

# Oracle Database Appliance



By Alex Gorbachev

**O**racle Database Appliance is the first, and so far the only, Oracle engineered system targeted on the lower segment of the database market. The innovation of Oracle Database Appliance (ODA) is in its architectural simplicity — it's as simple as you can imagine a two-node Oracle Real Application Cluster (RAC) could be. Thanks to its simplicity, Oracle has designed the database appliance to be an inexpensive and reliable all-in-one database platform. In this article, you will learn the architecture of this appliance, where it is useful (and where it is not) and the independent benchmarks I performed.

ODA is a single 4U device that contains all the required infrastructure for a RAC cluster, including two reasonably powerful database servers, shared storage and cluster interconnect. It comes as a fully integrated solution and takes minimal effort to have a database up and running on it. The only prerequisite for deployment is to allocate required IPs for ILOM (Integrated Lights Out Management) cards and public network (including VIPs and SCAN IPs if RAC is in use) and configure DNS to support SCAN if RAC is used. The configuration and installation takes only few screens and one to two hours for a full blown two node Oracle RAC cluster.

## Server Nodes

The Oracle Database Appliance chassis includes two redundant power supplies that power all the components of the system. This is more efficient compared to individual power supplies for each component. The chassis also contains all the required connectivity such as storage and the power backplane.

Each database server inside an appliance is represented by an independent mainboard with CPUs, memory and interfaces either built into the mainboard itself or as PCI Express expansion cards. All this (plus cooling and two hard disks used for the operating system) is packaged in a small chassis called a “system controller” by ODA creators or a “server node” for end users. You can think of them as server blades, because they are powered by the central power supplies and are simply plugged into the appliance connected through the backplane.

The fastest common x86 cores are found on the Intel Xeon X5600 series processors, at least until Intel Sandy Bridge CPUs become common in the enterprise servers. Note that I'm talking about performance of the core — this

is what you should be looking at since Oracle Database Enterprise Edition is licensed per core and not per socket. ODA uses the most powerful processor in the 95-watt TDP range: X5675 with six cores running at 3.06 GHz. The next CPU in this series would be X5680, which requires 130 watts, so it's more difficult to cool. All ODA customers should apply patch bundle 2.1.0.3.0, as it fixes an important BIOS limiting CPU performance. You can find more details in my blog at <http://bit.ly/ODA21030>.

With a total of 24 powerful cores, ODA has plenty of processing capacity for the majority of database applications I see customers running these days — even the poorly written applications that tend not to burn CPU cycles very effectively. These are, unfortunately, the most common applications out there. The 96 GB of RAM in each server of the ODA is a good match to the CPU capacity and should satisfy the demands of most medium scale database applications that I see in real life.

Each server node also has an embedded ILOM card that provides remote control of the server node and remote access to the server console — a manageability feature you would expect from any modern server. ILOM cards can also be integrated with Oracle Grid software, providing more reliable I/O fencing.

ODA has plenty of Ethernet ports for external connectivity. Other than ILOM network ports configured as part of you management network and two internal 1 Gbit Ethernet (GbE) for interconnect, there are two external 10 GbE ports and six 1 GbE ports configured as bonded in pairs by default. At least one of the bonds must be used as public network and others can be optionally configured for backup, external storage, DR and other uses.

## Cluster Interconnect

Oracle Database Appliance doesn't use InfiniBand to perform disk I/O or as a cluster interconnect. Instead, it relies on simple SCSI and Ethernet technologies. There's nothing bleeding edge here, but it's a proven rock-solid approach. Does the absence of InfiniBand reduce the performance of ODA? Not really.

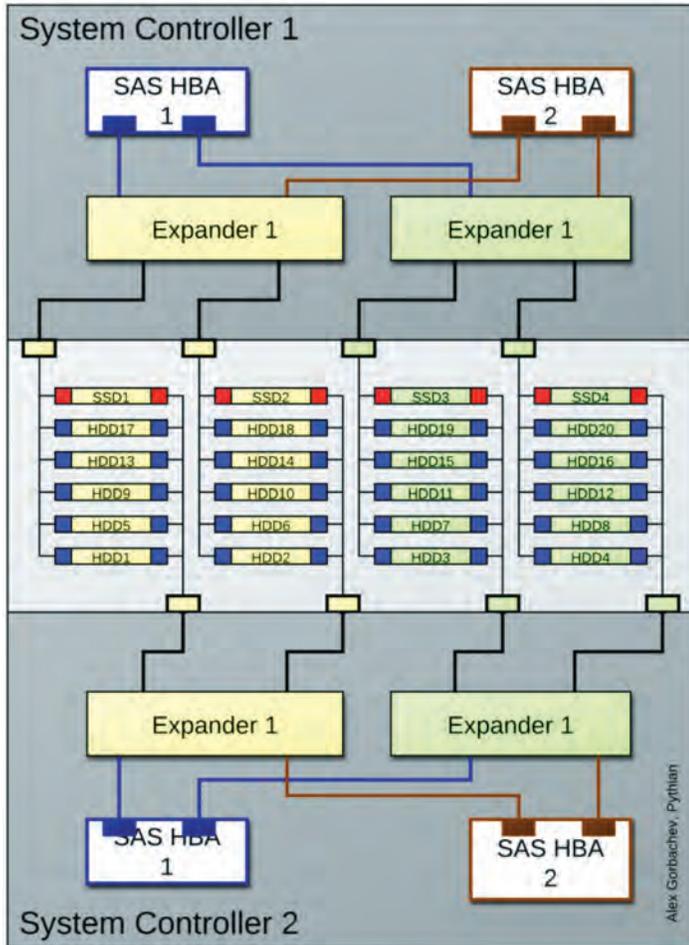
For the cluster interconnect, ODA uses two 1 Gbit Fiber Ethernet links. There is no network switch in between, and two fiber ports on one server node are connected directly to two ports on another server node. In my experience with Oracle RAC deployments, a two-node RAC cluster generally does not saturate a single gigabit link processing OLTP workloads. If the workload is not RAC optimized and causes a lot of cache fusion traffic, a bottleneck is usually hit somewhere else before interconnect is saturated. Tracking “gc %” wait events on ODA and Exadata, I see they have comparable timings. InfiniBand really shines at higher scale, where saturated Ethernet latencies usually climb up rapidly. However, this usually goes beyond a two-node cluster.

When it comes to data warehousing workloads running parallel operations with slaves on both nodes of the cluster, saturating both gigabit links is possible. However, I normally don't see much benefit to spreading parallel operations across just two nodes — there is more overhead of managing internode parallel processing than gains, especially when the bottleneck is not on the server but in the I/O subsystem. This is more likely the case for ODA, as we will see later. If you are processing parallel queries on the appliance, consider keeping each query on one node.

## Shared Storage

The storage subsystem of an Oracle Database Appliance resembles that of a standalone server with internal disks. Each server node has two Serial Attached SCSI (SAS) Host Bus Adapters (HBAs). These HBAs are connected to

24 disks via two SAS expanders. The difference from normal local storage is that each disk has two SAS ports, so each server node is connected to each disk via one of those ports. This is my favorite architectural feature of ODA — it solves the shared storage problem in such an elegantly simple way, as schematically shown in Figure 1.



**Figure 1: Storage Subsystem of ODA**

Since each of the two SAS HBAs has two paths to each disk — one via each expander — multipathing is configured on the servers. (Don't worry. It's taken care of by the installer.)

So, what are those 24 disks I mentioned earlier? Twenty disks are traditional spinning disks. They are 600 GB high-end 15K RPM SAS disks — the same disks as high-performance disks in Exadata. These disks are used for database storage as well as for the Flash Recovery Area (FRA).

Oracle Appliance Manager (OAM) can configure two predefined disk layouts: 40/60 when 40 percent is allocated for the database and 60 percent for FRA; and 80/20 with 80 percent allocated to the database and 20 percent for the FRA. Oracle Database Appliance is using HIGH REDUNDANCY ASM diskgroups, providing triple mirroring for its storage. Triple mirroring turns 12 TB of raw disk space into 4 TB of usable space (actually 3.8 TB when accounting for a single disk failure) split between FRA in the configured proportion.

Four other disks in ODA are SSDs from STEC (MLC version with SAS interface <http://bit.ly/ODA-SSD>). SSD disks are dedicated for only one purpose in the database appliance: the online redo logs.

I published an article in the February 2012 issue of NoCOUG Journal (publicly available at <http://bit.ly/NoCOUG12Feb>) where I discussed storage performance of ODA based on my benchmarks and described some of the reasons behind ODA design. I urge you to read that article for the complete story. Let me summarize the results here.

Thanks to SSD storage dedicated to redo logs, ODA is capable of processing thousands of redo log writes per second, enabling database instances to handle thousands of transactions per second. I have later performed stress tests of LGWR (the process responsible for writing to the redo logs) using a real database rather than an I/O benchmark tool and was actually able to reach several thousand transactions per second using a synthetic stress test (<http://bit.ly/ODASStorage1>). This is actually the level of high-end database applications, but real life database applications will most likely reach either CPU bottleneck or I/O bottleneck (not SSD but other 20 HDDs) of ODA before reaching that transaction rate since real-life transactions usually do more than just an insert of a single row. Another important advantage of SSD is that commit performance is very consistent whether a database processes 50 or 1,000 transactions per second. For small OLTP transactions (up to tens of kilobytes), database processes should be able to commit in less than 0.5 milliseconds.

Based on my benchmarking tests, 20 spinning disks can deliver up to 4,000 random IOPS per second if the entire disks are used for data (80/20 configuration) or up to 6,000 IOPS if only outer 40 percent of the disk is used for data (40/60 configuration) with average response time within 15 millisecond range. I was able to reach up to 2.4 GB per second bandwidth doing large sequential I/O from a single node.

### Database Software

ODA doesn't need special software and runs Oracle Database Enterprise Edition. It can be installed as a single node, but the full potential of the ODA is realized when customers use RAC or RAC One Node. Any Enterprise Edition options (e.g., partitioning) can be licensed on ODA just like on any other generic x86 platform. The same 0.5 multiplier applies to ODA licensing cost, so a single CPU license will actually cover two cores of an ODA.

One of the unique advantages of ODA is that customers can choose to enable only a subset of CPU cores available on each server and license only those enabled CPU cores. This is currently not possible on any other x86 platforms unless Oracle VM with static partitioning is used. This feature means that customers can start with a single CPU license Oracle Database Enterprise Edition and two enabled cores using one server node only and scale up to a total of 24 cores on both server nodes using Oracle RAC. Customers running older servers with dual-core CPUs will often face a significant license upgrade when upgrading to modern CPUs. The flexibility in core licensing in ODA can save significant sums of money during such upgrades. Note that once a certain number of cores are enabled, scaling back by disabling them to reduce licensing costs is not allowed. Think of the feature as a "one-way flexibility."

### "Secret Sauce"

Since ODA has no storage cells like in Exadata, there is no Exadata Storage Software. It means customers don't need to pay for it (currently listed at US\$ 10,000 per physical disk spindle), but they also won't get the unique Exadata Smart Scan and other Exadata-specific features. This is precisely why it's wrong to call ODA a mini-Exadata, as some analysts initially did.

ODA does have some unique "secret sauce" software coming with it, though: Oracle Appliance Manager (OAM) is a set of tools that simplifies installation and ongoing support of ODA by automating many common routine operations

*continued on page 6*

including configuration and deployment, patching, storage management and appliance diagnostics. The reason this package can function reliably on ODA is because of its fixed configuration. OAM developers can be confident that it always runs on the same hardware, is configured the same way and uses correct versions of software components and firmware. Because the appliance is configured in exactly the same way for thousands of customers, a significant level of repeatability can be achieved.

The predecessor of OAM is the OneCommand tool that was first used on Exadata. Since ODA is so much simpler and uses proven technology and architecture, you can expect OAM to be reliable from the very beginning. I can confirm that assertion having had firsthand experience with both Exadata and ODA.

## Applications of ODA

Since ODA is designed for simplicity, it comes with some limitations on its use:

- No Oracle VM or other virtualization technology can be deployed on ODA. Consolidation of databases is achieved by co-locating multiple databases on the same two nodes cluster or as schemas consolidation within a single database.
- While ODA can scale from two CPU cores to 24 CPU cores, it cannot scale the cluster beyond two nodes. The interconnect and shared storage inside ODA are not designed for expanding outside of the box.
- There is no opportunity for growth of the internal shared storage. I expect Oracle to expand internal storage capacity in the future as more and more customers ask for that.
- ODA comes with Oracle Enterprise Linux operating system only. It's not currently possible to install Windows or Solaris x86.
- There are no extra PCIe expansion slots available. All external connectivity must be done via either 1 GbE or 10 GbE network ports.
- Server nodes can only be used as database server, i.e., you cannot reuse one of the nodes as an application server. This limits usability of ODA for ISVs who want to bundle their solution on the single device unless they use only database like Oracle Application Express applications.

Ideal use for ODA would be for deployment of RAC or RAC One Node as a two nodes cluster. Yes, you can use only a single node in ODA, but that would be just an overpriced single server configuration unless you know for sure you need to scale up to two nodes in the near future or need high availability of RAC soon.

Even with all these limitations, Oracle Database Appliance fits a surprisingly significant number of Oracle database environments that customers run on other platforms today. Most of today's databases are below 3 TB in size, and an SGA size of 60 to 80 GB is a major step up for most deployments. The interconnect and storage performance are adequate for covering a vast amount of workload customers typically run these days.

## Other Considerations

I should warn that in order to follow the n+1 capacity planning rule, customers should only plan on loading one of the appliance's server nodes so that in case of a node failure, 100 percent of the workload can be processed on a single node. However, I see that in real life, customers running two-node clusters can compromise in a degraded mode with lower capacity when some of their processing is slower or can be delayed until full capacity recovery. Using multiple Cluster Database Services is a good option to control. Scaled down services are provided by an appliance in the event of a server node failure.

External disk storage for backups can be provisioned via 1 GbE or 10 GbE network. Customers can either use NAS (network attached storage) via iSCSI or NFS protocols. If iSCSI is used, customers must use ASM on top of iSCSI devices to share the storage between two nodes. If customers want to take backups further to tape, then it would need to go through the appliance servers again, which is not optimal. Using NFS mounted storage is the preferred option in my opinion, as the backups can easily be transferred to tape or replicated to another site using storage vendor replication. You can also install a backup agent on the server node directly and back up to the external tape library or other media. Oracle Secure Backup Cloud Edition will let you back up to Amazon S3 storage, for example.

Because customers cannot add any expansion cards and ODA doesn't have any Fiber Channel HBAs inside, it's not possible to connect a SAN storage with Fiber Channel interface. To utilize FC SAN, customers need to add intermediary storage server connected to FC SAN on one side and exporting NFS mount points to the database appliance on another. This is basically a NAS head or NAS gateway that many customers already have.

Right before this article went to print, Oracle has announced official support of external storage for ODA and this addresses the most common customer concern that I've heard so far. It's possible to store database files and not just backups on external NAS storage using NFS. Oracle suggests using Oracle ZFS Storage Appliance (ZFSSA) as this is fully tested configuration. ODA customers can also use Hybrid Columnar Compression (HCC) for data stored on ZFSSA. I posted some details on my blog — <http://bit.ly/ODANFS>.

ODA is fully compatible with Oracle DataGuard, and customers can even dedicate a separate set of bonded NICs for DataGuard replication traffic exclusively. Active DataGuard is a way to scale ODA deployments beyond two nodes if there is significant amount of read only reporting activities that can be directed to standby ODA(s).

## Summary

At only US\$ 50,000 (as of writing of this article) in hardware list prices, Oracle Database Appliance is a very competitive database platform suitable for small to medium database deployments. By using a traditional commodity servers, storage and network hardware, customers generally spend more to build a comparable configuration just on the hardware itself. Then they need to invest in an integration of all components and, in the end, get a unique configuration with its potentially unique issues. Replacing it all with a single inexpensive 4U database appliance running the same configuration as thousands of other customers around the world is very reassuring. Additionally, the Oracle Database Appliance provides a unique flexibility to license only a subset of available cores, enabling customers to achieve significant licensing savings upfront. If you are interested in more details, check the ODA article on the Pythian blog at <http://bit.ly/PythianODA>.

### ■ ■ ■ About the Author

**Alex Gorbachev** is a Pythian CTO, a respected figure in the Oracle world and a sought-after leader and speaker at Oracle conferences around the globe. He also regularly publishes articles on the Pythian blog. Gorbachev is a member of the Oak Table Network and holds an Oracle ACE Director title from Oracle Corporation. He is the founder of the Battle Against Any Guess (BAAG) movement promoting scientific troubleshooting techniques. More of Gorbachev's Oracle Database Appliance resources can be found at <http://www.pythian.com/oda>.