White Paper

Choosing a Collaborative Visualization System for Your Organization

Table of Contents

1.0 Introduction

Your competitors are getting tougher and smarter. Your customers are demanding more. Your investors are demanding a better return. Improved productivity is a mandate. The ability to solve the toughest problems and the flexibility to adapt your processes to your organization's changing needs is the difference between success and failure.

You recognize that adoption of a multi-disciplinary collaborative workflow with advanced visualization capabilities is one of the keys to your organization's survival and growth. This paper presents you with an overview of the latest visualization technologies and how you can take advantage of them while avoiding known pitfalls in the adoption of a new framework.

2.0 The Ideal Collaborative Environment is Both Natural and Scalable

A key driver behind an organization's decision to purchase a collaborative visualization system is the acknowledgment that effective decision-making is most easily accomplished when all of the appropriate experts have an opportunity to be involved in the process. With a collaborative approach, the solution can be optimized for each of the appropriate disciplines simultaneously, thereby reducing the number of times the team has to go back to the drawing board. Collaborative visualization of data, designs, and models provides a natural framework and a common "language" for each of the domain experts to share their expertise, iron out points of contention, and negotiate solutions.

The ideal collaborative framework supports individuals working at their desks and allows them to efficiently integrate their work in with that of their team members. It also allows multiple teams to aggregate their work into a larger assembled whole. This full-scale integration allows teams from different disciplines to collaborate and solve cross-functional problems earlier in the development process. This collaboration can take place in conference rooms, in interactive sessions with other parts of the organization distributed around the world, and in theater-type settings for review of major milestones for the entire organization. The ideal framework adjusts readily to changing user requirements and a wide range of data sizes without onerous system administration or expense.

3.0 Two New Alternatives in Visualization Leverage PC Components

Historically, visualization solutions have been based on high-performance system architectures built with special-purpose custom semiconductors and proprietary graphics boards. The graphics capabilities of these solutions can achieve very high levels of quality with the transformation of very large amounts of data into visual representations that are easily comprehensible. Users of these special-purpose technologies report dramatic productivity improvements and well-documented return on investment (ROI). Widescale adoption was limited, however, due to the expense of the solution, which typically exceeded $300,000 U.S..

Multidisciplinary collaboration is a proven method for reducing the time to a quality solution. An ideal collaborative framework supports team members in a flexible manner.

More recently, consumer-level graphics technology driven by the PC gaming industry has grown powerful enough to address an increasing number of the individual tasks in a professional workflow. This has led to the creation of hybrid solutions, which incorporate PCs or PC components, to compete with the special purpose visualization solution. These hybrids are designed to leverage the economics of the PC world while addressing the requirements of a multi-disciplinary collaborative environment.

The two hybrid solutions that are currently employed are:

- graphics clusters based on multiple PCs and leveraging industry-standard CPUs and graphics processing units (GPUs)
- a newer technology, known as scalable graphics, which leverages industry-standard CPUs and GPUs in a high-performance shared memory architecture

The key difference between these two hybrids is in the supporting system architecture, with the first built on a networked cluster of PCs and the second built on a modular, high-performance architecture where the system memory is shared between all computational resources. How this difference impacts the results for your organization is detailed below.

4.0 Graphics Clusters – Linking Single Systems

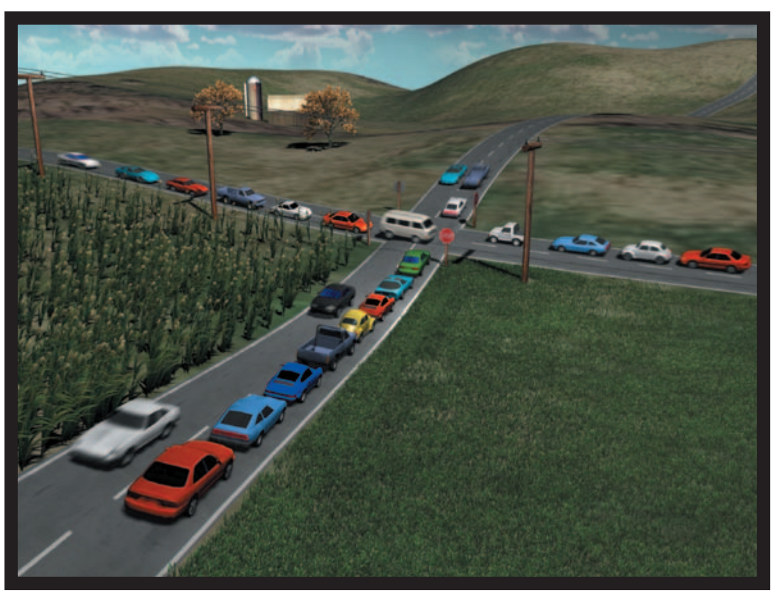
The increasing performance of the low-cost PC has led to the introduction of a visualization hybrid based on a dedicated network of individual PCs, called a cluster. Initially clusters were used in the high-performance computing world for raw computing capacity. Later, graphics cards were added to the PCs, creating a graphics cluster. Organizations that attempt a new workflow with graphics clusters report highest success when the size of the data that is exchanged between the individual PCs in the cluster is very small and they have ready access to a relatively inexpensive labor pool for system administration and software development.

Architectural constraints of graphics clusters include:

- Inadequate inter-PC infrastructure
- Software architectural issues
- The intrinsic memory limits of a PC
- Limited flexibility and reliability
- Operating expenses and hidden costs

(For analysis of the architectural considerations, see section 6.3: Underestimating Total Cost of Ownership.)

Clusters are efficient solutions for applications that are composed of lots of small jobs, such as compute tasks that are parallelized. If a node can work on its part of the problem with minimal communication with other parts of the system, good performance is typical. If, however, the computational resources need to communicate with each other across the system interconnect to achieve the desired result, then the communication burden quickly stalls the performance of the individual resources in a cluster architecture. These stalls can slow down the image update rate to the point where interactivity is lost. You can think of this as traveling across the continent on a country road. With limited traffic, you can make good progress, but multiple cars on the road can easily create traffic jams and there are points of delay at intersections between roads. Simply adding another road to the route does not easily resolve the traffic delays because of the intersections.



Traffic on a cluster architecture is similar to that on a country road. You can make good progress with limited traffic, but points of intersection on the road slow the pace. Additional traffic exacerbates the delays through the intersections.

The scalable graphics solution with shared memory is like an easily configurable superhighway.

5.0 Scalable Graphics – A High-throughput Shared Memory System Designed for Visualization

The most recently introduced hybrid solution and the first one to truly combine both the capabilities of the traditional special-purpose visualization solution and the economies of today's desktop technologies is the scalable graphics solution. The starting point for this hybrid is the scalable high-throughput system architecture of the traditional shared memory solution, which bypasses the limitations of the cluster approach. Entry prices for these systems are less than $45,000 U.S., substantially lower than the traditional special-purpose visualization systems, yet they include the key benefits of their higher-priced predecessors.

The advantages of the scalable graphics solution include:

- High-performance system with single system image (SSI) enabling linear scaling
- System designed to support multiple graphics pipes
- Flexibility and reliability
- Ease of programming and ease of system administration
- Large memory capacity
- Highly competitive total cost of ownership

5.1 High-performance System with Single System Image (SSI) Enabling Linear Scaling

The scalable graphics solution is a high-performance, integrated system that provides the power and flexibility to work on an organization's largest problems. At the same time, industry-standard components like CPUs and GPUs are leveraged in innovative ways to keep the costs down. System performance scales linearly as additional resources are added to the system, e.g., memory, CPUs, GPUs. The shared memory architecture means that computational resources are linked to create simultaneous access to data anywhere in the system's combined memory space. This shared memory eliminates the need for time-consuming data transfer.

5.2 System Designed to Support Multiple Graphics Pipes

The scalable graphics solution was specifically designed for collaborative visualization. It is a multiple pipeline visualization system that supports flexible video formats and image synchronization functionality, both in hardware and in software, that are critical for the deployment of an immersive visualization environment.

5.3 Flexibility and Reliability

A scalable graphics solution can range from very small systems designed for power users to very large systems, which can be logically subdivided for time-sharing across an entire organization. Shared memory systems can be partitioned on a moment-by-moment basis through easily reconfigurable software to focus all of the system's resources on one problem or to split the system's resources to work on lots of smaller problems. In addition, hardware components can be added in a modular manner to fit specific requirements, e.g., I/O or memory. In fact, a scalable graphics solution with shared memory is like a superhighway. As you have more data and get congestion, you can simply add another lane to the highway. Getting from one lane to another is simple and efficient. The lanes can even be assigned and reassigned for different uses, such as designating car pool lanes for use during peak hours.

5.4 Ease of Programming and Ease of System Administration

A scalable graphics system, although modular, is defined as a single system, and thus easy to program, service and administer. For instance, there is only one copy of the operating system to maintain. Because the memory space is shared, large models fit in to the memory space without additional programming. In addition, the system has a built-in interconnect so you don't have to buy or manage this functionality separately.

5.5 Large Shared Memory Capacity

The 64-bit architecture and the shared memory capability of the scalable graphics system combine to enhance the performance and data management capabilities of the scalable graphics system. Systems with global shared memory allow access to all the data in the system's memory directly and quickly without having to move data through I/O or networking bottlenecks. This reduces the time and resources required to visualize and interact with large datasets.

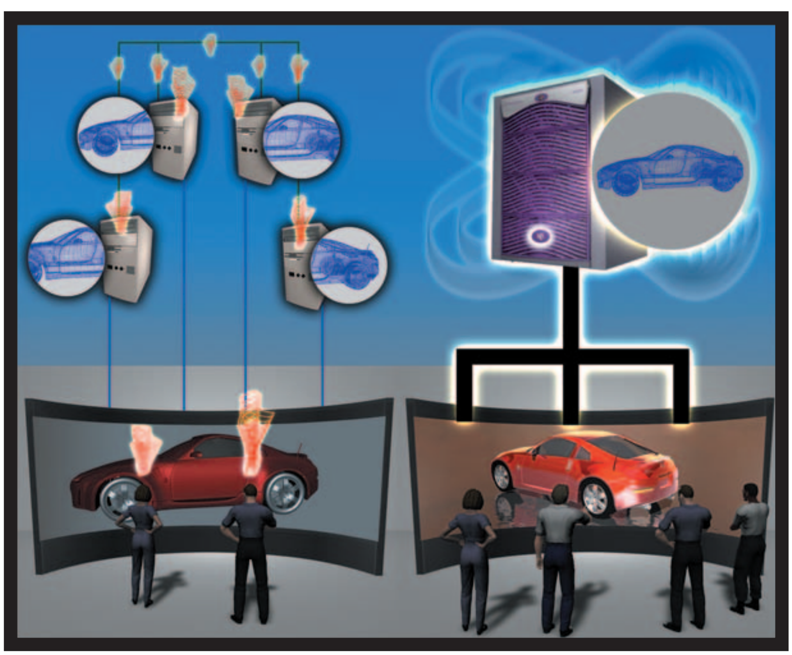
5.6 Highly Competitive Total Cost of Ownership

The scalable graphics system offers a lower pricepoint for advanced visualization than previously available. When its flexibility for time-sharing the system across an organization is exploited, it can deliver per user pricing that is comparable to typical workstation pricing. At the same time, however, it lets users interact with any of their largest datasets in a way that they could not on a PC.

The cost model becomes even more attractive when total-cost-of-ownership and operating expenses are factored in to the equation. There are many cost advantages associated with the fact that the scalable graphics solution is a single system with a modular structure. You only need one copy of the operating system and you don't have to create additional system interconnects. You have the flexibility to only purchase the components you need, as you need them, and those components are tested and warrantied for the system. Contrast this with a cluster, where additional capabilities may require the inclusion of an entire additional PC (which has been designed for obsolescence) in to the cluster network. In a scalable graphics system, memory, CPU, I/O, or graphics can be added independently, reducing costs and administration requirements.

The scalable graphics solution supports collaborative frameworks between cubicles, in conference rooms as a team room solution, across the entire distributed organization with key technologies like Visual Area Networking[1], and in theater environments for management reviews. Thus, a scalable graphics solution provides the most powerful, the most cost-effective, and the most flexible option for most organizations.

[1] For more information about Visual Area Networking, see www.sgi.com/visualization/van



Key differences between the two hybrid approaches to collaborative visualization are the capabilities of the underlying architecture:

- The cluster approach often does not have the capability to integrate large-scale multidisciplinary models.
- The scalable graphics architecture is designed to bypass the limitations of the cluster approach and create an efficient and flexible framework for large-scale visualization.

6.0 Pitfalls to Avoid When Choosing a Visualization Framework

As organizations move to new workflows and incorporate new technologies into their existing infrastructure, there are known pitfalls to avoid. Among these are:

- Inadequate assessment of overall workflow
- Overlooking version control requirements with shared data
- Underestimating total cost of ownership

6.1 Inadequate Assessment of Overall Workflow

At the highest level, the most important thing to consider is the willingness of the users to embrace the new workflow. If the new workflow does not deal with fundamental requirements then it is likely to fail, despite your management team's sponsorship.

Real-world performance of the collaborative visualization framework is critical from both a productivity and an acceptance perspective. If the new framework introduces bottlenecks or critical pieces of the workflow are not interactive, then users will become frustrated and reject the new processes. When evaluating a framework, measure the performance of the key applications running on the system as it will actually be configured rather than relying on theoretical performance measures or raw performance benchmarks. For instance, copying large data sets from one computer to another or from a shared storage system to a local drive can easily take more time than running the actual analysis, yet these issues are rarely considered when buying new technology.

Solutions that require the users to completely change their style of work, such as abandoning their desktop, may not be readily accepted. Some change is required to create collaboration, but the less intrusive it is and the more natural it is at the outset, the better. Collaboration technologies that work both at the desk, leveraging existing desktop assets and in a conference room or larger setting remove adoption hurdles.

6.2 Overlooking Version Control Requirements with Shared Data

As an organization's workflow migrates to a multi-disciplinary approach, users will be working with larger data sets that enable a holistic view of the problem or final product. The benefit is that issues are highlighted earlier in the process and there is less revisiting of decisions. Managing the larger datasets puts heavier demands on the computing infrastructure, however, so technologies that minimize the proliferation of different versions of the model become imperative. There are many mature technologies that can be added to existing environments to minimize the impact of these integrated datasets and keep overall workflow performance high. Visual Area Networking, for example, allows interactive collaboration between distributed users without data replication or transfer.

6.3 Underestimating Total Cost of Ownership

Another significant pitfall to avoid is basing your technology purchasing decision purely on the acquisition costs rather than analyzing the costs over the life of the technology. The hybrid graphics cluster offers a low-cost solution at introduction, but additional costs for programming, software licenses, system administration, power, and cooling, as well as the productivity costs of workflow inefficiencies must be factored in. For example, each node in a cluster requires its own copy of the operating system and each must be maintained, adding complexity. Further, with large data visualization, the graphics cluster quickly arrives at the point where adding additional systems to the cluster does not achieve the desired result. (For a description of the inter-PC infrastructure, see section 4.0: Graphics Clusters -- Linking Single Systems.)

In addition to the communication burden inherent in the cluster architecture, consider:

- Software architectural issues
- The intrinsic memory limits of a PC
- Limited flexibility and reliability
- Operating expenses and hidden costs

6.3.1 Software Architectural Issues

Cluster architectures are well suited to "divide and conquer" type applications, but they perform very poorly on problems that do not divide up into small independent chunks. Interactive, collaborative visualization, by its nature is not easily divisible. While theoretically data can be decomposed in to smaller chunks to be easily managed by the PC architecture, in practice logical model elements do not decompose easily within graphics pipe boundaries. With visualization applications, it is not feasible to program for a single view of the problem. The view will change dynamically as the user rotates and manipulates the model.

Furthermore, recomposition of the elements into a coherently visualized whole is not a trivial project. The PC architecture was not designed to handle the large sizes of integrated multi-disciplinary models or to provide the image synchronization necessary in large-scale collaborative environments. The programming requirements to adapt an application to a cluster environment must be factored in to the purchasing evaluation.

6.3.2 Intrinsic Memory Limits of a PC

The intrinsic memory limits of a PC significantly reduce the level of interactivity that is achievable with visual data. The memory limitations also force unnatural reductions in model fidelity or size. Memory can be added to the cluster, but because it is not shared between resources, it does not help get a node over its individual limit. Some clusters have very large aggregate memory, but most of that memory is wasted because it is simply a copy of the same data that the other nodes already have. As a result, the cluster architecture can restrict an organization's ability to achieve multidisciplinary collaboration because some members of the team are using higher-performance systems and applications that are beyond the scope of a PC.

6.3.3 Limited Flexibility and Reliability

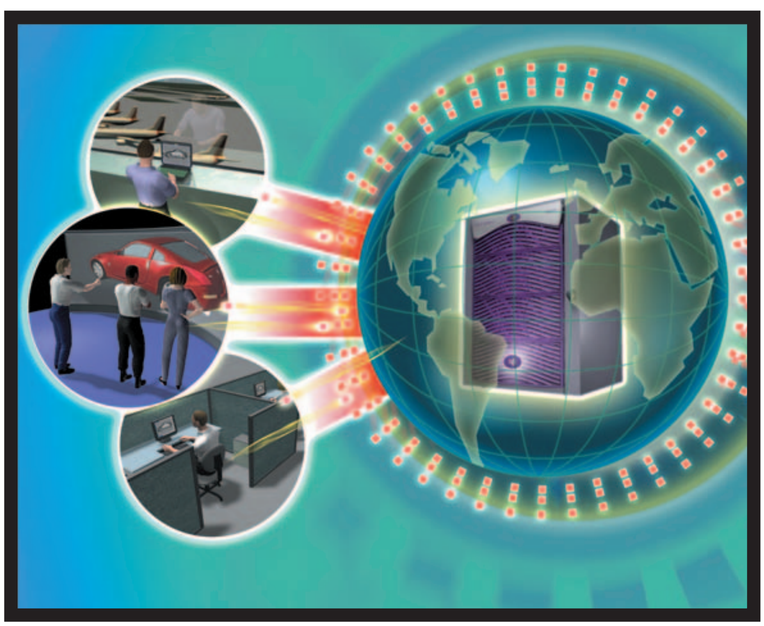
Clusters are limited in their flexibility despite the fact that they are composed of individual elements. Once a cluster is created in a specific network topology, it is not easily reconfigured for individual access. This means that clusters lack the architectural flexibility to easily support the varying levels of collaboration that are natural to an organization. In the ideal collaborative framework, you can assign all of your resources to a particular problem when needed, while allowing subdivision of the resources at other times depending on the workflow.

Another issue is that while some tradeoffs in reliability are acceptable in low-cost PCs for individual use, this trade-off has broader implications when it is a corporate resource. A single PC failure affects a single user, but when the PCs are tied together as a cluster, a failed PC means an inoperative cluster, which means lost productivity for the entire organization. Similarly, parts in each system may age at different rates, but if one key component fails, every similar component in the cluster may have to be replaced for consistency. For example, if a single GPU fails, every GPU in the cluster may have to be replaced in order to obtain a consistent image. This, in turn, may require new software.

6.3.4 Operating Expenses and Hidden Costs

Organizations that have implemented graphics clusters report that the operating costs of graphics clusters include many costs over the long-term that were not visible at the outset. This includes additional system administration, the high replacement cost burden mentioned above, additional software licenses for each node in the cluster, lost productivity due to poor reliability, and the cost of long-term maintenance. The strongest advocates of clustering for HPC or visualization have been found in the educational realm where large numbers of graduate students are available to administer and reconfigure clusters much more affordably than full-time system administrators in commercial or government organizations.

Superior solutions will improve productivity in the context of the entire workflow and will reduce costs over the lifecycle of the investment even as the size and complexity of your organization's data grow.



Scalable graphics supports a natural workflow with flexibility that includes multisite, interactive collaboration. The resources can be focused on one large problem, allocated to multiple team rooms, or assigned to individuals in their work areas, even across continents.

7.0 The Power of Visualization – Providing a Common Collaborative Framework

In driving your organization to the next level of success, you can employ the advances of the scalable graphics solution to create streamlined, multidisciplinary workflows without disrupting the current work rhythms of your team. Creating an infrastructure that is cost-effective and provides the flexibility to grow is now within the reach of every organization. The scalable graphics solution with its high-performance, flexible architecture delivers an economical, easily maintainable infrastructure for collaborative visualization that can dramatically improve your growing organization's productivity in the 21st century.

Corporate Office
1500 Crittenden Lane
Mountain View, CA 94043
(650) 960-1980
www.sgi.com

North America +1 800.800.7441
Latin America +55 11.5509.1455
Europe +44 118.925.7500
Japan +81 3.5488.1811
Asia Pacific +1 650.933.3000

3661 [06.28.2004]
J14618