Interview with George Linscott

George Linscott is president of CLAM Associates, the IBM Business Partner that assists IBM in creating High Availability Cluster Multiprocessing/6000 (HACMP/6000). CLAM began as a contract programming shop for IBM, completing AIX® networking, graphics systems, and X-Windows projects. Now that CLAM is intimately involved with HACMP/6000, it offers a full range of services for organizations planning an HACMP/6000 installation.

You are one of the founders of CLAM. How did the company get started?

Linscott: My background is in programming. I worked for IBM for five years in the Academic Information Systems group. During that time, I assisted with MIT's® Project Athena. The goal of Project Athena was to roll out 1,000 interconnected workstations from several vendors to undergraduates—so students could walk up to any UNIX® workstation and write a paper. This was the first time workstations from several vendors had been installed to cooperate with one another.

Many people from business came to see what we were doing. That's when I developed an interest in applying UNIX workstations to commercial applications (they were used almost exclusively in scientific and technical areas in the 1985–1986 time frame).

When did you start CLAM?

Linscott: We formed CLAM in 1987 as a software and services company. Three of us came from IBM: Les Comeau was my manager at IBM and Arnie Miller worked with me on the UNIX development team.

We did a lot of contract programming, such as porting NFS® to AIX PS/2®. The company started to move along—we had two or three employees! In 1990, we began working with HACMP/6000. Today, 85 people work here.

Where did you come up with the name?

Linscott: CLAM represents the first letters of the last names of the founders. We called the company CLAM Associates because we visualized getting all the really good people we knew together in one place—so we could do really great things.

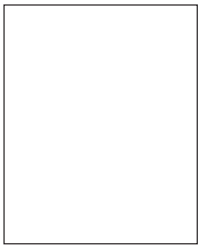
How about a brief description of HACMP/6000?

Linscott: There are two parts to HACMP. The HA part provides high availability—with two or more servers, each working with different databases. If one crashes, the others can take over the disks and restart the database. Companies that are running their businesses on computers should have HA so that downtime is minimal.

The CMP (cluster multiprocessing) part comes in when you have two or more servers concurrently accessing the same disk subsystem. We have done benchmarks against NCR® and HP™ that can outperform a Symmetric Multiprocessor (SMP)—depending on the workload—for very complex transactions like you would see in TPC-C, where the locking is minimal compared to the overall transaction. The CMP part provides extremely cost-effective scalable computing for large businesses.

So database design is really critical.

Linscott: Exactly. Even the best system can be "brought to its knees" by a poor database design. Designing a database system in this type of environment takes considerable thought. You can have four nodes in an HACMP cluster today, but if your design has a hot spot in one table, performance is significantly degraded. Database design-



George Linscott

ers prefer HACMP because they don't have the nightmare of partitioning (having more than one database). But they have to avoid hot spots. CLAM does a large amount of benchmarking to help customers get the best performance from HACMP.

You can really pump transactions through an HACMP cluster. Imagine up to four RISC System/6000s—each with hundreds of megabytes of memory—concurrently accessing a disk. Each has its own database cache and a separate connection to the network.

Could you tell us the history of the HACMP product?

Linscott: The original architects were Lynn and Ann Wheeler, a husband-and-wife team. The need was created when IBM made a commitment to a New York newspaper to create a high-availability system for their publishing needs. After a couple false starts in Austin, IBM asked the Wheelers—who had experience with high availability on MVS™ and were IBM employees in Austin—to do the job. The Wheelers said they would, if they could relocate to a beautiful IBM lab in Los Gatos, California and hire anyone they needed to help with the effort. Ultimately, management agreed. They contracted with CLAM (which happened to be me at the time) to help with the project.

Database designers prefer HACMP because they don't have the nightmare of partitioning.

When the Wheelers asked IBM customers if they would buy a UNIX-based high-availability solution if it existed, they received an overwhelming response. Based on that, IBM decided to make HACMP available to everyone.

It sounds like HACMP really kick-started CLAM Associates.

Linscott: We grew by meeting our commitments to IBM. We demonstrated competence; then IBM asked us to do more and more. When the product was released, we began to help with installations too. Since no one was doing training, we started conducting classes. We found that customers in a high-availability environment need special support; so we started doing 7-days-a-week, 24-hours-a-day (7x24) support for them.

Our main business today is services. Assisting with HACMP/6000 is our foot-in-the-door with many customers. Once we are there and demonstrate our competence, we may be asked to help with client/server design and implementation. We are also developing some of our own products now.

Tell us more about your systems integration and education business.

Linscott: We have discovered that to provide high availability, we need to know a lot about AIX as well as the network and clients. We take that expertise and help users design high-availability systems; then we perform the classic role of systems integration. Most customers are moving down from mainframes. We have a large project with Citicorp® where CLAM designed and developed a client/server application working with Citicorp and their IBM team.

CLAM is supporting the HACMP/6000 sales process. Do you also sell HACMP?

Linscott: Yes, we just opened our Commercial Systems Center in Austin. We will showcase AIX solutions along with IBM. As a Business Partner, we market products in Austin as well as from our Cambridge office.

What are the issues surrounding HACMP for developers?

Linscott: High availability is not a shrink-wrapped solution. You don't just buy a set of software, install it on a machine, and forget it. You need to do the same amount of planning as you would for a mainframe-based solution. High availability is a design point in which commercial systems need to be well thought out—from the server to the clients. It is really important to understand your entire operating environment. Many installed systems grew over time with very little planning, making it difficult to get the total picture. You would be amazed at the number of people we talk to who do not understand all the parts of their system. If you really want a high-availability solution, you have to understand how all the parts fit together.

You also need to understand the available technology alternatives, such as network topology and the capabilities of different communications protocols. For example, TCP/IP is more highly available than User Datagram Protocol (UDP).

How does the cost of supporting distributed systems compare with the same application on a mainframe?

Linscott: I believe that the maintenance cost of the systems is at least equivalent to the cost of maintaining similar mainframe-based solutions.

Time after time we see people frustrated; they have underestimated the effort needed to maintain some high-availability solutions. Keeping hundreds of client machines up and running is a lot of work.

Isn't there some problem with high-availability systems losing transactions if a switchover occurs?

Linscott: As with any OLTP system, transactions that are in flight when a crash occurs may be lost. Once HACMP has recovered from the crash and reconfigured the cluster, the database can perform log processing to ensure data integrity. Although transaction monitors could play a role here, we have yet to see widespread use in commercial installations. In a client/server environment, the client can have a role in providing transaction fault tolerance.

CLAM has a client software library called `clinfo` that provides the application with the state of the cluster. This library uses Simple Network Management Protocol (SNMP) to communicate between the client and the server. It keeps a current state of the connections and machines in the cluster. If the server crashes as the client is running transactions to that server, the application can redirect the transactions to the survivors. If the client application doesn't know which transactions have been committed yet, a restart is required.

You can build a fault-tolerant transaction-processing system by having the ability to restart transactions at the client. We are currently building some to show that it is a viable solution.

We are now building load monitoring into `clinfo`. Applications can perform load balancing, which directs transactions based on the load on the server. This software is available for AIX and Microsoft® Windows™.

How can readers learn more about this?

Linscott: Just call CLAM at (617) 621-2542 and ask for John Ranta. We have white papers describing how you can build a fault-tolerant transaction-processing system.

Could you give us an overview of HACMP support by the database vendors?

Linscott: This can be answered at several levels. First, any database that can recover from a hard shutdown and maintain data integrity is an instant candidate for high availability. From the standpoint of the database, an HA fallover looks like a very fast reboot. This means that all the major databases are suitable for the HA environment. Next, you want to examine the architectures available. There are two "religions" in the database world. The architectures of DB2/6000™, Sybase®, Informix®, and Progress® are based on a shared-nothing philosophy. Each machine owns its disk. Any machine that wants access to the tables on another's disk must go through the network. Oracle® and Mumps are based on a shared-disk philosophy. Also, Ingres® was the first to explore concurrent processing on a UNIX platform. CLAM ported the Ingres database to work in the CMP mode of HACMP. Today, Ingres does not support this product.

HACMP adds value even in a shared-nothing world. Given a replicated architecture, making a node highly available is still a winner in terms of minimizing downtime and masking minor failures. In a partitioned scheme, high availability is essential for each partition. In a shared-disk architecture, HA provides the mechanism for maintaining multiple highly available database servers and adds the requisite disk-sharing subsystem.

You can build a fault-tolerant transaction-processing system by having the ability to restart transactions at the client.

Are you seeing fault-tolerant customers migrating to HACMP?

Linscott: For many years fault tolerance was the only approach available to customers who needed high availability. Today, if you ask many fault-tolerant customers how long the server is up, they say 99.999% of the time. But then ask them how often clients or users on terminals experience an outage; many say once or twice a week (due to software crashes or network problems). At that point, we start talking about a total system design, which is necessary to provide front-of-screen availability. You see, it really doesn't matter how long that server is up; what really matters is how much of the time the person at the screen can work. HACMP is the most cost-effective way to accomplish that today.

Have HACMP-like configurations changed the basis for choosing shared-nothing versus shared-disk architectures?

Linscott: A shared-nothing strategy can be very scalable in a read-mostly environment. You can have 10, 20, or 40 machines that are all cooperat-

ing together because they are shipping around functions and data. But remember, these architectures take many years to put in place. These decisions were made five or six years ago when the theory about shared disks was that it did not scale well because of the limitation of a 2 MB/second throughput rate from disks. The architects did not consider that disk rates also move forward. So now, we have 10 MB/second SCSI-2 disks and IBM 9333 serial drives that are 32 MB/second. And vendors are talking about fiber-connected disks running at 100 MB/second in the future. With Fiber Channel Standard (FCS), you could have 64 processors hooked into a shared-disk system. Suppose the disk subsystem is a disk array with intelligent caching—it only makes the shared-disk approach appear better. For write-intensive OLTP applications, shared-disk clusters provide better scalability.

HACMP is nothing more than exploiting all the things that are already in AIX.

What's your advice to developers who know AIX, but have little or no experience with HACMP/6000?

Linscott: Developers need to do paradigm shifts quite often. HACMP is just another paradigm. Often programmers are afraid to make those paradigm shifts because when you're the best at what you do and move to a new environment, you end up at the bottom of the heap. Remember that HACMP is nothing more than exploiting all the things that are already in AIX. If you know AIX, you know HACMP/6000.

What competes with HACMP?

If you know AIX, you know HACMP/6000.

Linscott: NCR has a product called LifeKeeper™, which some analysts think is a better product. But when we go through a point-by-point comparison with these analysts, they conclude that HACMP is much better. So what's going on? Well, NCR has done a very good job at marketing their product. IBM is changing its marketing plans this year to address that issue.

HP has a product called Switchover/UX. I can't say that we have ever lost an account to HP because of Switchover/UX versus HACMP.

Is there a market for third parties to build off HACMP?

Linscott: Absolutely. Companies with products that require high availability are integrating HACMP into their offerings. Industries that we have worked with so far include retail, medical imaging, insurance, telecommunications, transportation, and power management. I expect to see many more companies reselling HACMP in the coming year.

What's in the future for high-availability systems?

Linscott: High availability is currently a thin layer over the operating system. As time goes by, more and more high-availability functions will be integrated into the subsystems of the operating system. There will still be a need for highly available application support above the operating system. HACMP should take advantage of evolving technology. Suppose IBM comes out with a four-way SMP. Customers will realize tremendous scalability by plugging SMP nodes into HACMP clusters.

Future capability that I would like to see is "disaster survivability"—clusters spread out geographically. Machines could be in separate buildings or separate counties. Since natural or man-made disasters can take out a big business for an extended amount of time, there is a great opportunity to protect businesses with geographically dispersed clusters. The internode communications may use Fiber-optic Data Distribution Interface (FDDI) from a telco at 100 MB/second. Bell Atlantic® offers plug-in FDDI in the Washington, DC area now. FCS and Asynchronous Transfer Mode (ATM) are other transports that are suitable for internode communications in clusters. In the next year or two as more telcos offer this, disbursed clusters will become a reality. This is actually a type of replication plus additional high-availability logic.

Is HACMP ready for prime time, or are there still missing pieces needed to create a complete solution for a customer?

Linscott: It's absolutely ready for prime time. If you are running a database today, you can bring in HACMP without any additional development and just a little systems integration. It will give you much higher availability than one processor. Clients will have to reconnect after a switchover, but that's better than being down for hours. IBM has over a thousand customers that are doing very well with it.