

# Oracle Net High-Speed Interconnect Support

*An Oracle Technical White Paper  
May 2005*

*Oracle10g*

# Oracle Net High-Speed Interconnect Support

Overview.....	3
Challenges.....	3
Infiniband Introduction.....	4
Oracle Net High-Speed Interconnect Support.....	5
Oracle Internal Performance Benchmarks	<b>Error! Bookmark not defined.</b>
Deployment and Configuration.....	7
Infiniband Hardware Information.....	9
Conclusion And Further Reading.....	11

# Oracle Net High-Speed Interconnect Support

## OVERVIEW

A new feature in Oracle10g, Oracle Net High-Speed Interconnect Support, improves network performance significantly by taking advantage of industry-standard high-speed networking technology such as InfiniBand. This document describes the performance bottleneck experienced in today's data centers, introduces InfiniBand Architecture as a solution and provides overview of few deployment scenarios. Finally, it identifies several InfiniBand hardware vendors & products available in the market today.

## CHALLENGES

Today's high-volume Internet applications in a typical data center environment require a rapid and reliable exchange of information between

- Application servers & backend database server
- And possibly between the nodes of an application/database server cluster, to synchronize operations and/or to share data.

The performance and scalability of these applications depend upon an efficient communication facility. Traditional network architectures do not provide the performance required by these applications, largely due to the host-processing overhead of kernel-based transport stacks and latency of the physical link.

Today's predominant LAN and Internet protocol (i.e., TCP/IP) sacrifices latency for reliability, generality, and distance. In a data center environment, as the middle tier software, the application servers process the business logic and communicate with the Web servers and with the database servers on the backend. The communication on the front-end is typically carried out today over TCP/IP connections, which puts a heavy burden on the CPU utilization of the application servers, as well as on back-end database servers. In some configurations, measurements have shown that processing the TCP/IP stack consumes 80% of the CPU cycles,<sup>1</sup> which does not leave much CPU capacity for the original purpose of application servers: to run applications, and the original purpose of database servers: to serve data.

---

<sup>1</sup> Source: TCP/IP at Near-Gigabit Speeds, Andrew G, Jeff C, and Ken Y, Duke University, December 1999

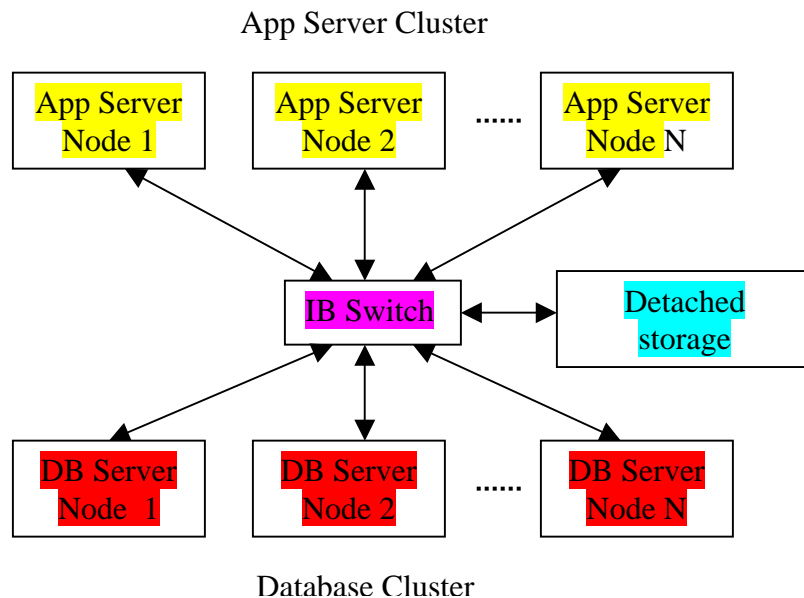
## INFINIBAND INTRODUCTION

InfiniBand Architecture (IBA) is an industry standard, channel-based, switched fabric, high speed interconnect architecture with low latency and high throughput.

A PCI-X InfiniBand host channel adapter (HCA) has two 10Gbps ports. These ports can support an aggregate throughput of 20Gbps, up to 20x bandwidth improvement over a typical Fast Ethernet card.

This large bandwidth and low latency allows sharing of same physical link for communication between the nodes of a cluster, communication to back end database servers and Network Area Storage (NAS). Thus, aggregating network pipe technology in the data center helps achieving in reduction of complexity and cost.

Figure 1: InfiniBand in Data Center



Sockets Direct Protocol (SDP) is an InfiniBand specific protocol defined by the Software Working Group of the InfiniBand Trade Association (IBTA). It defines a standard wire protocol over IBA fabric to support stream sockets (SOCK\_STREAM) networking over IBA. SDP utilizes various InfiniBand features (such as remote DMA (RDMA), memory windows, solicited events etc.) for high-performance zero-copy data transfers.

Most of the SDP/Infiniband vendors provide two mechanisms for taking advantage of SDP

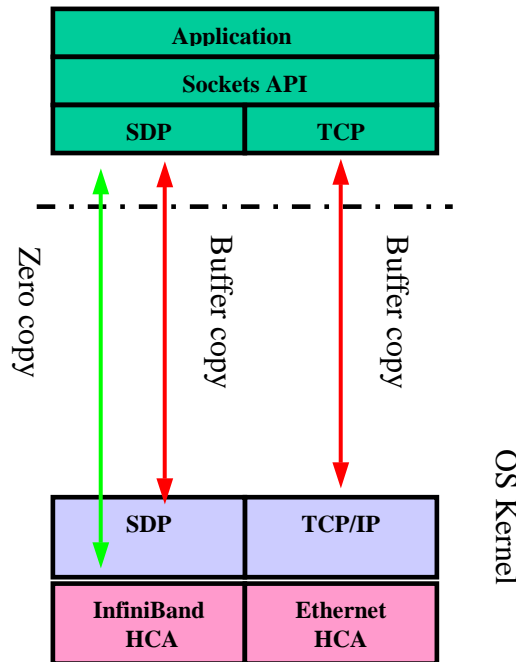
- Applications explicitly create SDP connection (henceforth will be called Native SDP)
- Convert TCP connection to SDP at run time without any change to application (henceforth will be called Transparent SDP)

Most of the SDP implementations support two I/O modes

- Buffer copy: user data gets copied from user space to kernel space.
- Zero Copy: eliminates buffer copy by sending user buffers directly to HCA

Please refer to <http://www.infinibandta.org/specs> for more information about the InfiniBand Architecture and SDP protocol.

Figure 2: SDP Implementation Illustration



### ORACLE NET HIGH-SPEED INTERCONNECT SUPPORT

The new functionality provided by Oracle Net High-Speed Interconnect Support can be divided into two categories: SDP protocol support and asynchronous I/O support.

The SDP protocol support enables the databases, application servers, and CPUs to operate more efficiently: the databases spends less time waiting for work, the application servers spend less time waiting for responses, and the CPUs have more cycles free for other work. As a result, customers see

improved application performance from existing servers (both database servers and application servers) with the addition of InfiniBand hardware, without the cost of upgrading to expensive hardware.

Asynchronous I/O support, another benefit of Oracle Net High-Speed Interconnect Support, allows Oracle Net send and receive buffers to be passed directly to the operating system kernel, eliminating CPU-intensive memory copying operations. This frees up CPU cycles in addition to the improvement brought by SDP support, particularly for applications with a large amount of network traffic.

***Availability & Supported Operating Systems***

For 10g Release 1, SDP protocol support is available only on the Linux platform. Asynchronous I/O support is available on Windows, and VMS platforms, using shared servers for TCP and IPC protocols. Asynchronous I/O support for SDP is currently undergoing testing, the table below summarizes the support for High-Speed Interconnect features on different Operating Systems.

Figure 3: Operating Systems supported

	Asynchronous I/O Support	SDP Support
RedHat AS2.1	Not supported for Any transport	Supported
RedHat EL3	Undergoing testing for SDP	Supported
Windows	Supported for TCP, IPC	Not supported
VMS	Supported for TCP, IPC	Not supported

***Support for Pre-Oracle 10 Clients***

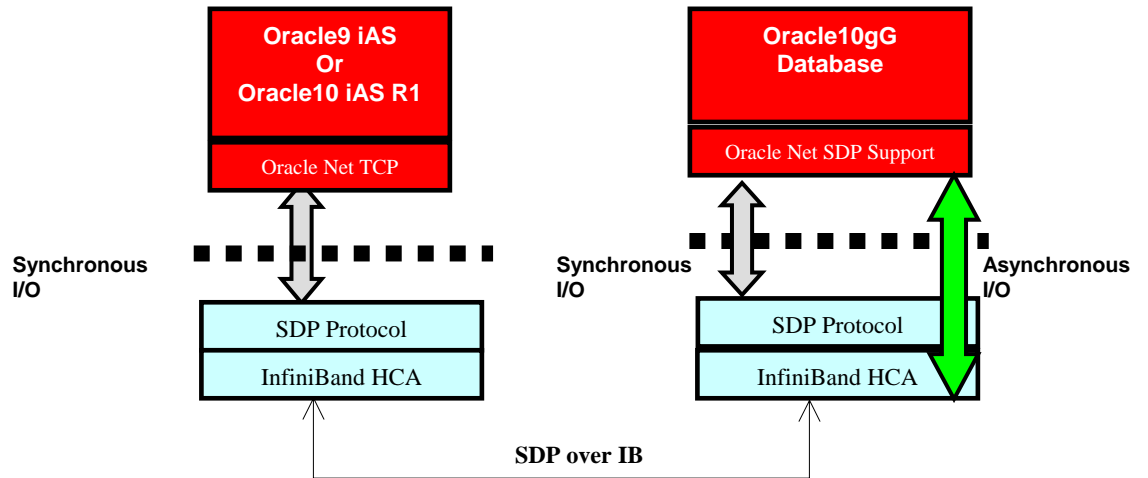
Pre-Oracle10 clients/Servers can use Transparent SDP method to take advantage of SDP.

**Note:** Oracle 10iAS R1 uses oracle 9i db client & hence can use only transparent SDP. However Oracle 10.0.2 iAS, to be available in summer of 2004, would have Oracle 10g db client. Also note that the thin JDBC drivers will have to use transparent SDP to take advantage of SDP.

Figure2 below illustrates a possible communication stack between Oracle9i/Oracle10iAS R1 Application Server and Oracle10g Database using this new feature. Oracle9iAS/Oracle10iAS R1 transparently uses the existing TCP protocol support. The SDP protocol provider efficiently converts TCP connection to SDP transparently. The Oracle10g Database uses the SDP

protocol support and uses Synchronous or Asynchronous I/O provided by Oracle Net High-Speed Interconnect Support. This enables Oracle9iAS/Oracle10iAS R1 to communicate to Oracle10g Database via SDP protocol over InfiniBand network to improve performance.

Figure 4: Oracle communication stack using SDP



## DEPLOYMENT AND CONFIGURATION

There are multiple interfaces to connect to an Oracle Server from a client and based on the method employed the SDP configuration will differ. Listed below are the different methods to connect to DB server

- Connectivity utilizing Oracle Call Interface drivers: supports both Native & Transparent SDP for 10G DB client & only Transparent SDP for pre 10 Oracle Clients
- Connectivity utilizing JDBC OCI driver: supports both Native & Transparent SDP for 10G DB client & only Transparent SDP for pre 10 Oracle Clients
- JDBC thin driver: can only use Transparent SDP connectivity

**Note:** To take advantage of Native SDP Support in 10G Release 1 of Oracle Database, specify "Protocol=SDP" in the connect descriptor. For more information refer to Oracle "Net Services Administrator's Guide for 10g Release 1", section "Configuring SDP Protocol Support for Infiniband Network Communication to the Database Server".

Figure 6 shows connections between Oracle9iAS/ Oracle10iAS R1 and Oracle10g Database, Figure7 shows JDBC OCI connections between Oracle 10.0.2iAS and Oracle 10g Database, and Figure 8 shows connections between Oracle10g Client-based middle tier and Oracle10g Database.

Figure 6: Interconnect between Oracle9iAS/ Oracle10iAS R1 and Oracle10g Database using transparent SDP on application server and Native SDP on database server.

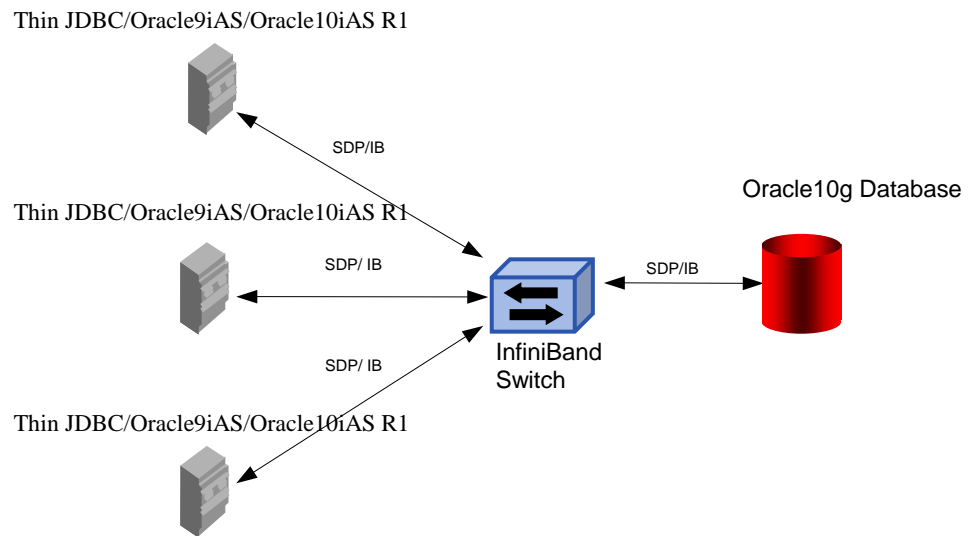


Figure 7: Interconnect between Oracle 10.0.2iAS and Oracle10g Database using JDBC OCI driver using Native SDP on both application server and database server.

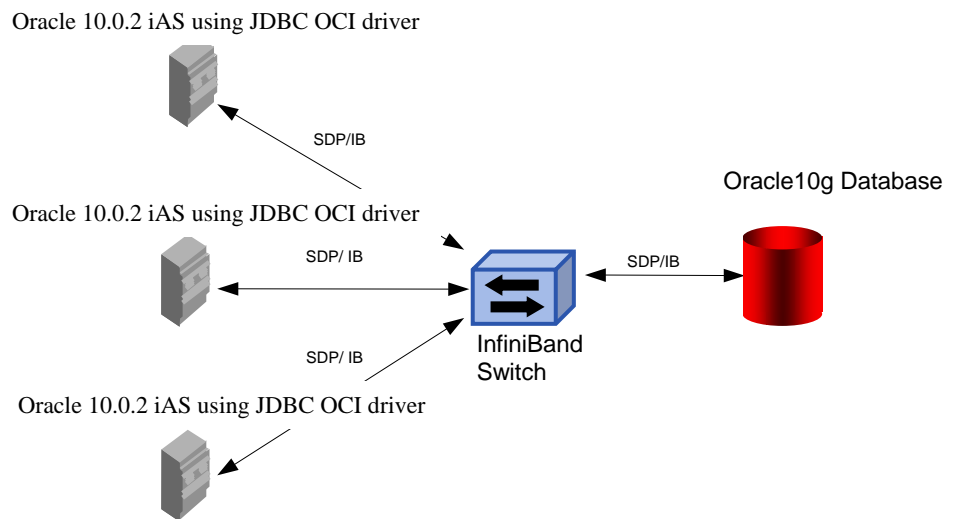
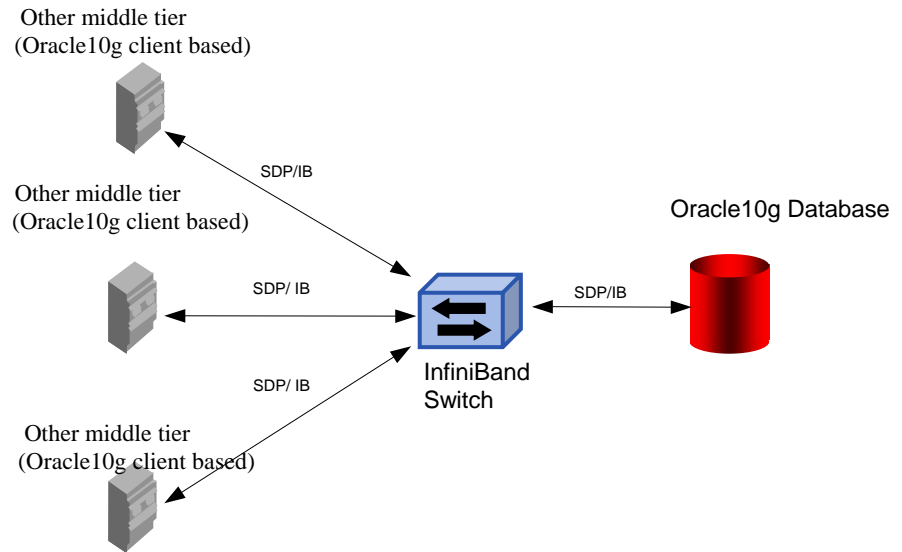




Figure 8: Interconnect between other middle tier and Oracle10g Database using Native SDP on both middle tier and database server.



### INFINIBAND HARDWARE INFORMATION

In addition to the installation of Oracle database software, customers need to install InfiniBand HCA's, and InfiniBand switch based on their respective deployment scenarios. Infiniband hardware vendors are expected to release product versions which support Oracle Net Services in the near future. Below is more information about InfiniBand hardware.

#### *Topspin InfiniBand Switch and Host Side Adapter Family*

Topspin provides a complete end-to-end InfiniBand server switching and I/O consolidation solution for boosting Oracle performance and cutting server I/O costs in both the application and database. This includes a family of InfiniBand switches, Fibre Channel to InfiniBand and Ethernet to InfiniBand gateways, InfiniBand server adapters and a suite of high performance drivers developed in conjunction with Oracle.

Topspin's switch family includes the Topspin 90 Switched Computing System, with twelve InfiniBand ports and the ability to add a Fibre Channel or Ethernet gateway for efficiently connecting storage and LAN into the network. The Topspin 360 provides up to 72 ports and the ability to add multiple Fibre Channel and Ethernet gateways in the same system.

See table-5 for Oracle Net performance benchmarks using Topspin SDP/Infiniband solution. More information about Topspin's product offering is available at <http://www.topspin.com/solutions/hca.html>.

Topspin InfiniBand products are available from every major system vendor, including Dell, IBM, HP, NEC and Sun, as well as from Topspin directly.

### ***InfiniCon InfiniBand Switch and Host Side Adapter Family***

InfiniCon Systems offers an full array of InfiniBand solutions, including Host Channel Adapters (HCAs) with a complete suite of high performance upper layer protocols for all major platforms, a wide array of InfiniBand switching solutions, Fibre Channel and Ethernet gateways, management software and tools, for designing and deploying 10 - 30 Gbps switching systems which enable outstanding Oracle performance.

InfiniCon's product portfolio includes the 12-port InfinIO 5000, a complete clustering and I/O virtualization system available for Oracle 9i and Oracle 10g database applications. The entry-level InfiniBand system available with integrated, dual Fibre Channel and Ethernet gateways, the InfinIO 5000 connects clusters seamlessly into storage and LAN networks, while supplying a low latency, high bandwidth infrastructure that yields high application and database performance with enterprise-class availability.

To streamline operations, the InfinIO 9000 family, provides from 24 up to 288 InfiniBand connections in a single switching system, all with integrated I/O capabilities.

For more information about InfiniCon, please visit <http://www.infinicon.com>

### ***Voltaire InfiniBand Switch and Host Side Adapter Family***

The Voltaire ISR 6000 is a modular chassis that supports 6-18 InfiniBand ports, and includes virtualized SAN and LAN connectivity. The Voltaire ISR6000 is 1U device that provides InfiniBand switching, FC SAN connectivity and connectivity to TCP/IP network. Voltaire's routers also include the unique capability of TCP/IP termination which reduces the networking overhead for application servers running Oracle Net.. More information about this product is available at <http://www.voltaire.com/isr6000.html>.

Voltaire also provides the HCA 400 and associated protocol suite to enable high performance Oracle implementations for both application server and database solutions. The HCA 400 is a dual port, 4x InfiniBand, low-profile adapter HCA that enables PCI, PCI-X and soon-to-be-available PCI Express based servers to access the full performance of high-speed InfiniBand fabrics. The Voltaire software stack allows both new and legacy applications to run transparently while benefiting from InfiniBand. More information about this product is available at <http://www.voltaire.com/hca400.html>.

It is expected that customers will be able to also get InfiniBand hardware along with their host machines directly from major system vendors in the future.

### **CONCLUSION AND FURTHER READING**

Oracle Net High-Speed Interconnect Support is well designed to support InfiniBand Architecture and other future high-speed networks. This feature increases the throughput and reduces CPU utilization for the communication between application server/other middle tier and database server. And therefore it directly addresses performance and scalability challenges facing data center today.

Readers who wish to obtain more details about requirements, configuration, and deployments are encouraged to consult the corresponding section of *Oracle10g Net Services Administration Guide*.



White Paper Title  
May 2005  
Author: Stella Kister

Oracle Corporation  
World Headquarters  
500 Oracle Parkway  
Redwood Shores, CA 94065  
U.S.A.

Worldwide Inquiries:  
Phone: +1.650.506.7000  
Fax: +1.650.506.7200  
[oracle.com](http://oracle.com)

Copyright © 2005, Oracle. All rights reserved.

This document is provided for information purposes only and the contents hereof are subject to change without notice.

This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle, JD Edwards, and PeopleSoft are registered trademarks of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.