Technical Report

INCREASING THE MANAGEABILITY OF THE ORACLE GRID

Oracle Database 11g Release 2 and Real Application Clusters

IN AN EXPERT REVIEW of Oracle Database 11g Release 2 and Oracle Real Application Clusters (Oracle RAC), WinterCorp concluded that manageability of the database grid or cluster as a whole—and the databases within it—has been significantly increased. In two common cluster management tasks, WinterCorp observed substantial reductions in both the number of steps required of a skilled database or system administrator and in the elapsed time required. For an Oracle Database 11g Release 2 and RAC installation on a four-node cluster, a 40% reduction in the number steps required and a 34% reduction in the time required was measured, in comparison to the same operation using Oracle Database 11g Release 1. In this case, more than 100 steps were eliminated. For installing a software patch on a 4-node cluster, the reduction in steps and time was over 90%.

In the opinion of WinterCorp, these advances will result in significant savings in skilled labor for Oracle customers running clusters or grids—as well as lowering the risk that important tasks will be performed incorrectly. Since skilled labor is often the largest and fastest growing component of total cost of operation (TCO) for databases, WinterCorp believes that many customers will see a measurable benefit by adopting Oracle Database 11g Release 2 and RAC. In many cases, customers should be able to redeploy skilled DBAs and system administrators to tasks that are more rewarding and result in more positive business impact.

In addition to these two tasks for which direct measurements were made, Oracle Database 11g Release 2 and RAC related software contains several other major advances in manageability. These are also discussed in this paper.

METHODOLOGY

WinterCorp was retained by Oracle to provide an independent review of Oracle Database 11g Release 2 and RAC. WinterCorp observed demonstrations of each product feature; analyzed the steps required to perform each major operation; and, interviewed Oracle technical personnel about usage, limitations and other aspects of each feature. This report contains WinterCorp's description, analysis and conclusions concerning these major advances in this latest release.

Oracle has had an opportunity to request technical corrections to this report. However, WinterCorp retains final editorial control over this report and is solely responsible for its contents and conclusions.

WinterCorp
www.wintercorp.com

SPONSORED RESEARCH PROGRAM

ABOUT WINTERCORP

WinterCorp is an independent consulting firm that specializes in the performance and scalability of terabyte- and petabyte-scale data management systems throughout their lifecycle.

Since our inception in 1992, we have architected many of the world's largest and most challenging databases in production today. Our consulting services help organizations define business-critical database solutions, select their platforms, engineer their implementations, and manage their growth to optimize business value.

With decades of experience in large-scale database implementations and in-depth knowledge of database products, we deliver unmatched insight into the issues that impede performance and the technologies that enable success.

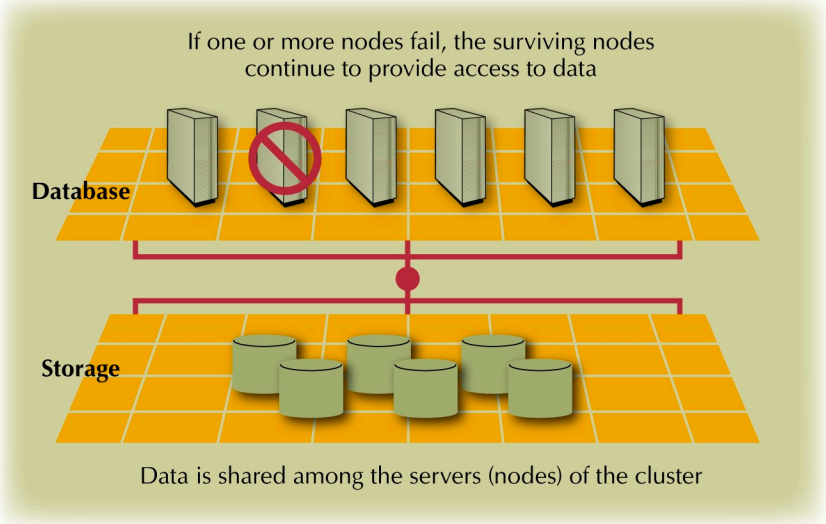
411 WAVERLEY OAKS ROAD, SUITE 328
WALTHAM, MA 02452
781-642-0300

visit us at
www.wintercorp.com

CLUSTERING

Oracle Real Application Clusters (RAC) was introduced by Oracle in 2001 to support the sharing of Oracle databases by multiple instances of Oracle, each of which can run on a separate node of a server cluster.

Figure 1: Oracle's Real Application Cluster



The fundamental architecture of Oracle RAC provides a database cache, which is shared by the multiple instances accessing a given database, using a feature named *cache fusion*. As a result of cache fusion, database changes made by one instance become visible to others as appropriate, providing the same database update semantics that Oracle provides within a single instance. In general, applications and Oracle queries which run on a single instance of Oracle will run on an Oracle RAC cluster without change.

Early versions of Oracle RAC were focused on increasing availability and scalability, with features such as: (a) node failover—if an Oracle instance on one node failed, instances on other nodes could continue to provide database services; and, (b) database scalability—an ability to increase or decrease database processing capacity by adding or removing nodes.

Adoption of Oracle RAC across Oracle's customer base has been widespread, with thousands of copies installed. One of the fundamental benefits of Oracle RAC is that it enables the deployment of large scale operations on configurations employing many small, inexpensive servers. This provides a hardware cost benefit to Oracle customers that is amplified because capacity can be purchased only as needed—a benefit often not present in large SMP deployments.

In addition, the built in failover capability of Oracle RAC ensures that customers can deploy robust database operations—even when using commodity components which are individually less reliable than high end enterprise hardware.

Thus, Oracle RAC enables a "scale out" strategy for users of Oracle database applications. Oracle RAC has enjoyed widespread use by Oracle customers for both transaction processing and data warehousing.

FROM CLUSTERS TO THE GRID

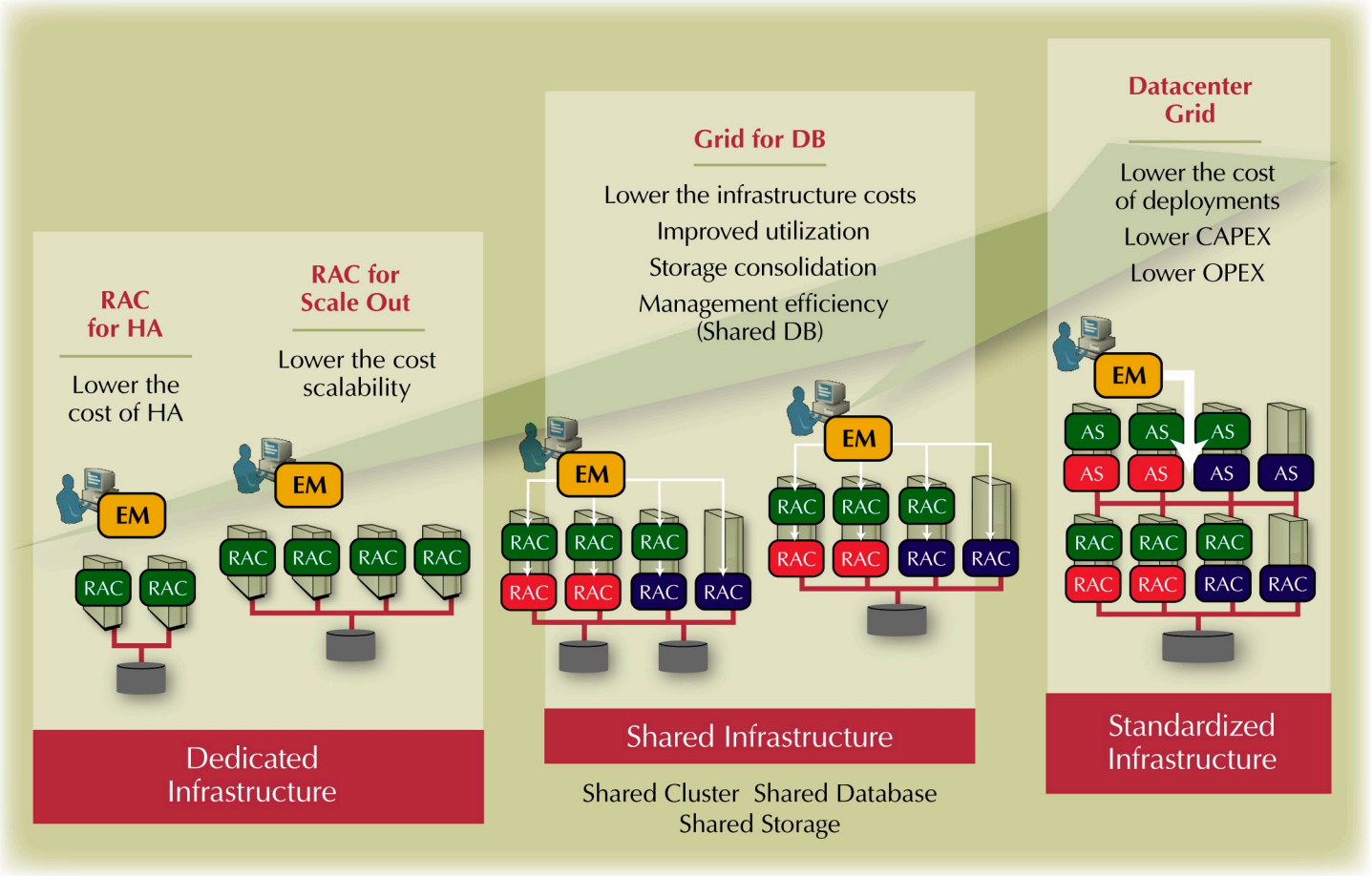
In the years since its introduction in 2001, Oracle RAC has been enhanced with features to increase its robustness and manageability. In addition, on behalf of users running larger scale Oracle operations—frequently consisting of many databases—Oracle has incorporated the concept of a database cluster into the broader idea of a "grid." A grid consists of a potentially large number of servers and storage systems working together to support multiple applications and databases. Within the grid, there are three tiers: the application tier, the database tier and the storage tier. Oracle RAC provides the software that enables multiple servers at the database tier to interact with multiple databases in the storage tier in virtually any combination.

The development of Oracle RAC, and of the grid concept within the Oracle product line, has facilitated the operation of clusters and grids on an increasingly large scale. Nevertheless, a cluster or grid environment—with potentially many servers running one or more Oracle instances (Oracle RAC is architected to support up to 100 nodes)—has entailed some inherent complexity. Provisioning new instances; maintaining software and databases; performance tuning and troubleshooting; and other activities—have all required database administrators (DBAs) with Oracle RAC-specific knowledge and skills.

SIMPLIFIED MANAGEMENT

As a consequence, Oracle has moved forward on a number of fronts—within Oracle Database; within Automatic Storage Management (ASM); and, within Oracle RAC—to simplify the management of clusters and grids using Oracle Database 11g Release 2. Much of this occurs via enhancements in Oracle RAC, which will be the central focus of this report. However, key related advances in ASM and in Oracle Database 11g itself are also briefly discussed here.

Figure 2—Evolution of Oracle RAC to GRID



ORACLE RAC ONE NODE

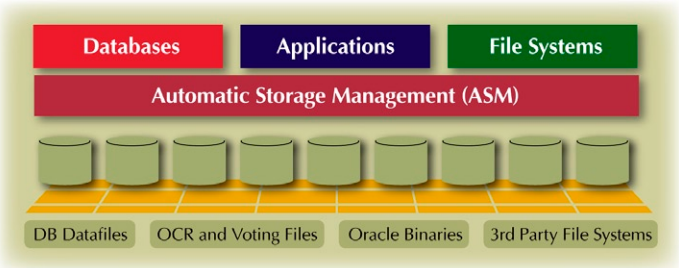
Oracle RAC One Node is a single instance version of Oracle RAC now available with Oracle Database 11g Release 2. Oracle RAC One Node provides failover and other services to single instance Oracle databases at a lower price point than full Oracle RAC. Oracle RAC One Node includes failover, which allows a second server to continue offering services on a database when the server or instance hosting that database fails. Oracle RAC One Node supports the live migration of running instances from one node to another, enabling customers to balance their workload across servers and take nodes down for service, etc. Oracle RAC One Node also supports rolling patches—a maintenance technique whereby database instances are online migrated to alternative nodes enabling database and operating system software to be patched or upgraded on the original node. Thus users can perform software and hardware maintenance without interrupting service to end-users. Oracle RAC One Node can also be online upgraded to full Oracle RAC, in case the operation expands to the point where it is desirable to deliver database services concurrently from multiple nodes of the cluster.

AUTOMATIC STORAGE MANAGEMENT (ASM)

Management at the cluster or grid level has also been simplified in Oracle Database 11g Release 2 by changes elsewhere in the Oracle infrastructure (e.g., outside of RAC)—for example, via the enhancement of ASM, Oracle's automatic storage manager. ASM was introduced in Oracle Database 10g to simplify the management of database files, providing such services as managing volumes and files systems without need for third-party software.

With Oracle Database 11g Release 2, ASM manages the entire Oracle storage tier, as shown in *Figure 3*

Figure 3: With Oracle Database 11g Release 2, ASM Manages all Data



—including standard storage architectures and Exadata storage systems—in a largely automatic fashion. This includes management of storage for the entire cluster or grid; storage volume management; cluster file management services; and, storage management for all Oracle databases. Through a single ASM interface, files and databases can be created for use on any database node or set of nodes; storage space is managed via single interface for the cluster; and, storage provisioning is handled at the cluster level. The inclusion of the cluster file system within the scope of ASM means that the files used for managing the cluster, including OCR and voting files, are also handled by ASM.

As in prior releases, ASM automatically handles striping, mirroring and rebalancing data over the set of available storage devices as configurations and workloads change. With many other storage management products, administrators must devote a substantial amount of time to balancing the workload over the storage devices as certain data sets become more frequently—or less frequently—used. ASM automatically rebalances data under its management in response to changes in the underlying storage devices.

In a new feature for Oracle Database 11g Release 2, data classified by an administrator as hot or cold is placed accordingly on different areas within in a given disk drive: data that is accessed more frequently (hot data) is placed on the outer tracks of the disk where access times are shorter; cold data is placed where access times are longer. In general, ASM will now more fully account for the I/O throughput requirements of data and manage its placement accordingly. In some scenarios, intelligent data placement will increase disk throughput by as much as 59% and reduce the number of spindles required by as much as 37%.

Another significant feature of ASM is that disk drives or LUNs can be added while the system is online. Once the drive has been added, ASM will automatically configure it and distribute data to it, to take advantage of both the extra storage capacity and the extra I/O capacity present in the drive. All this occurs while Oracle Database continues to run.

ORACLE RAC

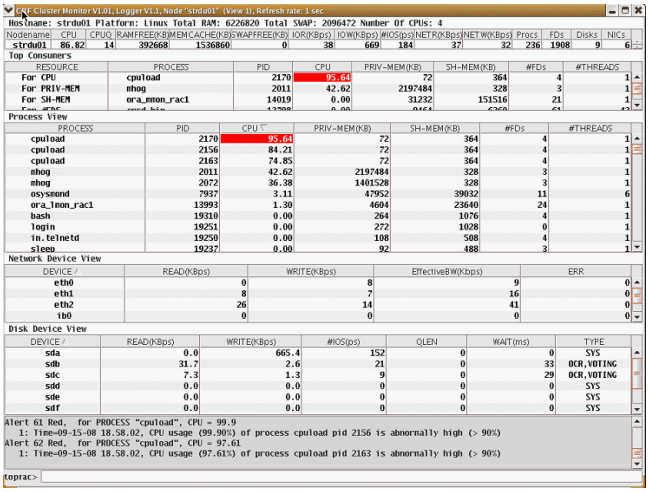
In addition to other features, Oracle RAC now offers substantial advantages in installation; adding or replacement of storage devices; adding, dropping or

replacing server nodes; performance management and tuning; and, software maintenance.

CLUSTER HEALTH MONITOR

The cluster health Monitor (CHM) is a new facility in Oracle Database 11g Release 2 that monitors resources across the cluster and helps the DBA identify, understand and correct problems. For example, if a rogue piece of software causes a node to hang, Oracle RAC or Oracle Clusterware will detect the absence of node "heartbeat" and remove the node from the cluster. The DBA, having been alerted to the development, will want to know why it happened.

Figure 4: Oracle Cluster Health Monitor Identifies Unusually High Processor Consumption



Source: Oracle Corporation

Using CHM, the DBA can replay the state of the cluster, starting from before the removal of the node, to see how the problem developed and take appropriate action. In effect, CHM provides the DBA with a cluster-level TIVO capability that can be used for root cause analysis. Plans call for CHM to become more proactive and automatic over time, thus further simplifying cluster management.

SIMPLIFIED INSTALLATION AND PROVISIONING

In Oracle Database 11g Release 2, the software of an entire grid can be installed in a single operation. The system checks automatically to determine whether software pre-requisites are in place and performs intelligent error handling and repair. For example, if key kernel parameters are not set, the installer sets them. If SSH is not set up properly, Oracle sets it up. Both the probability of an incorrect install—which appears to be okay but does not function properly—and the probability of an install which must be done over due to human error—are both greatly reduced.

WinterCorp observed an estimated 40% time savings in the installation of a 4-node cluster database. The install, which took 216 minutes with Oracle Database 11g Release 1 was reduced to 127 minutes with Oracle Database 11g Release 2. This was achieved mainly by reducing the number of steps (e.g., system commands) from 315 steps to 208 steps—a reduction of 34%.

More specifically, this install, after building the hardware cluster with Oracle Database 11g Release 1, consisted of the following major tasks:

1. Install OS (includes required packages)
2. Create raw partitions for CW files and set permissions
3. Set kernel parameters
4. Set shell limits
5. Create users and groups
6. Create directories
7. Set up SSH
8. Set up IP name resolution
9. Configure NTP
10. Run CLUVFY post HW/OS and pre-CRS install
11. Research and fix errors
12. Install CRS
    a. Specify directories, networks, nodes, OCR and VD locations
13. Run CLUVFY pre-DB install
14. Install DB
15. Configure ASM
16. Configure database (DBCA)

Notice that several of the major tasks above must be repeated by the DBA or Sysadmin for each node of the cluster.

The same result is achieved with Oracle Database 11g Release 2 via the following six major steps:

1. Install OS (includes required packages)
2. Set up IP name resolution (GNS)
3. Create Oracle user
4. Install Grid Infrastructure
5. Configure ASM
6. Install and configure database

Among the key areas improved in Oracle Database 11g Release 2 are:

- SSH setup script reduced setup time by 28 mins
- CVU fix-up scripts also partially reduced OS setup by 8 mins
- Combining the Clusterware with ASM under one home reduced time by 38 mins

While the labor savings (comparing the time for a flawless, successful install for the two versions of Oracle) will surely be valued by Oracle customers, WinterCorp believes that the automation and simplification here is even more significant.

Consider a DBA who is performing an install with an earlier release such as Oracle Database 10g Release 2. Should he or she improperly set permissions, or fail to set up SSH equivalence on all nodes, the install would continue and the problem may surface much later in the install process as some other type of error. These problems could take hours or days to diagnose, and then require wiping the installation and starting from scratch. In a large cluster, manually configuring all the nodes for installation was so difficult it could take several days to complete successfully. Oracle Database 11g Release 2 virtually eliminates such problems by automating these manual, time-consuming and error-prone tasks.

There are other advances in installation for Oracle RAC in Oracle Database 11g Release 2, including the availability of a complete *de-installer*. Previous versions of the de-installer left some elements of an Oracle installation behind, which under some circumstances would require additional effort before a reinstallation would be successful. Now that the de-installer is complete and available for all platforms, customers should find it easier to recover from an installation that goes awry.

GRID MANAGEMENT

Oracle Enterprise Manager (EM) has been extended to manage and monitor the entire grid, including all the components of Oracle Clusterware and all the databases and cluster file systems in the grid. Thus, instead of requiring command line instructions, DBAs can use EM's graphical interface to monitor and manage the grid. EM will allow the DBA monitor and manage all the clusterware (e.g., software components involved in operating the cluster, including the cluster file systems) and all the application resources. EM will automatically discover new targets that have been added to the cluster and allow the DBA to manage them.

In addition, Oracle Database 11g Release 2 includes a new ASM Configuration Assistant (ASMCA) for configuring both ASM and the cluster file system automatically.

ROLE SEPARATED MANAGEMENT

Oracle Database 11g Release 2 now separately support the roles of dba, sysadmin and storage admin—making it simpler for customers to train personnel for these roles and easier for individuals with these roles to work together to provision, maintain and operate clusters and grids.

For example there are new storage centric tools and privileges for the system administrator or the storage administrator, including an ASM configuration assistant that makes it easier to add, remove and configure storage devices.

There are new knowledge tools for the system administrator, including training courses and documentation, separate from the equivalent tools for the DBA. Thus the system administrator, who often is concerned with the operating system and the network rather than the particulars of the databases, can get more directly to the information he or she needs to accomplish a task or solve a problem.

Similarly, the installation process is now designed to respect the separation of DBA and system administration roles. Also, ASM now better supports role separation by enabling ASM tasks to be accomplished via the command line interface more familiar to system administration personnel. This is in contrast to earlier versions, in which ASM tasks were performed by means of SQL statements familiar primarily to DBAs.

GRID AUTOMATION

Oracle RAC also contains features to manage at the grid/cluster level, thus making cluster infrastructure invisible for many purposes.

For example, it is now much simpler to add and drop nodes. When a node is dropped or added, network configuration, name resolution and storage configuration are all accomplished automatically.

Cluster management is now substantially policy based. Services can be deployed to clusters, rather than to specific nodes. Further, application-required minimum resources are provided automatically, whenever possible.

Cluster patching is now automated. This dramatically reduces the work and the complexity involved in

applying a patch (e.g., a update to the database software that occurs between releases) to the Oracle infrastructure software in the cluster. WinterCorp witnessed a demonstration of this in which the number of steps required to apply a patch to the clusterware and databases in a 4-node cluster was reduced from 46 steps in Database 11g Release 1 to 4 steps in Oracle Database 11g Release 2—a reduction of 90%.

As with the simplification of installation, WinterCorp believes that Oracle customers will value the time savings significantly, but are likely to find the reduced complexity and potential for error to be even more significant.

PERFORMANCE TUNING

In Oracle Database 10g, performance tuning was enhanced via the introduction of Automatic Database Diagnostic Monitoring (ADDM) and the Automatic Workload Repository (AWR). These tools made it much easier for Oracle DBAs to identify, prioritize, analyze and correct database performance problems.

In Oracle Database 11g, ADDM support has been extended to the cluster and is now called *Global ADDM*. Global ADDM automatically identifies the most "globally significant" performance issues (e.g., those SQL statements that are consuming the most resources by virtue of the a combination of sub-optimal performance and frequency of execution). By default, Global ADDM runs hourly and identifies not only problem SQL but also global cache interconnect issues, lock manager congestion issues, global resource contention issues (e.g., I/O bandwidth problems) and skew in instance response times.

Global ADDM will not only identify such issues for the DBA, it will in many cases suggest corrective actions. For example, problem SQL will in some cases be corrected by the automatic creation of better plans. In other cases, the system will suggest improving performance via addition of index or updating of statistics. Much of this system generated tuning requires no change to the source SQL query, which is a great advantage when running packaged applications that the customer has limited or no ability to change.

Global ADDM can save a DBA a lot of time diagnosing and fixing performance issues. Imagine a 16-node cluster, with one node hosting a run-away query that is consuming a large amount of resources. That query could impact the entire system by consuming storage bandwidth and delaying inter-cluster messaging. With releases prior to Oracle Database 11g R2, the DBA would have to drill down into each node in the cluster hoping each time to identify the misbehaving query, often needing to examine all 16 nodes. With Global ADDM, he or she can view a global report that compares the performance of all the nodes and identifies those performing out of line with others. The DBA can then drill down with Global ADDM to identify the query at the root of the issue, whereupon Global ADDM may suggest a fix. This single-click diagnosis can result in a major time savings for a DBA, especially for a DBA managing a cluster.

CONCLUSIONS

Enhancements in Oracle Database 11g Release 2 and related advances in Oracle RAC and ASM will provide customers with a major simplification in the management of Oracle clusters and grids. Software installation and patching are much automated and simplified, with built in system error checking and correction. Grid management and monitoring are enhanced, with the result that many day-to-day functions can now be performed once at the grid level, rather than being repeated for each node of the grid. Grid performance and tuning is substantially simplified and automated with Global ADDM. And separation of the DBA role and the System Administrator role is now supported, with the result that operational responsibilities in larger installations will now be more readily organized, staffed and managed.

The magnitude of these changes, where they have been demonstrated to WinterCorp and measured, are substantial. Time to install Oracle infrastructure software on a 4-node cluster was reduced by over 40%, with a correspondingly large reduction in the number of steps to be performed and the complexity of the operation. Similarly, time to install a patch on a 4-node cluster was reduced by 90%. WinterCorp expects these savings to be even larger on larger clusters, where the reduction in complexity will be yet more dramatic.

Overall, the advances in manageability in Oracle Database 11g Release 2 and RAC are substantial and should help customers redeploy skilled DBA staff to more satisfying tasks with more positive business impact.