaboration and Efficiency Through Education

- ✓ ***ISSJ*'s editorial mission is to showcase proven solutions** that will guide, motivate, and inspire senior IT and business management leaders in the planning, development, deployment, and management of successful enterprise-wide security and storage solutions.
- ✓ ***ISSJ*** brings together all key elements of data storage and protection, and presents compelling insight into the benefits, efficiencies, and effectiveness gained by focusing on these two critical areas of IT simultaneously.
- ✓ ***ISSJ*** is an objective, critical source of information that helps storage and security managers make informed management decisions about what is working today, and what they need to plan for tomorrow, and is the only publication that focuses exclusively on the needs of IT professionals who are driving the enterprise storage architecture/infrastructure while pursuing and incorporating the latest security technologies.
- ✓ ***ISSJ*** achieves our mission by delivering in-depth features, practical "how-to" solutions, authoritative commentary, hard-hitting product reviews, comprehensive real-world case studies, and successful business models, written by and for professional IT storage and security practitioners.

SAVE 50% OFF!

(REGULAR NEWSSTAND PRICE)

Only $39.99 ONE YEAR 12 ISSUES

www.ISSJournal.com

or

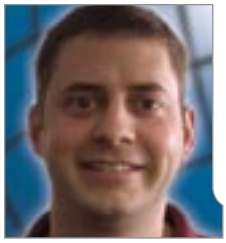
1-888-303-5282

SYS-CON MEDIA

The World's Leading i-Technology Publisher

The Management Option

BY CAMERON PURDY



WebSphere provides a number of out-of-the-box session management options, including a new in-memory replication option in WebSphere 5. Successful use of session management requires some engineering foresight, and optimal use of session management requires an understanding of the options that WebSphere provides.

On the engineering side, the rules for successful session management are simple:

- Make sure that all attributes that are put into a session are serializable. This allows you to later use distributed sessions if necessary for scalability or failover reasons.
- If you get a mutable attribute out of the session and change it, make sure that you explicitly put it back (using the setAttribute method) so the container knows for certain that the session attribute was modified. This is unnecessary for nondistributed sessions, or when using advanced session products, but you should still always follow this rule.
- Use the session attributes only for user- or session-related data, and never for data that can be stored or cached globally. Sessions used for global data end up increasing significantly in size, and the data is often duplicated across many sessions, thus wasting memory.
- Keep track of how big your sessions are, and try to minimize that size. Some of the session management options work acceptably only for small sessions.
- You will most probably have to do some extra work to support URL encoding for session identity. URL encoding is often used as a backup approach for session identity when session cookies do not work, which can happen when for example cookies get rejected by a browser's privacy settings. (For more details on how to support URL encoding, see http://tinyurl.com/33dsg.)
- Remember that an HTTP session is not guaranteed to be there, and that it is not in any way guaranteed to be transactionally consistent. For example, if a server fails, the session – or part of it, or even just the most recent updates – may be lost. The application must be built in a resilient way, expecting that such things could happen.

When configuring WebSphere, the key to obtaining HTTP session management nirvana is, like most things in software, knowing the requirements before deciding on the solution. Further, like almost all other tasks in software, the optimal solution is usually the simplest one. What are your requirements in the following categories?

- ***Session tracking:*** How do you want the session identity to be managed by the client? Should it be invisible in a browser cookie, or encoded into the URL via URL rewriting? WebSphere supports both cookies and URL rewriting. (There is an additional option for SSL-based session identity, but it usually requires the additional use of a cookie or URL rewriting to make it work.)
- ***Session size:*** How much data is in your HTTP session object? Is it tiny (less than 250 bytes including all names and values) small (less than 2KB); medium (less than 50KB); large (less than 500KB); or huge (above 500KB)? The larger the HTTP session object, the fewer the choices that are practical – or even possible.
- ***Session persistence:*** Should the session data be stored using a central store, such as any JDBC database? This may allow the session data to survive the failure of a server, or even the temporary shutdown of the entire application (i.e., all clones). Also, if the session is to be persisted, what is the tolerance for losing data? WebSphere supports an option that allows data to be lost in order to provide more acceptable performance for session persistence; you can choose whether you want better performance (the TIME_BASED_WRITES option) or more surety that the data is saved (the END_OF_SERVICE option).
- ***Session mobility:*** In a cloned or other multiserver environment, should WebSphere allow a session that was last accessed on one server to be accessed by another server? This is called session mobility and is usually the basis for session failover.
- ***Session failover:*** In a cloned or other multiserver environment, should session information survive the loss of any single clone or server? This can be accomplished in WebSphere by using either session persistence or memory-to-memory replication of sessions.

–continued on page 49

Cameron Purdy is the president of Tangosol, Inc., publisher of Tangosol Coherence, an in-memory caching and data management solution for clustered J2EE applications and application servers. Cameron is also the specification lead for the JSR 107: JCACHE - Java Temporary Caching API. In his free time, Cameron frequently writes, speaks, and blogs on industry issues, and is an active participant in the Java community. cpurdy@tangosol.com

TANGOSOL™

Spread the Load

With TANGOSOL COHERENCE™ many servers work as one. Partitioned data management ensures that each server carries its own load and contributes to the overall capacity and fault-tolerance of the cluster. Application capacity planning becomes a simple exercise in arithmetic.

www.tangosol.com

Coherence™ is pure Java software that reliably and cost-effectively scales J2EE applications across any number of servers in a cluster. Coherence provides reliable coordinated access to in-memory data that enables safe data caching and user session management for applications in a clustered environment.

Sales +1.617.623.5782 or sales@tangosol.com
Support support@tangosol.com or http://www.tangosol.net
Pricing US$1,995 to US$4,995 per Coherence production CPU
Development licenses are offered free of charge.

Jumping 2 Enormous Eggplants.

A new way to look at J2EE.

Cyanea/One gives you unprecedented visibility into mission-critical application performance. With the transaction-based performance management capabilities of Cyanea/One, you can drill down to the method level to determine and resolve performance issues. Cyanea/One monitors complex, multi-platform, service-oriented environments 24 x 7 from a single web-based console, all without having to touch the application code. If you are developing enterprise applications, you know the hurdles you sometimes have to overcome; with Cyanea/One, your applications will reach new heights.

For more information, visit www.cyanea.com/j2ee or call 1.877.CYANEA8.

Cyanea®