Graphics Leader SGI Reinvents Itself

January 30, 2002

Summary

SGI's January 29, 2002 announcement marks a significant initiative in making the company's technology more pervasive, affordable, and in line with market demands and application usage trends. The cornerstone of the announcement is the sharp focus on lower price/performance, both on the scalable Onyx class systems as well as on the desktop platform with the Silicon Graphics Fuel Workstation.

SGI is the company that first made graphics fashionable. With a long history and commitment to delivering industry-leading feature-rich graphics, SGI remains a step ahead of its competition in delivering high-end graphics solutions, often far ahead of the broad market expectations curve. SGI is reinventing itself again and realigning its energies and drive toward delivering its core competency in scalable solutions based on its ccNUMA (cache coherent Non-Uniform Memory Access) architecture and graphics heritage.

Technology Building Blocks

With its Onyx class systems, SGI is extending its software technology – OpenGL Vizserver – to launch a new technology initiative – Visual Area Networking – designed to make remote visualization and real-time virtual collaboration a closer reality from a systems perspective. With collaboration among multiple users spanning geographical distances and sharing limited resources a necessity, today's infrastructures are strained when they must display complex visuals over the network. These large data sets and assemblies exceed local graphics resources and require customers to spend excessively on hardware and software resources to facilitate high-speed communications and provide effective control to remote users. The OpenGL Vizserver fosters remote collaboration among multiple users (v2.0 supports two simultaneous users), providing them with the ability to visualize, share, and interact with the data, irrespective of their platform or operating environment. SGI's announcement spans support for handheld devices (running embedded Windows and Linux operating systems), beyond its existing support for Windows, Linux, IRIX, and Solaris desktop users to visualize data from an Onyx class system. (Collaborative visualization is likely to emerge as a major market opportunity for various players in the industry.)

The Onyx 300 extends SGI's InfiniteReality3 graphics offering by leveraging the NUMAflex design (modular building blocks) and features rich graphics in a scalable, high-performance compact-computing solution. The Onyx 300 allows customers to pack more performance in a denser platform while significantly lowering price and making InfiniteReality3 graphics much more affordable. In a dramatic shift for its Onyx 3000 series, SGI is introducing low-cost graphics to create a multipipe solution focused toward delivering rapid geometry and fill-rate performance scaling using a compositor module. The InfinitePerformance graphics based on SGI's VPro V12 graphics allows customers greater collaboration and performance potential with the low-latency, high-bandwidth architecture of the Onyx 3000. On the desktop front, SGI is revamping its price/performance standing

D.H. Brown Associates, Inc. ***www.dhbrown.com***

Our research program in Workstations and Graphics makes this Technology Trends available to all our subscribers. Those interested in this program should contact marketing@dhbrown.com.

with the new Silicon Graphics Fuel workstation, which has a powerful combination of high-memory bandwidth architecture (with DDR memory), latest MIPS R14000A processors, and VPro graphics.

Market Response Expected

The product and technology announcements from SGI should strike a significant chord with customers, both established and newer ones. First, it demonstrates SGI's commitment to continue innovating on new products and technologies, and secondly, it delivers solutions that solve real-world problems crucial to businesses that yield return on investments and competitive advantage. With Visual Area Networking, customers can challenge traditional approaches and develop new infrastructures to enable collaboration among remote users. This can be done without incurring significant overhead costs, and while maintaining total data integrity. Finally SGI is taking a firm stance behind MIPS and IRIX, the two key platform and system software components underlying SGI's products and future. While SGI's progress has been slower on the finessing aspects of processor performance, its recent introduction of the 600 MHz MIPS R14000A processor with competitive raw chip performance promises renewed vigor. SGI has demonstrated plans to support well over 1 GHz clock speeds in the near future.

Despite its long-standing presence in government (defense and research), education, energy, and enterprise organizations, with its focus primarily on technical computing (simulation, visualization), SGI has surprisingly stayed off the mainstream radar. Most of SGI's existing customer base deploys its systems toward primary application usage, critical to product development and maintaining a competitive advantage. The new products are likely to extend SGI's presence in newer markets, especially with the Visual Area Networks initiative.

Competition from HP and Sun

On the competitive front, HP is also developing an infrastructure similar to SGI's Visual Area Networking with its e-Utilica offering, combined with the HP visualization center sv6. Besides its investments in systems and technology, HP also benefits from a wide array of system-management tools (Netaction, OpenView) that might give it a competitive edge over SGI, although SGI remains ahead in the race. Sun is also gearing its products toward a similar goal, but has yet to put all the visualization pieces together.

Overall, D.H. Brown Associates, Inc. (DHBA) expects that SGI's expertise in graphics, combined with its market presence with early adopters, gives it a full lap advantage over the competition. The recent dis-investment in commodity Intel/Windows-based servers and workstations gives SGI a more focused approach toward delivering on its own intellectual property (IP). Further asset and market consolidation gives it a more stable financial platform increasing its viability and competitiveness. The key markets SGI is targeting with the new products offer a strong upturn and a large market opportunity. As discussed earlier,

DHBA does expect competition to react feverishly and to try to win mind and marketshare in this growing and very profitable niche. Long term, SGI still needs to drive further with services and partnerships with leading software application-tool vendors as well as management software players to put the entire framework in place.

Visual Area Networking

Real-time collaborative visualization has always been an elusive dream for hardware and software players. Increasingly, organizations around the world are facing the task of sharing visualization resources at specific locations with users who are often geographically dispersed and limited by their network bandwidth and graphics hardware. Several different, incompatible software applications and a myriad of operating systems and platforms add to the difficulties of long-distance collaboration. The engineering design market has been at this juncture for some time and various software vendors in the CAD market currently offer a wide array of applications to solve this problem.

Controlling actual CAD data and menus is the key. This requires ownership of the application or at least some ability to manipulate the data (interoperability). Some solutions depend on taking advantage of client/server architectures, with the data residing on a single server. Others take a peer-to-peer approach with fat clients. Most of the software tools in this arena are specific to the vertical market they target (e.g., MCAD) and require expensive add-on software to existing applications to collaborate effectively. Further, the opportunities for remote visualization extend well beyond the traditional design market into other disciplines including education, scientific research, government, and oil and gas. This presents opportunities for hardware vendors, both system and storage.

SGI is currently at the forefront among system vendors tackling this problem and has launched a new initiative, aptly dubbed "Visual Area Networking" to incorporate the underlying technologies. While the name implies similarity to Storage Area Networks (SAN), Visual Area Networking is similar only in the sense that it centralizes the resources – such as the massively scalable Onyx 3000 series visualization system – into a single or strategically positioned location. However, the similarity ends there. SANs require their own dedicated fibre-channel networks and talk mostly to back-end servers. On the other hand, Visual Area Networks are full-fledged infrastructures designed to serve users, either local or remote, and provide visualization as a utility.

The objective behind the vision is to:

- Provide advanced visualization capability such as that achieved by a local Onyx or Reality Center user to any remote client, irrespective of its platform or operating system.
- Allow many such users to effectively communicate with each other and interact with (manipulate) the data as if they were local to the server.

Many vendors including SGI have already addressed the first objective by supporting major desktop operating environments including Windows, Linux, and the various UNIX variants, with some caveats. Wireless tablets running Windows or Linux are also supported with OpenGL Vizserver v2.0. SGI plans to add support for other embedded operating systems including Palm OS and Windows CE, once their compute, display, and networking (such as 52 Mbit 802.11a) become sufficient to support interactive environments. In this case, the application and data mostly reside on the server. The compressed image is either sent in a proprietary format or an industry-accepted standard such as JPEG, and the pixels are assembled at the client device with minimal overhead. No rendering is required. Users can then zoom, pan, rotate and markup the image. The current release of SGI's OpenGL Vizserver software supports up to 16 users, each on one of the InfiniteReality3 or InfinitePerformance graphics pipes of the Onyx 3000 series system, or up to eight users on an Onyx 300 system. The key advantage of this approach is that the integrity of the data is maintained so no user can, even inadvertently, change data. The biggest challenge is to allow multiple users to interact with the data simultaneously, leading to the second objective.

The second objective is obviously more complex and difficult to address. Although the current OpenGL Vizserver software supports multiple simultaneous users on each graphics pipe, only one user can control the application. Some software CAD tools (e.g., Cimmetry) track and manage markups by creating sessions, so that changes can later be compared, verified and merged (or purged) by the local user. Other peer-to-peer model-based software (e.g., Alibre) allow changes to the data repository, which is available on each fat client that uses local resources. However, this implies duplication, putting the burden on maintaining an up-to-date copy of the data, besides security issues. While SGI continues to work with third-party software vendors to support peer-to-peer applications where distributed data makes sense, the company is firm on the importance of the client/server model for its OpenGL Vizserver software. The primary advantage relies on the computing horsepower and ccNUMA architecture of the Onyx family systems with InfiniteReality3 or InfinitePerformance graphics. Although some software-based tools provide additional features specific to the vertical, SGI's OpenGL Vizserver provides other benefits that are widely applicable:

- OpenGL based (works with any OpenGL-compatible application).
- Application transparent (does not require application to be modified to take advantage of OpenGL Vizserver).
- Client independent (platform and operating-system agnostic). Host application needs to run on MIPS and IRIX.
- The OpenGL Vizserver v2.0 allows two users to simultaneously interact with the data, extending its existing capabilities. Future versions will expand this capability to multiple users.

SGI already provides the ability to authenticate users employing a standard UNIX sign-on mechanism in addition to Kerberos, DCE, and others. Supporting SSL and VPN will also

be a crucial requirement as security becomes a major concern, particularly as more users are allowed to access and share the data, through the supply chain or partner portals. SGI does provide the ability to log per-user usage so that the manager can verify and add accountability. Besides SGI, HP and Sun are also exploring remote collaboration and have already put some pieces of the puzzle in place as discussed earlier.

Visual Area Networking extend the grid-computing concept to visualization, making these resources available to all users on the grid irrespective of their local graphics capabilities. Note that such a network does not imply dumb terminals. The majority of today's desktops possess a good deal of compute and graphics resources. Rather, the focus is on delivering massive capabilities, beyond those supported by current desktop alternatives.

Onyx 300 with InfiniteReality3 Graphics

The Onyx 300 adds InfiniteReality3 graphics support to the Origin 300 server (announced last quarter) to create a scalable graphics-visualization platform at a more affordable price point than the current Onyx 3000 series solution. The Onyx 300 combines the C-brick[1] and I-brick[2] supported in the Onyx 3000 series to create a compact compute model that leverages the NUMAflex design. Besides providing good scalability with up to 32 CPUs, the design allows the segregation of key system components, like processors, I/O, and memory, each with the ability to scale independently of the other.[3] This allows systems to be either compute intensive or I/O intensive depending upon the type of application, and also eliminates bottlenecks. Each compute module of the Origin 300 contains either two or four 64-bit MIPS processors, up to 4 GB of memory, and two disk drives in a 2U-rack form factor.

	Onyx 300 InfiniteReality3	Onyx 3000 InfiniteReality3
Processor	64-bit MIPS R14000	64-bit MIPS R14000
L2 Cache	2 MB	8 MB
Maximum No. of Processors	32	512
Maximum Memory	32 GB	1,024 GB
Single-System Image	Yes	Yes
ccNUMA Architecture	Yes	Yes
X-brick Support[4]	No	Yes
Maximum No. of Graphics Pipes	8	16
No. of Raster Managers	1, 2, or 4 per graphics pipe	1, 2, or 4 per graphics pipe
Texture Memory	256 MB per graphics pipe	256 MB per graphics pipe
Frame Buffer Memory	80 MB – 320 MB per graphics pipe	80 MB – 320 MB per graphics pipe
Entry U.S. List Price	$119,000	$164,000

The Onyx 300 uses the same G-bricks[5] found in the Onyx 3000 series, as shown in the table earlier, supporting up to eight InfiniteReality3 graphics pipes in a 32-processor system. The maximum number of graphics pipes in the Onyx 300 is limited by the NUMAlink module, which has eight ports. The Onyx 3000 series supports multiple R-bricks to support more CPUs and graphics pipes. Note that the maximum number of graphics pipes is not restricted architecturally. However, configurations exceeding eight graphics pipes have not been tested or certified.

As the table on the previous page indicates, the Onyx 300 provides limited capabilities compared to its more scalable Onyx 3000 offering. Both are highly competitive compared to other solutions in the market. Note that the graphics and CPU-expansion capabilities of the Onyx 300 are similar to the Onyx 3400, which also offers up to eight InfiniteReality3 pipes and 32 CPUs, but twice the main memory (64 GB), providing a higher CPU-to-memory ratio. Also, the L2 cache on each Onyx 300 processor is one-fourth of that in the Onyx 3000 series, a critical factor for users dealing with large data sets. The Onyx 300 can support up to two InfiniteReality3 pipes and a full 32 CPUs in a single rack, as opposed to only one InfiniteReality3 pipe and four CPUs in a single rack for the Onyx 3400. Both require multiple racks for full graphics expansion as the G-bricks take a considerable amount of rack space. However, customers can aggregate all the compute power in a single rack while using the remaining racks for graphics.

The Onyx 300 with InfiniteReality3 graphics offers a similar pixel fill rate and geometry performance to the Onyx 3000 with InfiniteReality3 graphics, besides other graphics features including a full 48-bit RGBA, eight-way multisampling, Digital Multiplexing (DPLEX) support, and a high-bandwidth subsystem. Nevertheless, Onyx 300 is a more compact solution compared to the Onyx 3000 series at a lower, more affordable price point. The Onyx 300 offers the lowest priced InfiniteReality3 solution for customers and will often supplement entry Onyx 3000 series configurations. There are, however, differences in the systems, most notably with the expansion capabilities. The Onyx 300 also does not support the X-brick, and therefore lacks support for the various XIO-based adapters including HIPPI, VME, and GSN. Customers should evaluate their requirements and measure each system's capabilities against them.

Onyx 3000 with InfinitePerformance Graphics

The InfiniteReality3 graphics remains SGI's premier high-end offering, but customers often do not take advantage of all the bells-and-whistles associated with this feature-rich solution. However, customers do want a scalable multipipe solution that takes advantage of the shared-memory, scalable, and modular architecture of the Onyx 3000 series, and scales performance and resolution capabilities more cost effectively. For these customers, SGI has created the new InfinitePerformance graphics, leveraging its desktop investments with the VPro graphics.

The InfinitePerformance graphics is based on the VPro V12 graphics, found in SGI's existing Octane2 and the new Silicon Graphics Fuel workstations. It has been slightly modified for the new architecture and to support multipipe configurations, but is binary compatible with the desktop product. This way, customers can move off (deploy) applications developed using the desktop solutions onto Onyx 3000 effortlessly. The VPro V12 graphics offers high geometry and fill-rate performance compared to other UNIX graphics solutions in the market. It also offers a strong set of features including 128 MB of unified frame-buffer memory with up to 104 MB of texture memory, 48-bit RGBA, and 96-bit hardware-accumulation buffer. Up to two InfinitePerformance graphics pipes are housed in a single V-Brick with dual-channel support on each pipe. Each pipe can either drive a single display or multiple pipes can be combined to drive a single display.

The other key component of this solution is the new hardware compositor, a zero-latency module that combines two or four InfinitePerformance graphics into a single digital or analog video output. Up to 16 InfinitePerformance graphics pipes are supported in a single Onyx 3000 series system using multiple levels of hardware compositors. The compositor allows performance scaling of both geometry and fill rate with each additional graphics pipe to offer a high-performance solution. The hardware compositor is stereo capable, and supports genlock and frame-synchronization capabilities.

SGI provides numerous software APIs, including OpenGL Performer (high-level scenegraph API), OpenGL Multipipe, and GLX Hyperpipe (configuring and managing multiple pipelines) for customers to take advantage of its scalable, parallel pipeline-graphics solutions. All of these APIs are supported on the Onyx 3000 with the InfinitePerformance graphics.

The Onyx 3000 with InfinitePerformance graphics is intended for customers wishing to have a cost-effective, scalable, and performance-oriented graphics solution that takes advantage of the massive bandwidth and architecture of the system. Each increment of the processing power adds 3.2 GB/sec. of memory bandwidth, 1.6 GB/sec. of interprocessor communications bandwidth, and 2.4 GB/sec. of I/O bandwidth. This is a key advantage over other competing solutions in the market that offer limited system and graphics bandwidth with high latencies. It provides the lowest entry-priced Onyx solution compared to the InfiniteReality3 price points. The InfinitePerformance graphics solution will complement the high-end InfiniteReality3 solution and will continue its focus on visual realism.

Silicon Graphics Fuel Workstation

SGI's desktop products are renowned for their graphics prowess, but with the Silicon Graphics Fuel workstation, SGI extends that legacy to delivering lower price/performance as well. The single processor-capable Silicon Graphics Fuel workstation revamps SGI's desktop offerings with a new high-bandwidth memory architecture, powerful MIPS R14000A processors, and VPro graphics. Key features of the Silicon Graphics Fuel workstation are as follows,

- Uni-processor capable supporting either 500 Hz MIPS R14000A with 2 MB of L2 cache or 600 MHz MIPS R14000A with 4 MB of L2 cache processors.
- Up to 4 GB 200 MHz DDR memory.
- 3.2 GB/sec. memory bandwidth.
- VPro V10 and V12 graphics support (dual-channel display option support for V12).
- Up to three 3.5" Ultra160 SCSI drives, 219 GB maximum internal storage capacity.
- Four integrated 64-bit PCI slots.
- Internal CD-ROM and tape-drive options.

The Silicon Graphics Fuel workstation offers considerable performance advantages over the existing Octane2 workstation. Besides the improvements associated with the faster processors (support in Octane2 will be available later this quarter), the Silicon Graphics Fuel workstation also sports a 200 MHz front-side bus, coupled with 200 MHz DDR memory, boosting the overall system bandwidth. The graphics bandwidth across the X-bridge is still 1.6 GB/sec. Key feature differences compared to the Octane2 are the standard 64-bit PCI slots, internal CD-ROM, and tape-drive options in the Silicon Graphics Fuel workstation. Unlike the Octane2, the Silicon Graphics Fuel workstation lacks standard XIO slots.

At an entry price of $11,495 (U.S. list), the new Silicon Graphics Fuel workstation significantly improves SGI's price/performance standing compared to the Silicon Graphics O2+ and Silicon Graphics Octane2 workstations. SGI needs this price/performance improvement as it competes against UNIX workstations, most of which have improved their respective price/performance in recent months with the introduction of newer processors. SGI will continue to maintain its low-end O2+ product line. Its dedicated following is attracted by its Unified Memory Architecture capabilities. The Silicon Graphics Octane2 is focused on customers running multithreaded applications or those taking advantage of its XIO options, such as the DMediaPro DM2 high-definition and standard definition video I/O option. The new Silicon Graphics Fuel workstation is expected to boost SGI's mainstream market presence, both with its price tag and support for more industry-standard components, such as PCI and DDR memory.

[1] C-brick: CPU Expansion Module; includes four MIPS CPUs with maximum memory density of 8 GB per module.
[2] I-brick: Base I/O Expansion Module; includes PCI Fibre Channel controller, system disk, CD-ROM, PCI-based digital audio card, 3 PCI slots, and Ethernet, IEEE-1394, serial, and USB ports.
[3] Memory and processors are packaged together into one compute module, so memory requirements exceeding 4 GB require additional compute modules.
[4] X-brick: XIO Expansion Module; includes four XIO slots to support HIPPI, GSN, VME, etc. cards.
[5] G-brick: Graphics Expansion Module; supports up to two InfiniteReality3 graphics pipelines.