# HP's Strategy for Scalable Visualization

Kevin Lefebvre
Hewlett-Packard Company
3404 E. Harmony Rd.
Ft. Collins, CO 80525

phone: 970-898-4382
fax: 970-898-6318

kevinl@fc.hp.com

## Abstract

Hewlett-Packard has developed a vision for scalable computing and visualization that includes a solution roadmap that meets the most demanding business requirements in computing and visualization. This paper describes the problem that technology users are facing and the strategy HP is following to produce visualization products to meet these users needs.

## Introduction

Because businesses today operate in an increasingly competitive global market they need to adapt to new requirements in order to survive. They need to:

- decrease the product development cycle time and costs
- while increasing product quality, performance, and features

Successful businesses are changing the way they operate in order to meet these new challenges. Teams develop whole product designs rather than individual components while the members of these teams are cross functional and globally located. External partnerships are formed to allow the teams to focus on their core competencies. Concurrent design processes are used rather than sequential to decrease design time while expensive time consuming physical prototypes are discarded and more virtual prototyping is used.

In this new environment better tools are needed for managing and distributing large designs, communicating and visualizing these designs to senior management and potential customers, and collaborating on these designs between multiple geographically disperse teams. More computational and visualization power is needed to make these tools, processes, and in turn the company itself successful in this new way of doing business. Over time the computation and visualization needs will undoubtedly increase. HP is working with these companies to define and develop computational and visualization solutions that meet their current needs and scale as their requirements increase.

**Vision**

HP's scalable computing and visualization vision is to provide on demand access to the complete computing and visualization resources owned by an enterprise to better leverage their investments in IT equipment. The intent is to provide transparent access to these resources from any point in the enterprise and to deliver visualization results to any point in the enterprise. The purpose of this vision is to more efficiently utilize the available resources within the enterprise and combine the individual resources together to solve or visualize problems too large for any single system.

## Building Blocks

HP is in a unique position being able to deliver a range of computing, visualization, and networking solutions. From the desktop to the backroom compute center, from a single monitor to high-resolution projection display, and from the Intranet to the Internet, HP has the products to meet businesses needs. Coupled with support and consulting services, HP can deliver complete solutions.

**Computing**

HP has continually delivered improvements in computational capacity through improvements in processor architecture, increases in clock speed, and number of processors in a product. Computation scaling over time will always be available via processor upgrades but this may require replacing existing systems with newer ones.

At the SPU level HP provides a range of computation capability products. These products vary in clock speed and number of processors. Workstations are available in desktop, deskside, and rack mounted configurations with 1-2 processors. Server configurations are also available in commercial or technical market configurations, these systems support 4-64 processors.

**Scalable Computing**

Computational scaling is available in both workstation and server based configurations. In the case of servers, scalability can also be achieved by adding more processors till capacity of the particular system is reached. For workstations compute ranch configurations exists. Using the rack mountable dual processor J6000 20 systems can be mounted in a single 2.0 meter rack.

Figure 1 shows examples of some of the HPPA based systems HP markets.

**C3600** | **J6000** | **J6000 compute ranch** | **L-class server** | **Superdome server**

**Figure 1 - Example Workstation and Server Products**

### Graphics

On the visualization side HP has a solid history of increasing graphics functionality and performance while reducing cost and size. The current Visualize-fx architecture has gone through several chip and board implementations. HP's current Visualize graphics lineup that supports OpenGL includes the *fxe*, the $fx^{5}$ and $fx^{10}$.

With the current Visualize-*fx* architecture, graphics performance can be scaled at the chip and board level. In the latest incarnation of the architecture the geometry and rasterization engines per chip have increased while the physical chip size has been reduced. With the current round of products scaling along the performance axis can be achieved by moving from the $fx^{5}$ to the $fx^{10}$, along the image quality axis by moving from the $fx^{10}$ to the $fx^{10}$b which has support for scene anti-aliasing. These level of scaling are accomplished through board designs that utilize more geometry, rasterization chips or memory. HP also supports scaling along the resolution axis, multiple graphics cards can be joined to present a single large resolution frame buffer. The cards can reside in the same SPU or in different SPUs interconnected via a high speed LAN. This technology is transparent to the application and the user. HP delivers this technology in the form of the Visualize Workgroup and Visualize Center. If more resolution is needed this additional systems can be utilized to expand the frame buffer in both x and y.

As can be seen the Visualize architecture can be scaled at the expense of chip development or new board designs. For an extreme high-end visualization system the return on investment of chip and board development may not justify the expense due to the lower volumes of these types of systems unless the system can be sold at a very high price. Assuming the ROI is sufficient and the product is developed there is a limit to how much can be put on a single board given physical size of the board, and the power and cooling constraints of the system the board is connected to. While a multi-board solution may solve the single board physical size limitations and possibly the power problem it introduces other problems of high-speed board interconnections and puts more requirements on the system the graphics is connected to. Finally if all the design problems and resolved and the product is built and delivered there are still visualization users that will come up with larger more complex models that require more compute and visualization power than is available in the high-end system.

**Scalable Visualization**

HP's strategy is to combine our strengths in scalable computing and high-end graphics to produce a scalable visualization solution. By using standard high-end graphics cards coupled with the standard compute ranches or servers and the basic technology of the Visualize Center, HP can deliver a scalable visualization solution. As more visualization or compute power is needed addition rendering systems or graphics pipelines (J6000/*fx*[10]) can be added. As technology improves SPU and/or graphics can be upgraded. A visualization user can start with a minimal system and grow it over time as their visualization needs increase. Figure 2 shows how these base products can be combined to create a scalable visualization solution.

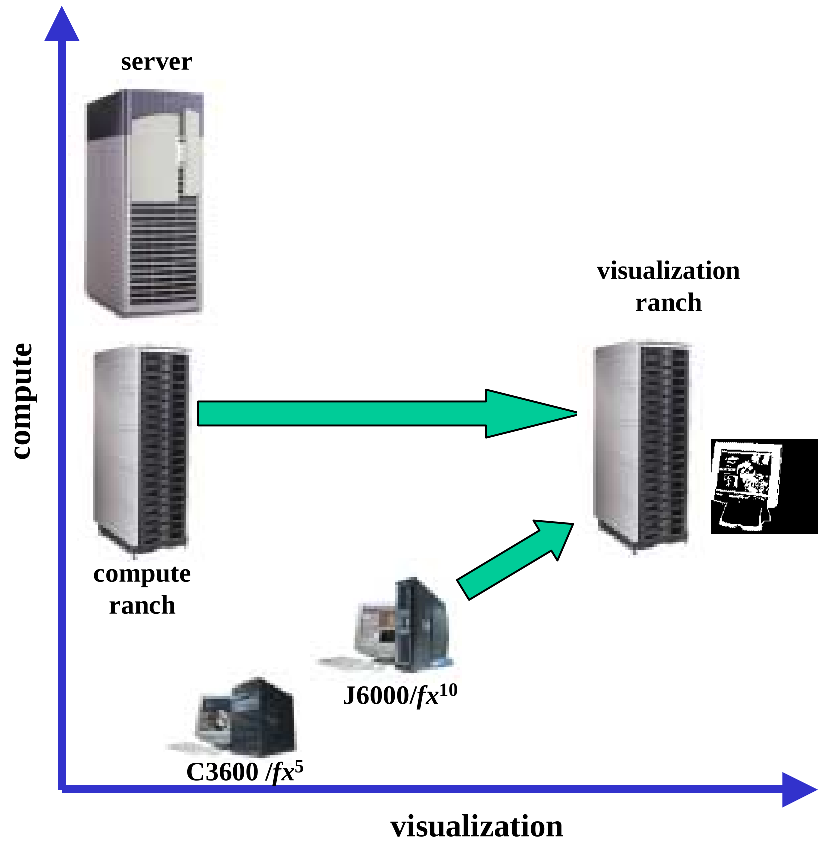


**Figure 2 - Scalable Visualization Components**

## The Technology of a Visualization Solution

The technology behind HP's scalable visualization has been designed to meet or exceed the performance and image quality needs of the most demanding user and to be able to grow and scale as their needs increase over time. The key attributes that make this system unique from existing scalable solutions include:

- scalable performance, image quality, and resolution
  - rendering resources can be added over time as needs increase
  - these resources can be dynamically configured to contribute to performance or image quality needs
  - the system can be assembled to drive a seamless multi-screen high resolution environment
- use of standard SPU and graphics components
  - scalable system immediately leverages the industries investment in components (SPUs and graphics) that are in the mainstream of technical advancements

For users of this technology the key benefits of this architecture include:

- The applications the customer uses today can run on this system unmodified and take advantage of the scaling and image quality improvements immediately
- The use of standard components with the ability to add, re-deploy, or upgrade these components provides significant investment protection
- The individual SPUs of the scalable system can be used as computational resources when not being used for visualization

**Architecture**

The goal of the scalable visualization architecture was to meet the performance and image quality improvement needs required by the users of these types of high-end visualization systems. The architecture also had to allow for the changing needs of performance and image quality in a single session where the resources were fixed. Finally it needed to be able to scale as the users needs increased over time without complete replacement of their existing system.

To meet these needs the architecture was defined as a set of high performance 3D rendering pipelines. These pipelines could be individually configured to contribute to performance acceleration or image quality improvement.

Performance acceleration is achieved by partitioning the screen in to sub-regions and assigning a pipeline to each sub-region. Each pipeline is responsible for rendering only the image geometry that is visible in its sub-region, geometry outside the sub-region is quickly discarded. In this model fewer polygons are transformed and rendered per pipeline while multiple pipelines operate in parallel, hence in a perfect world a 1000 polygon model distributed across 10 pipelines would result in each pipeline rendering 100 polygons. If each pipeline completes its rendering in equal time then a 10X-performance improvement can be achieved. Unfortunately it is not a perfect world and the polygons to not always distribute evenly across the pipelines, nor do they all finish at the same time, but with the use of dynamic pipe re-balancing effective use of all the pipelines can be achieved. With additional pipelines the screen can be divided into smaller regions, more pipelines can operate in parallel, each pipeline has fewer polygons, and greater performance improvements can be had up to the limit of how fast the application can pass geometry changes to the graphics system.

Image quality improvement is achieved via a scene anti-aliasing technique known as jitter. Scene anti-aliasing is not done through screen partitioning but rather by having each pipeline render the same image with the rendered image moved or jittered by some fractional sub-pixel amount differently on each pipeline. The final image is arrived at by taking the complete images from each pipeline and summing value of the corresponding individual pixels together and dividing by the number of pipelines thus getting an average pixel value based on some number samples.

In both modes, performance and image quality, some agent is necessary to take the output of the individual pipelines and combine them appropriately. Either the agent must piece together the individual pipeline sub-regions into a single image in accelerate mode, or sum and average the complete images coming from the pipelines in scene anti-aliasing mode. In the scalable architecture this is the responsibility of an image compositor.

The complete scalable visualization technology consists of a rack-mounted cluster of graphics enabled workstations. The systems are interconnected with a high speed LAN between all the SPUs and their graphics card outputs are connected to a digital compositor. Modifications to the OpenGL and Xserver software enable applications to distribute their rendering across the systems transparently while the compositor assembles the renderings from the individual systems into a single image. Figure 3 shows a high-level block diagram of the scalable visualization architecture. It also shows how the components are interconnected for passing rendering commands from the application to the rendering pipelines and images from the rendering pipeline to the compositor.

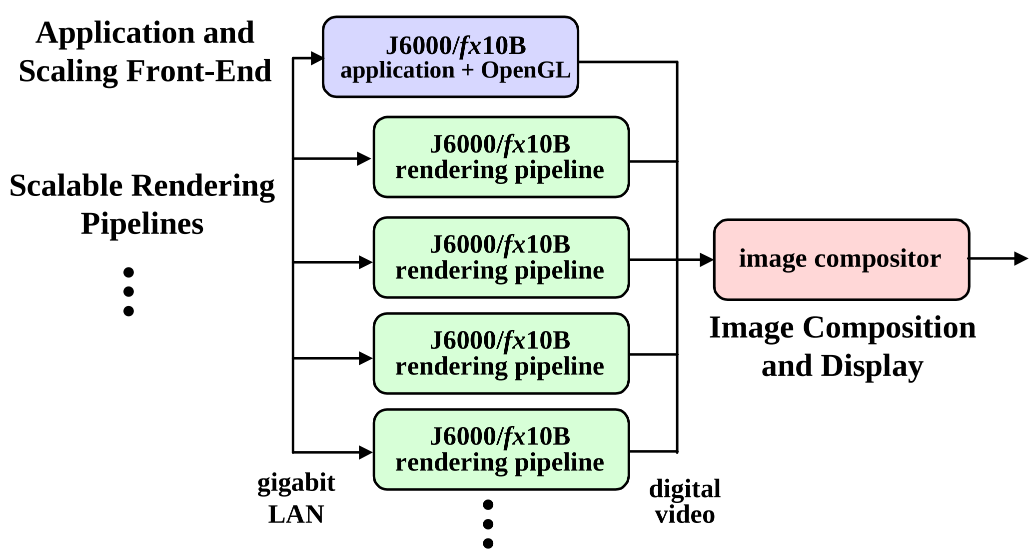


**Figure 3 - Scalable Visualization Block Diagram**

The primary components of the of the architecture include :

- Application master - this system runs the application and is responsible for collecting and distributing the 3D rendering commands. Depending on the application needs this system could be a simple rack mounted workstation or be one of various types of multi-processor servers.

- 3D rendering pipeline - each pipeline contributes to the rendering of the entire image either by only rendering a portion of the image in order to accelerate rendering or by rendering the entire image jittered by some sub-pixel offset so as to contribute to scene anti-aliasing. By increasing the number of pipelines the performance and/or image quality can be greatly enhanced.
- SPU interconnect - the application master and rendering pipelines communicate via a high-speed connection. Software on the application master and rendering pipelines interact to maximize the utilization of this connection. As connection technologies improve (bandwidth increases and latency decreases) the class of applications that run well on this system will expand.
- Graphics interconnect - the graphics cards communicate their rendered image to the compositor via a digital video interconnect.
- image compositor - this hardware component takes the digital video streams from all the rendering pipelines and reconstitutes a single image merging the sub-regions generated by the pipelines in accelerate mode or adding and averaging the pixel values generated by the pipelines in scene anti-aliasing mode. The application master system communicates with the compositor to define what type of image data is coming from each pipeline and how it should reconstitute the image.

### Screen Space Division/Image Composition

Scalable visualization provides a very flexible scaling environment. Figure 4 shows an example of the compositing operation.

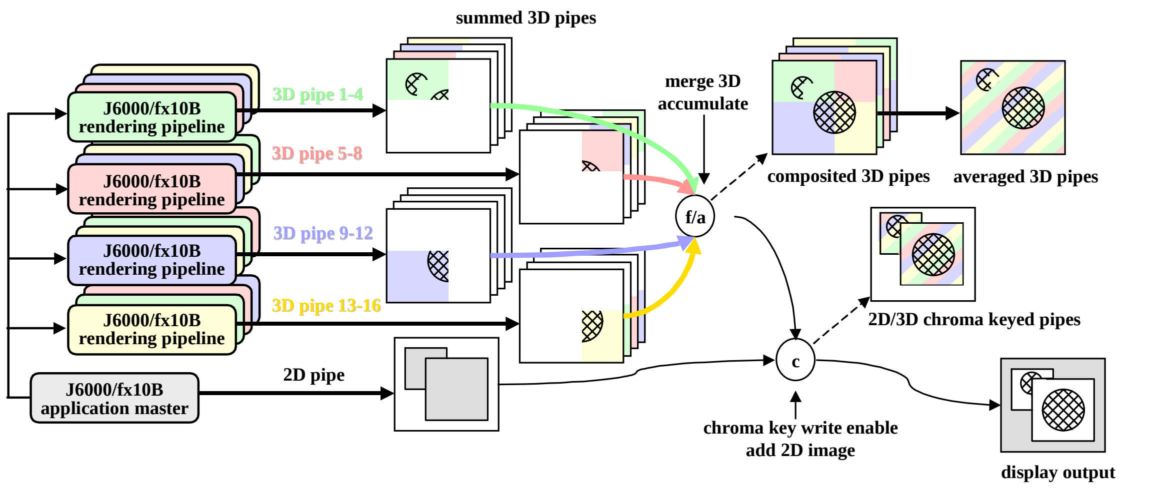


**Figure 4 - Compositing Operation in Mixed Accelerate/Scene AA Mode**

The multiple rendering pipelines can be configured to increase rendering performance or improve image quality. As visualization needs increase over time (performance or image quality) additional rendering pipelines can be added to the system. In some situations space is already available in the rack, compositor, and network switch, all that is needed is the J6000/*fx*[10] pipeline. In the figure 4 a 16-pipe system is configured to split the screen in to 4 sub-regions. Each sub-region has 4 pipelines rendering the same jittered quarter image producing 4 scene anti-aliasing samples. The compositor simultaneously sums and averages the appropriate quadrants while also piecing back together the complete image. A 16-pipe system could be operating in any one of the following modes:

- 16 pipe screen division - full acceleration mode
- 8 pipe screen division with 2 scene aa samples per pixel
- 4 pipe screen division with 4 scene aa samples per pixel
- 2 pipe screen division with 8 scene a samples per  pixel
- 16 scene aa samples per  pixel - full scene aa mode

The system can be dynamically switched at any time by user or application control between the various modes depending on what aspect is more important. In addition the *fx*[10] super-sample scene anti-aliasing feature can be enabled, effectively 4X'ing the number of samples per pixel (1, 2, 4, 8, 16 in the above mode list goes to 4, 8, 16, 32, 64). Enabling this mode does decrease *fx*[10] performance somewhat and reduces available texture memory but it is available if needed.

**Visualize Center**

As with the original 3-pipe Visualize Center that increased screen resolution by dedicating a pipe per display , this new scalable visualization architecture can be configured to support high resolutions. Figure 5 shows the system configuration to support a scalable Visualize Center.

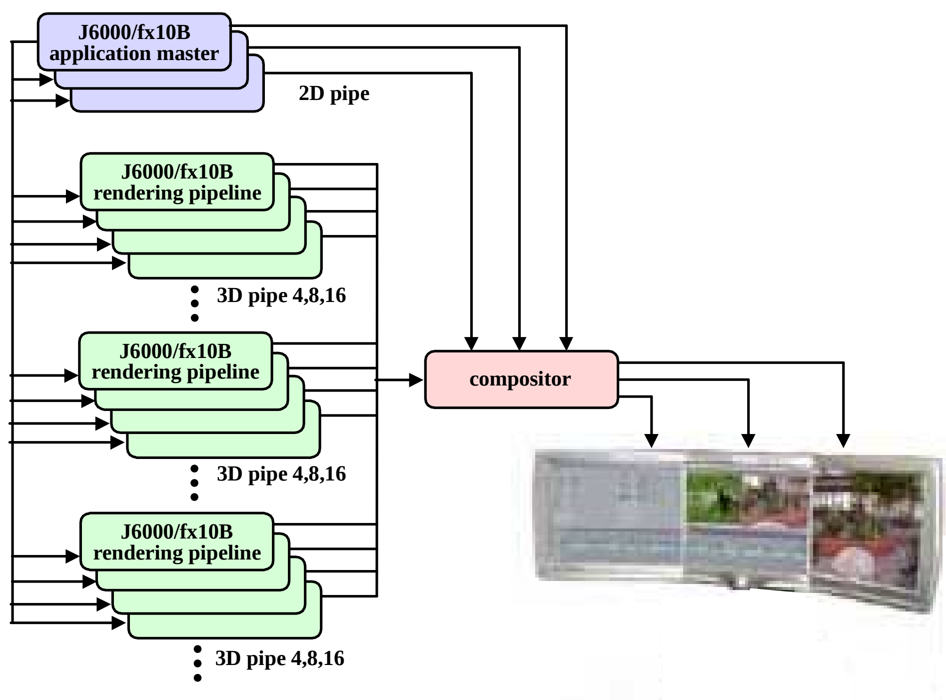


**Figure 4 - Compositing Operation in Mixed Accelerate/Scene AA Mode**

## Conclusion

The scalable visualization technology is being developed for those customers who require visualization capabilities beyond what is available from standard single pipeline solutions. It has the ability to scale at the same time in performance, image quality, and resolution as demanded by the customer. These scaling characteristics can be increased over time with additional pipelines or by upgrading of the individual components. Upgraded components can be re-deployed into less demanding compute/visualization uses thus protecting the customer's investment costs. The system utilizes standard components, which reduces overall system and support costs. By using standard components it also takes advantage of technology improvements in the areas that are changing the most rapidly. Applications can run in this new environment unmodified and take advantage of the scaling and quality capabilities. Finally when the pipeline systems are not in use for visualization they are available for use as computational resources.