

Do Diligence Partners, Inc. E-Commerce Scalability Management Outline

Topic: Technical Roundtable on Scalability of Internet-based Applications

The purpose of this forum is to provide technology executives with the opportunity to review and share their experiences and thoughts on the issue with their peers.

The original discussion agenda promised:

- Current state of internet-based applications scalability - lessons learned, major challenges, etc.
- Comparison of Microsoft SQL/Windows 2000 and Unix/Oracle server-based technologies
- Outlook on application architecture scalability

Agenda:

Meeting Intro and Objectives – Kevin Burns

Introduction – Tom McKittrick

Scalability defined

Scalability management as a core competence – an example

Discussion

Status and objectives of each participant

The components of scalability management

Selecting the right data base ??

Wrap-up

The future of scalability management

Notes

Some scalability management products and services

Do Diligence Partners, Inc.

E-Commerce Scalability Management Outline

Introduction

Scalability

The ability of a system to adapt to increases in demand while maintaining or improving on response times.

A systems' ability to support an increased workload while maintaining adequate performance.

Scalability Management

Design, planning, measurement and management that results in ensuring system scalability.

Effective internet-based applications scalability begins with a strong application architecture and design that integrates compatible software components and considers the hardware and network infrastructure. Success is ensured through constant monitoring of workloads and system resources, and timely reaction to workload changes to ensure consistent performance is maintained.

In other words, you have to plan this carefully when you build an e-commerce application, then pay attention to intercept and correct problems and prevent surprises.

Scalability management as core competence – an example

Let's talk about how one organization is effectively managing scalability

The environment. The application consists of low transaction volumes, but high intensity transactions. It is a small application in terms of code, but complex in terms of transaction logic. The service is sold to large, multi-national corporations and used internally and by their suppliers. Multiple customers are supported on the same configuration, hosted at this time by a Managed Service Provider.

The Process

1. Dedicated manager of hosting (formerly performance management & capacity planning), charged with ensuring consistent high-level performance regardless of changes to the application and expansion of the customer base. No reliance on hosting vendor for this support, based on self-analysis of lack of experience and expertise at hosting vendors.
2. The CTO has extensive performance and capacity planning experience.
3. The 4-tier architecture isolates functionality to enable software and hardware analysis and tuning to overcome bottlenecks (client, enterprise server, application server, database server).
4. Stable platforms have been selected for implementation: iPlanet Web Server, BEA WebLogic Applications Server, and Oracle 8i Database Server (Oracle 8.1.6 at this time).
5. Hardware resources have been configured to ensure best performance. For example, they have configured enough memory / cache to permit the entire database to reside in memory on the database server.
6. The characteristics of the application's major workloads have been identified through close coordination with application designers, along with monitoring and analysis of the operational system.
7. Performance analysis and tuning are important parts of the quality assurance and testing process that accompanies new application releases and the addition of major applications functionality. This consists of stress and load testing. They simulate different loads using **Segue Silk** – record activity, edit scripts, playback.

Do Diligence Partners, Inc. E-Commerce Scalability Management Outline

8. Conduct customer workload analysis during 1 to 2 month customer trial period. After 5-6 months, when the application has been rolled out to the customer's entire enterprise and supplier base, they re-evaluate the workload. Re-evaluate periodically thereafter for each customer workload.
9. Evaluated hosting companies to pick the one that yielded the best upgrade, recovery, and security strategies at affordable costs. **Digex** was selected because:
 - Guaranteed hardware fix in 4 hours – including complete replacement, if needed. Part of fixed price (vs. fix yourself or \$150 / hour charge)
 - Guaranteed hardware upgrade (or addition of servers) in 30 days
 - Guarantee against denial of service attacks (as a result of TruSecure Service Provider (includes HIPAA) and SAS70 Audits)
10. Manage the hosting provider closely to ensure that service, reliability and upgrade guarantees are met.
11. Use real-time performance monitoring to identify resource consumption peaks that signal performance problems. Real-time monitoring using Freshwater Software's **Site Scope** (run by Digex for them). For CPU consumption, I/O, PageScan Rate (only way to tell if paging is getting out of hand on UNIX). Set target thresholds (example 200 pages/second – means not enough memory).
12. Use remote application availability monitor to logon from key remote locations at specific intervals and measure response. If degraded or no response, generate alert to hosting manager for immediate review.
13. Monitor daily utilization reports to further evaluate real-time anomalies and identify trends leading to poor performance. **CapaciTeam** service collects data, produces and emails reports. This has been successfully used to prevent runaway situations, such as Java code looping, consuming 30% of a processor.
14. Project 40% stable state utilization on any processor as the maximum before adding or upgrading processors; project a maximum of 3 processors in any tier before adding another instance of the application to support new users. Project 65 customers before that happens, but are constantly reevaluating this number as the applications and user workloads change.
15. Use workload scheduling to improve scalability. They publish a calendar that highlights periods of low to medium system activity so users can schedule their workload to ensure best responsiveness.
16. Will plan takeover of hosting in the future as soon as revenue permits. "Hosting is too critical to leave to an outsourcer over the long term."

Experiences managing scalability

1. Discovered a serious database I/O problem – data base update was causing peaks in I/O operations, taking too long. The data model had been optimized for space ("over-normalized), resulting in 5 way table joins, for example. In a subsequent application release of the software, they changed the data model to optimize for performance.
2. Suppliers on dial-up lines were experiencing long delays due to feature creep in the User Interface. Downloads of 1 ½ to 2 megs were causing disastrous response times. Eliminated some functionality, rewrote other features to improve performance.

Do Diligence Partners, Inc.

E-Commerce Scalability Management Outline

Discussion

Status and objectives of each participant

1. Is scalability a problem in your organization?
2. BRIEFLY discuss a critical scalability issue you have uncovered and resolved.
3. Do you self host or use an ASP or MSP? Name?
4. On 1-10 scale, where 10 means "we're highly experienced and expert at scalability, performance, and capacity management", where does your organization place itself?

The components of scalability management

Recognition

The need for effective management and control
The need to develop a core competency in-house

Management & Staff

Experienced experts provide a critical foundation
Management experienced in tuning applications and infrastructure
DBA's experienced in DB design and tuning for performance
Architects experienced in designing scalable solutions
Ensure collaboration to ensure success
Application requirements
Customer acceptance
Software design
Hardware architecture & deployment
Operations
Procurement
Performance and capacity planning

Applications architecture and design

n-tier applications
to suit the demands of the application
use components that scale to the needs of the application
Oracle 9 Real Application Clusters – Cache Fusion – consolidated data base
scalability
SQL Server 2000 - fully segmented data base performance
IBM DB2 – integrated information sources in distributed data bases
use established, production components
consider denormalizing data for performance
Introducing redundant data to reduce the number of joins needed to retrieve data from multiple tables. Example. If producing calculated values involves complex queries with many joins, add calculated columns to hold these values in existing tables.

Hardware selection and deployment

architecture that suits the application
SMP architecture for transaction processing
configure for maximum scalability
large db server memory
large cache
fiber-optic I/O
fast DASD
configure for fail-safe
RAID I/O

Do Diligence Partners, Inc.

E-Commerce Scalability Management Outline

redundant processors

Performance analysis and capacity planning

- establish this discipline as a core competency and develop a methodology
- use real-time tools and historic analysis tools or services to get a complete picture
- characterize workloads periodically and when new customers or functionality are introduced
- establish thresholds for key performance indicators & investigate exceptions
 - sustained processor load
 - sustained I/O loads
 - sustained paging rates
 - poor response times
- load test new software releases using simulated workloads
- consider influencing user workload scheduling through advertising, recommendation, and financial incentives

Operational Readiness

- Develop hosting vendor selection criteria (ASP, MSP - Digex, Loudcloud, Exodus, Allegrix, etc.) that consider scalability issues
- Establish service Level Agreements that provide
 - Short response time for:
 - Equipment upgrades
 - Outage recovery
 - Equipment repair or replacement
 - Protection from denial of service attacks

Selecting the right Database ??

Oracle vs. MS SQL Server vs. IBM

Some data points

1. Oracle 9i provides unlimited scalability

Nov 13, 2000. Oracle 9i, due to ship next spring, incorporates Real Application Cluster technology that lets IT departments add computers to a cluster without changing the application or the data, Oracle executives say. The technology achieves linear scalability beyond four nodes-doubling the number of users supported as the number of computers is doubled-without performance degradation, the company says. It also provides transparent application scalability by sharing frequently accessed data across all computers in the cluster, says Oracle.

The clustering advance features Oracle's patented cache fusion technology, which enables this architecture and also offers enhanced reliability.

Because this approach treats each node independently, applications still run even if a computer crashes. Oracle boldly claims that an application running on a 12-computer configuration will not experience a significant failure for more than 100,000 years.

ORACLE TOUTS 9i'S VIRTUES -- Makes bold promises about database's scalability, reliability, Computer Reseller News: Nov 13, 2000

2. Oracle & IBM Duke it Out

Do Diligence Partners, Inc. E-Commerce Scalability Management Outline

Oct 23, 2000. With Sybase and Informix recently losing market share and the SQL Server from Microsoft still not up to the level of IBM and Oracle products, the major competitors appear to be Oracle and Big Blue.

New front opens in database war - The ascendance of e-commerce has renewed the category's strategic role, InfoWorld: Oct 23, 2000, Ed Scannell, Tom Sullivan

3. MS SQL Server scores high – but will it work for you?

Oct 16, 2000. By taking the now well-demonstrated clustering path to performance, Compaq and Microsoft have scored the highest ever throughput score on the Transaction Processing Performance Council's TPC-C database benchmark, while maintaining one of the best price/performance ratios among all TPC-C results.

The firms' cluster reached stable peak throughput of 505,302 transactions a minute at a five-year cost of ownership of just under \$10.5m

Compaq and Microsoft beat an IBM result from July by taking advantage of slightly newer hardware -- in particular, 15,000rpm SCI Ultra 3 Compaq drives, which have been shipping for only a few months -- and using 192 CPUs compared with IBM's 128. The result illustrates the strength of shared-nothing approaches on data sets that can be logically broken up into smaller groups, or partitions. Shared-nothing clusters are made up of servers that manage their own local memory and storage. With this result, the top five TPC-C results are all on shared-nothing clusters.

However, as attractive as shared-nothing clusters are for some applications, many customers will not be able to use the technology, due to the centralised design of their databases. "About a third of our customers can take advantage of this scale-out architecture," said Mike Nikoliev, director of database engineering at Compaq. Other firms will need to redesign their databases or choose shared-disk software, such as Oracle parallel Server, which can accommodate a wider variety of data arrangements.

At peak throughput, the cluster was able to handle queries and updates from 432,000 simultaneously connected users and still keep response times to under 2.3 seconds for 90 percent of all requests. It ran 24 Compaq ProLiant 8500 servers, each with eight 700MHz Pentium III Xeon CPUs, SQL Server 2000 Enterprise Edition and Windows 2000 Advanced Server.

Clusters run faster, IT Week (UK), Tymothy Dyck

4. What's right for you? Oracle, IBM, or MS SQL Server?

Oct 16, 2000.

With SQL Server 2000, users can "scale out to a server farm to partition the workload of the database," Murchie says. "It's not right for every application, but it's a very economically attractive solution."

"IBM's data management's philosophy is that customers want integration of data, not centralization of data," Jones says. "Customers tell us that they have stuff in a lot of different places, and they don't have the ability to centralize it in one big database. They want it left where it is and managed where it is." Jones contrasts this with

Do Diligence Partners, Inc.

E-Commerce Scalability Management Outline

Oracle's approach, which puts everything in an Oracle table and lets Oracle be the application and database server.

"Oracle's leadership in security is light-years ahead of the competition," Jacobs says. New security features in Release 3 include support for SSL over HTTP connections to the database, and strong encryption for thin JDBC connections to the database, ensuring data privacy over the network. Integration with Entrust PKI enables Oracle environments to securely manage X.509 standard digital certificates.

Vendors Target E-Business Solutions -- Internet's growing influence spurs changes to database field, VARbusiness, Maria Kaplan

5. IBM and Microsoft respond to the challenge of e-commerce

IBM and Microsoft both have expended a lot of time and effort trying to get the reliability and scalability issues right. SQL Server 2000 is, by design, tied to the Windows 2000 operating system. That means you'll have to live with a single operating system, and any potential problems with that operating system, should you choose SQL Server 2000.

The close ties to the operating system present advantages in areas such as Microsoft's clustering technology. There's also full support for storage area networks, allowing the database to bypass the CPU and talk directly to the network. Benefits to this technique include faster database access and significant increases in overall database performance. Microsoft is working closely with third parties, including EMC Corp. and Compaq, on features such as serverless snapshot backups.

Scalability must be addressed from two perspectives: performance and size. Both companies rely on the underlying hardware and operating system to deliver the raw computing power needed to scale in terms of performance. DB2 will run on IBM's most powerful mainframe systems, providing orders of magnitude more performance than you would ever get on a Windows machine. Both products address essentially the same amount of data, making size a non-issue.

SQL Server 2000 and DB2 offer automated performance and tuning utilities that analyze a running database server and make adjustments to help it run faster. While the tools accomplish the same task, they go about it in slightly different ways. DB2 uses a cost-based algorithm to optimize the more critical items of the user's choosing. SQL Server 2000 attempts to match a pattern of database activity to that of a known workload and then react accordingly.

IBM has made great strides in optimizing its product to take advantage of the strengths of each platform on which it runs. While most of the code base for each different platform is essentially the same, a small portion of the code is reserved to deal with hardware differences. It's in these platform-specific areas that IBM seeks to get a big lift in performance. It appears to be paying off, as IBM has won a number of database duels and has taken on the Transaction Processing Performance Council suite of database-performance tests with great enthusiasm. The TPC tests attempt to measure the performance of a database program in performing very specific tasks such as insert queries.

DATABASES THAT FOCUS ON THE NET -- Updated versions of relational database management systems from IBM and Microsoft are Internet-aware, InformationWeek, Paul Ferrill

Do Diligence[®] Partners, Inc.
E-Commerce Scalability Management Outline

Do Diligence Partners, Inc.

E-Commerce Scalability Management Outline

The future of Scalability Management – a crystal ball

Reactionary in early deployment stages – concentration on time to market, feature delivery, and competition

Growing emphasis as e-businesses gain critical mass and their applications mature

New DB and Server Software will reduce problems - temporarily

Oracle 9i, SQL Server 2000

Video compression

Satellite services for point-to-multi-point transmission

New Hardware and Network Bandwidth will reduce problems – in step with needs

Better tools and services will emerge

Talent will be scarce – as always

Tradeoffs between cost and service will force difficult decisions

Notes

Some scalability management products and services

Load Testing

Silk Test, Segue, load and regression testing, www.segue.com

Loadrunner, Mercury Interactive, load testing,

<http://www-svca.mercuryinteractive.com/>

Workload Simulation & Load Testing

Silk Performer, Segue, www.segue.com

Site Load, Rational Software, www.rational.com

Workload Characterization, Forecasting, Performance Planning

CapaciTeam, workload forecasting and capacity planning software and services,

www.capaciteam.com

Athene, Metron, workload characterization, analysis, and capacity forecasting,

www.metron.co.uk

Keynote, e-commerce benchmarking and Web performance management services,

www.keynote.com

Real-time Monitoring

Patrol, BMC Software, www.bmc.com

Sitescope, Freshwater Software, www.freshwater.com

NetIQ AppManager Suite, www.netiq.com, www.webtrends.com

Pegasus, www.netiq.com, www.webtrends.com

I/Watch, Quest Software, Oracle & MS SQL monitoring, www.quest.com