COMPAQ
White Paper

September 1998
ECG098/0998

Prepared by Industry Standard Server Division

Compaq Computer Corporation

Contents

Virtual Interface (VI) Architecture - The New Open Standard for Distributed Messaging Within a Cluster

This paper will discuss the VI Architecture concept, its key components, its benefits, its relationship to the SAN, how Compaq envisions utilizing VI Architecture, and where VI Architecture is headed in the industry standards community.

Abstract: As business and technology requirements drive Intel-based servers to become clustered, a need for a high-speed, low latency communications service will be a requirement. This requirement for high-speed communication is further exacerbated as clusters grow from the current 2 node cluster to multi-node clusters which can approach 16 – 32 nodes in some currently proposed configurations. This service should not be encumbered by traditional multi-point protocols such as TCP/IP.

With these requirements in mind, Compaq has taken a leadership role in creating and defining the Virtual Interface (VI) Architecture. VI is a distributed messaging technology that has been authored by Compaq, Intel, and Microsoft. The specification, published December 16, 1997, provides an open architecture promoting clustered computing with a transport layer and Application Programming Interface (API) which can be utilized by software and hardware vendors to provide high-speed and low latency communications between servers in clustered server and storage configurations. This concept is called a System Area Network (SAN).

VI Architecture is designed to be hardware independent and compatible with current network interconnects such as ServerNet, Ethernet, and ATM. This ability to rely on existing communications adapters and media will help migrate existing protocols to VI Architecture. Since VI Architecture is also processor independent, it will allow the architecture to be implemented into silicon, therefore further improving communication performance.

Compaq will develop clusters of ProLiant Servers and/or NT Workstations to demonstrate its ability to perform in the high end of the enterprise market space. Compaq's initiative in developing and setting standards such as VI Architecture, in partnership with Intel and Microsoft, further strengthens its presence in the IT marketplace as a high-volume, low-cost supplier of quality x86 servers and workstations for enterprise level solutions.

Notice

Information from this paper was compiled from information found on websites and includes news releases, product summary sheets, product guides, and technical papers.

Virtual Interface (VI) Architecture – The New Open Standard for Distributed Messaging Within a Cluster

First Edition (September 1998)

Document Number ECG098/0998

Concept Overview

The Virtual Interface (VI) Architecture, authored and developed by Compaq, Intel, and Microsoft, is designed for hardware and software vendors and provides adapters, switches, middleware, and end-user application software that will seamlessly grow as servers or workstations are added and integrated into a clustered System Area Network (SAN). A SAN relies on high-speed, reliable messaging without the traditional communications protocol overhead as experienced by current LAN or WAN adapters and protocols, such as Ethernet, Token Ring, Fiber Channel Standard, Systems Network Architecture (SNA), or even Asynchronous Transfer Mode (ATM).

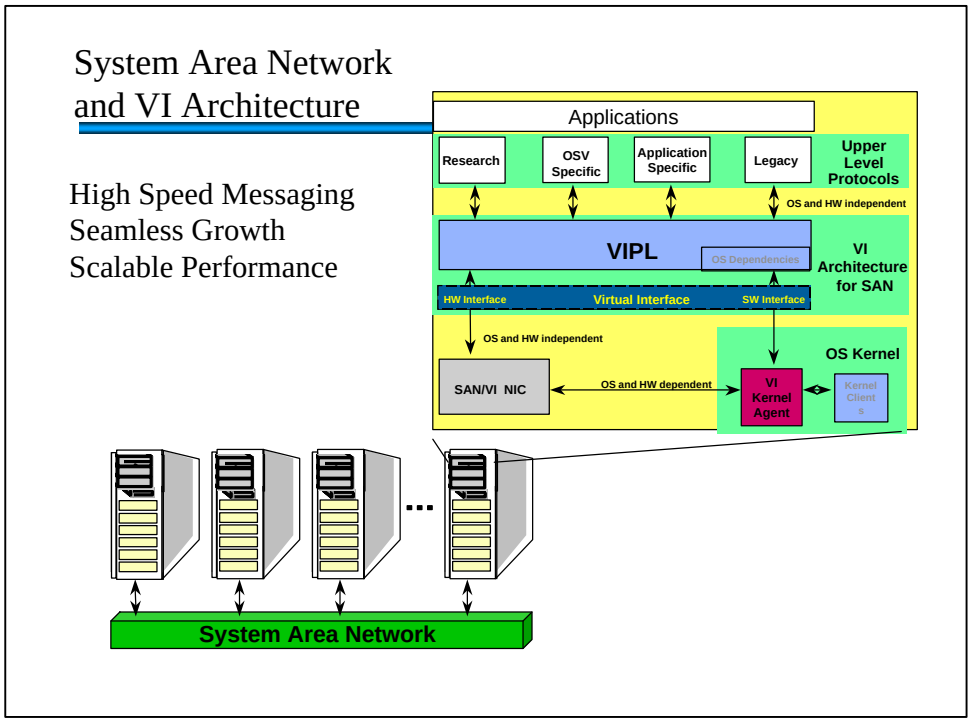
The VI Architecture is to be used where server-to-server messages deal with application and data availability, and it allows any server error or failure to cause an immediate transfer of the business critical application to another server for uninterrupted processing. The architecture also allows for parallel application processing and minimizes performance bottlenecks where more traditional communications protocols such as the RPC, UDP, and Sockets have caused application processing delays due to protocol stack handling and/or network traffic overloads due to collision detection processing. These slowdowns cause application delays or inefficient use of the cluster.

The VI Architecture specification has been reviewed by over 40 companies that have an interest in the specification and many have provided critiques and comments. These companies include hardware vendors such as Hitachi, HP, IBM, Pyramid Technologies, and Siemens-Nixdorf as well as software vendors such as Oracle, Informix, IBM Software Group, and SAP.

VI Architecture Implementation

Compaq has been working with Tandem Computers in developing a VI Architecture-based adapter by engineering VI Architecture specifications into the hardware adapter. This will provide the high-speed and low latency required ensuring superior performance in a cluster. Figure 1 illustrates how VI Architecture would be implemented within an industry standard-based SAN.

Figure 1. VI Architecture Structure within a clustered System Area Network (SAN)



The roots of VI Architecture are in server messaging without allowing the operating system to relinquish memory to an application if other applications or operating system services need memory. In other words, VI Architecture will reserve memory so that distributed calls from other clustered nodes will not have to wait for memory to be mapped.

One of the key benefits of VI Architecture is its ability to reduce or remove the protocol stack(s) for send/receive and read/write for applications, thus removing a large percentage of operating system and server processing overhead for functions such as LAN and WAN packetizing and depacketizing. The operating system will still continue to provide security and messaging setup for applications as well as less common distributed processes where kernal pass through and operating system services are used.

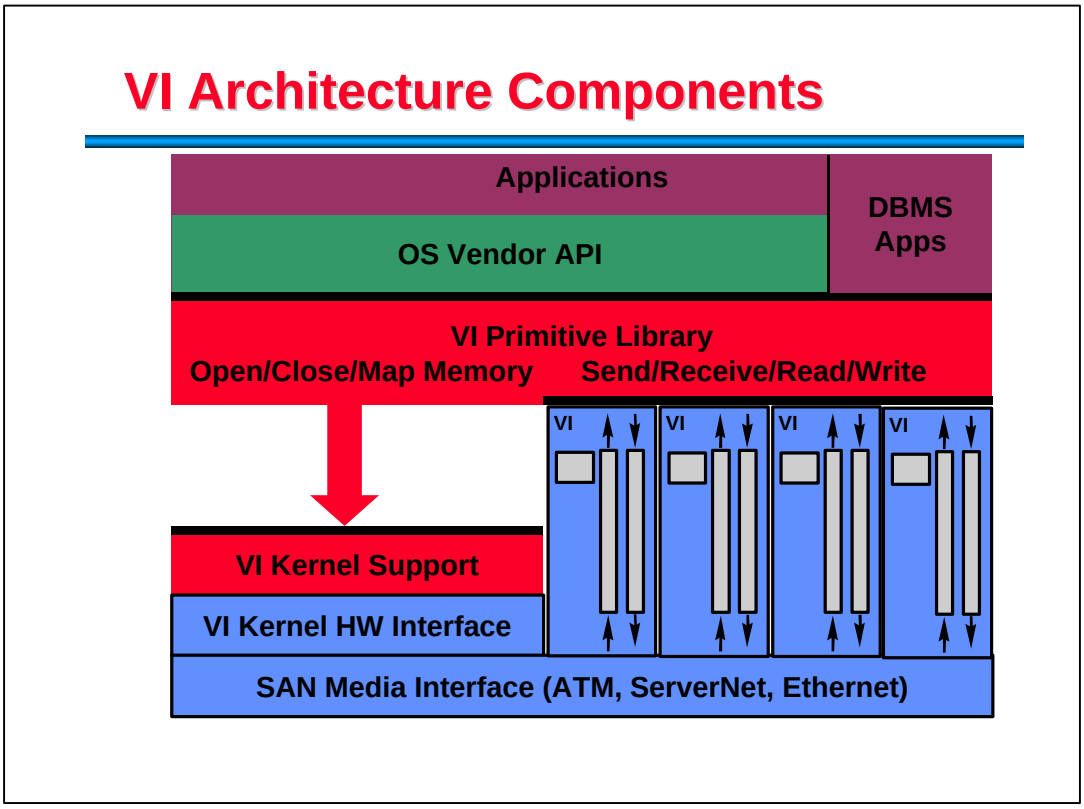
The VI Architecture provides a widely-supported set of commands to send reliable distributed messages between clustered servers, allowing applications to use reserved or pinned memory directly for message services between clustered servers. The VI Architecture also specifies a set of pre-defined queues and allows these queues to be resident on the VI Architecture compatible Cluster Interconnect Adapters such as future Tandem ServerNet adapters and OEM adapters that

will support the architecture. The architecture eliminates buffer copies and kernal overhead that have caused traditional networked applications to be performance bottlenecks in the past. This technology will further extend the concept of distributed and parallel computing.

Key Components

VI Architecture implementations will consist of three components: SAN media which will be equipped with VI Architecture registers and memory, the VI Architecture Primitive Library (VIPL) which will support application use of the architecture, and the VI Architecture operating system kernal support services. The relationships between these are shown in Figure 2.

Figure 2. Virtual Interface Architecture Components



Applications will use VI Architecture in the following manner:

- As an application is initiated, the application will initialize server or workstation memory segments through VI operating system kernal support. This action will establish an endpoint as well as initialize local VI memory registers and memory resident on the Network Interface Card (NIC) if it is VI-enabled. These queues are active as long as the application is active and provides a server or workstation memory segment that can be written to or read from. At this time, the application will register a portion of its virtual memory with the operating system kernal to be used to send/receive messages or data transfer.
- As the application starts to move data and act upon data either locally or through the VI-enabled NIC registers, the application will interface with the VI Primitive Library (VIPL)

which provides the interface between the application and the VI enabled NICs. It also supports and issues send/receive and read/write commands through the VI NICs as well.

Customer Benefits Utilizing VI Architecture

Customers using VI Architecture are assured of an open architecture promoting clustered computing and improving communication. Customer benefits include the following:

- **Lower cost per transaction**. The VI Architecture user will benefit in the reduced cost of transactions and application systems which will be achieved when VI becomes a broad industry standard. Vendors from Compaq, Tandem, IBM, Oracle, and Giganet are all producing hardware and software products that take advantage of the VI Architecture.

  With VI Architecture on existing servers and storage, customers will not have the need to increase server capacity as they have had to in the past, thus reducing cost. Also, with customers using traditional network interfaces for application corrections, applications will operate more efficiently in their use of buffer pools and networks, so overall communication costs will also be reduced.

- **Scalability**. Software vendors will be able to offer their applications without change to take advantage of VI Architecture as Microsoft integrates the function in their Winsock 2 interface, which is a direction they have stated. This results in immediate scalability for Winsock 2-enabled applications when VI Architecture suitable hardware is used in the systems interconnect. Thus, increased performance for existing servers and immediate cost benefit for the customer is provided.

Advantages of VI Applications

VI Architecture, as an emerging industry standard protocol for Distributed Message Passing and command processing, provides many technical advantages including:

High Degree of Application and High Availability

Applications can take advantage of the Distributed Messaging and Parallel processing of VI with very simple read/write and send/receive commands through a well understood transport layer like Winsock 2 for NT servers and workstations. These are also being extended for SCO UnixWare and Novell NetWare as well. This architecture can provide a high degree of application and data availability as well as increased application performance through high-speed systems interfaces at a read/write level for query and/or transaction processing in a Client/Server model. It can also enhance performance as parallel instances of the application run on two or more clustered nodes. Scaling can be even further enhanced by partitioning the application(s) into more efficient subprograms executing on different clustered nodes and passing dynamic information only as required to complete the transaction(s).

In simple test cases for message passing, dramatic results have been achieved over traditional TCP/IP and LAN media. In a database test case using IBM's UDBV5.0 with Extended Enterprise Edition for NT, it was observed that nearly linear scaling could be achieved in DSS operations across six ProLiant 6500 servers each supporting 70+GB of disk. This technology was voted Best of Show for Servers at the Hannover Technology Fair, April 1998.

More Effective Operations

As solution providers begin to use VI Architecture, larger clusters can be designed since intra-cluster control, through the VI enabled network, can allow for more data and hence more nodes to operate effectively than can be achieved with default TCP/IP or UDP communications.

As clusters grow from two node to multi-node, and applications become distributed for scalable performance, then inter-processor communication speeds become more critical. Clustered system management messages as well as data interchange between nodes can not be delayed by a higher form of communications interface and as application processing becomes dependent on high-speed distributed systems and interfaces, VI becomes a paradigm that solves a multitude of problems.

Lower Latency

VI has lower latency, in that kernal processing overhead is not required for message passing in a point-to-point switched environment. Applications read and write into and out of pinned memory regions not requiring a kernal process. This frees the processor(s) up to do application-related processing as opposed to overhead-related functions attributed to message passing. VI Architecture also minimizes machine cycles lost for interrupts dealing with message passing.

Conclusion

Compaq has taken a leadership role in creating and defining VI Architecture. VI is an industry initiative that provides nearly seamless growth for applications that can take advantage of its message passing and distributed processing features.

Compaq will develop clusters of ProLiant Servers and/or NT Workstations to demonstrate its ability to perform in the high end of the enterprise market space. Compaq's initiative in developing and setting standards such as VI Architecture, in partnership with Intel and Microsoft, further strengthens its presence in the IT marketplace as a high-volume, low-cost supplier of quality x86 servers and workstations for enterprise level solutions.

For More Information

To access additional Compaq white papers, please visit
http://compaq.com/support/techpubs/whitepapers

Related Compaq white papers include:

System Area Networks – A New Approach to Clustering

ServerNet – The High Bandwidth, Low Overhead Enterprise Cluster Interconnection

Compaq E2000 Platform Architecture: Meeting All Levels of Enterprise Computing Requirements with Standards-Based Solutions

Feel free to visit the Virtual Interface Architecture Web site at http://www.viarch.org