Backing up the Big Data Stack

ZFS-based Backup for Oracle Exadata Database Machine

By John Webster

May, 2013
The Stack Attraction

Shortly after the first round of converged systems was announced, we conducted a series of brief interviews of our user clients to determine how these announcements were being received. From these interviews we determined that:

1. Computational “stacks” i.e. server, networking, and storage components that were simply pre-integrated by a single vendor were not of significant interest. Users believed that their IT departments already had the required expertise on hand and that it was not worth paying the uplift in price for these pre-integrated appliances.
2. The vendor of the server within the “stack” was a significant consideration. It couldn’t be just anyone’s server. The vendor had to be known and approved.
3. Converged systems that were pre-integrated and optimized for a particular purpose or class of applications were of greater interest than those that were not. In fact, it was mentioned to us during these interviews that appliances built around Oracle database-supported applications would definitely be worth reviewing.

Since that time Oracle has been delivering a number of converged systems that are optimized for Oracle applications. These include the Exadata, Exalogic, and SuperCluster. Of these by far the most popular has been the Oracle Exadata Database Machine with more than 5,000 total units shipped to date. This becomes abundantly apparent as one looks at the growing scope and size of the Exadata platform. Exadata is now being deployed for an expanding list of applications beyond the “traditional” database workloads—OLTP and data warehousing—to Big Data and in-memory data analytics supported by a scale-out server/storage architecture. PB-scale implementations are now common as Oracle adds Big Data analytics applications to the list.

Big Data Protection in Converged Systems

While the number and variety of converged infrastructure/pre-integrated stacks is increasing, Evaluator Group notes that integrating the supporting operational and data management functions can still be left as a significant challenge to the user. One of these challenging but nevertheless essential operational practices is data protection. Accelerating data growth on all fronts is pressuring storage administrators to increase storage efficiency in order to ensure that all the data that requires protection can in fact be included under the data protection umbrella. Adding converged infrastructures can in turn add complexity to data protection operations.

Some of the growth in popularity of using the Exadata platform for Big Data analytics applications can be explained by disappointment in Hadoop among a percentage of enterprise IT administrators. We note that at least half of the Hadoop projects implemented to date by enterprise data center IT
administrators have failed to reach full production status. One reason is that Hadoop is perceived to be flawed with regard to its data protection mechanisms. Hadoop's underlying file system HDFS makes three copies of data by default. These can be used to recover from a number of failure scenarios within the Hadoop cluster. However, these are three full copies of data meaning that every TB of data ingested by the cluster results in three TBs stored. Making matters worse for these administrators is the fact that data corruption introduced by machine or human error can be propagated across the copies. And because the Hadoop open source developers community has yet to add snapshot capability to HDFS, there is currently no way to roll the Apache Hadoop cluster back to a known good state when this occurs.

In contrast, we note that established data warehousing and analytics platforms—Oracle’s included—have mature data protection environments supporting them. These environments include:

- Software to automate data compression, deduplication, and data replication to make protected and coherent copies of data in the Exadata database
- Infrastructure to enable the movement of backup data from Exadata to an external data protection storage device
- Storage devices to maintain the protected copies of information—in this case Oracle’s ZFS Backup Appliance

These capabilities make platforms like Exadata more efficient and secure repositories of large volumes of data than Hadoop. As a result, it has increasingly become the case that Apache Hadoop is instantiated as a highly scalable and versatile Extract/Transform/Load (ETL) staging area for data that is subsequently fed into the new generation of scale-out BI systems.

As a result of the convergence of transactional business systems with analytics, IT administrators will be managing critical business systems that are growing to PB scale and making the more or less normal compounded annual storage growth rates of 50-80% seem mild by comparison. Convergence is driven by many factors including:

**Business expansion and the demand from top-level management to create more business opportunities.** Previously, C-level management was more heavily focused on business preservation and risk management. Eighty percent of the IT budget was devoted to maintaining existing applications. The pendulum is now swinging in the direction of new application development and the use of data analytics to gain competitive advantage.

**The need to leverage additional data types.** These more voluminous types include unstructured data, rich media, and machine-to-machine data (RFID, GPS, etc.) that feed new applications and analytics systems.
The influx of data spawned outside of the data center. A huge wave of data generated by the web, social media, and mobile computing devices is now being captured inside the data center.

More information retained for longer periods of time. This need is driven by the traditional sources of data retention, including audit and compliance, as well as new application users needing to analyze change over time within their areas of focus (markets, populations, scientific research, etc.)

Given the accelerating demand for storage capacity we are seeing across industry segments, we believe that storage administrators should review existing data protection practices to determine whether or not they are sustainable. Inadequate backup processes create increased exposure to data loss and make recovery more complex and time consuming. If inadequacies are not effectively addressed in a timely manner:

- The growing amount of data included the backup process places increased pressure on storage managers who may already be struggling to ensure that enterprise data is adequately protected.
- The capital budget devoted to data protection that already accounts for half of many enterprise storage budgets will continue to grow.
- The amount of time required to protect the information may extend beyond what is practical from a business operations standpoint.
- The amount of data to protect may require more backup systems than can be physically accommodated.
- New IT projects could be impacted, as administrators realize that they are competing for already-strained data protection and budget resources.
- Capacity-based licensing thresholds can be exceeded—both on and off site—and turn into unpleasant budget busters.

Here we evaluate one way to address all of these issues for an Oracle Exadata environment.

Oracle’s ZFS Backup Appliance

The Oracle ZFS Backup Appliance (ZBA) is a high performance, purpose-built solution for backup and recovery that also is optimized for Oracle Engineered Systems, including Exadata. It can be managed by DBAs, storage administrators and IT data center administrators using Oracle Enterprise Manager and can be integrated with Oracle Recovery Manager (RMAN) automated backup and recovery utility that is built into Oracle databases. ZBA also can be integrated with the StorageTek Modular Tape Library for long-term archiving via optional FC connectivity.
Accessibility

ZBA falls into the category of Network Attached Storage (NAS) that is accessed via the NFS protocol. For integration with Oracle 11g, backup/recovery performance can be enhanced by enabling Oracle Direct NFS (dNFS). Access via NDMP is also supported for attachment to non-Exadata systems.

Performance

For best backup/restore performance, InfiniBand Quad Data Rate connectivity between Exadata and ZBA is provided, as is 10GigE. Performance also is accelerated by the implementation 64 CPU cores and 512GB DRAM per system as primary cache and can be further tuned on an ongoing basis to specific environments using the DTrace Analytics management tool (see below).

Oracle claims throughput rates of 20 TB/hr. using dNFS over InfiniBand for backup and 9.4 TB/hr. for restore—significantly greater than reported by competing data protection appliances in this category that includes those from EMC (Data Domain, Avamar) and NetApp. Concurrent backup and restore operations are supported so that restores can be performed when required without interrupting a running backup operation. Lastly, ZBA provides multiprotocol access to snapshots and clones (see page 6) for secondary database processing such as test, development or quality assurance allowing database backups to be used more for value-added work.
Advanced Functions

As its name implies, the ZFS Backup Appliance leverages the features and functions of the Oracle ZFS file system that forms its software foundation. Features and data management services built into ZFS and supported by the ZBA include:

Data Reduction Options

ZBA supports both data compression and inline deduplication functions that in some cases can result in a dramatic reduction in disk capacity consumed by backups depending on the type of data. ZFS deduplication is performed in-line and at the block level. It can be most effective for data reduction of VM boot images where duplicate blocks predominate. However, in-line deduplication is not recommended for live transactional workloads or heavy throughput workloads. In these cases, data compression also is supported. Each block is compressed independently using four different compression options:

- LZJB – for low CPU overhead compression
- GZIP-2 – a lightweight version of the gzip algorithm
- GZIP – offers higher compression but at some compute overhead penalty
- GZIP-9 – offers greatest compression of the four options, but can consume significant compute resources

As an example of its unique integration with Oracle Database, ZBA also supports the backup of in-database and Exadata-resident archives that have been compressed by up to 50x vs. the original by using Oracle Hybrid Columnar Compression (HCC). This particular capability is not supported in any form on other data protection appliances or other NAS platforms in this category. HCC technology is included at no extra cost and can help customers to:

- Compress historical data by 90 to 98 percent
- Increase application performance when querying historical data
- Dramatically reduce the cost of storing and managing Oracle Database data
- Use clone and snapshot copies of HCC-compressed data for test, development and quality assurance

This reduction in the storage capacity required for RMAN backups can equate to savings in the initial purchase of data protection storage and provide CAPEX and OPEX budget relief in the near term while postponing the future purchase of additional capacity to accommodate growth. In addition, the RMAN compressed data backup can target either disk or tape media or both as required for an added level of protection.
Clones and Snapshots

ZFS clones provide multiple duplicates of databases for application development, test, and performance troubleshooting. Cloning is non-disruptive to production applications and databases and can leverage RMAN backups already on ZBA to facilitate a storage-efficient test and development environment.

ZFS snapshots provide a read-only, point-in-time copy of the file system and the number of snapshots supported is practically unlimited. Snapshots are writable and only changes to the original are tracked to minimize capacity consumption and recovery time when using a snapshot to restore data. Snapshots can also be used to quickly export data copies to development and test environments. A Snapshot Manager supports scripted automation of snapshot creation and provisioning of snapshots for test/development and other applications. Oracle Database administrators can directly manage snapshots, clones and restores on the ZBA using the Oracle Snap Management Utility for Oracle Database UI.

DTrace Analytics

For real-time visualization of performance-related metrics and to troubleshoot and resolve bottlenecks, ZBA features a built-in application and workload analysis tool called DTrace. DTrace measures ZBA CPU, cache, protocol, disk, memory, networking, and system-related data, and presents the results graphically and in real time. Administrators have mouse-click drill down access to areas of concern from application to storage at the I/O and VM level. DTrace also supports multiple simultaneous application and workload analysis in real-time to help compare various aspects of system stress. Analysis can be saved, replayed and exported to other administrative tools for further analysis. Administrators can also remotely manage their ZBA environments from their Apple iPhones or iPads with a secure application that enables monitoring of storage service logs, I/O statistics, real-time analytics, component status, faults and recommended repairs.

Additional Advanced Features

- Triple parity RAID
- End-to-end checksumming
- Concurrent backup and restore operations
- Support for secondary database processing
- Predictive self-healing
- Diagnostics linked to Oracle’s on-line knowledgebase
- Remote data replicas and clone copies (requires additional software license)
Available Configurations

Oracle offers two fixed configurations of ZBA: a high performance option with high-performance SAS disks for throughput and 55TB of raw capacity; and a high capacity option with SATA disk for 132TB raw capacity and reduced throughput vs. the high performance option. Both configurations are scalable to 2.5 PBs of raw capacity. Oracle also provides an installation script to reduce set up time and the possibility of making errors during the setup process, as well as a complete installation service through its Advanced Customer Support group.

Conclusion

ZBA is a single-vendor solution for Exadata that can be deployed more quickly and at less expense than going through the administrative process of integrating a backup platform from another vendor. This eliminates the possibility of cross-vendor finger pointing when problems arise. No other elements, such as a media server or 3rd party backup software, or staff are required either. Management efficiency is optimized by using a single point, browser-based UI. And in addition to the optimization that Oracle has already done, performance can be further tuned on an ongoing basis to a specific environment with the aid of DTrace.

Finally, we note that accelerated backups of Oracle databases can result in reduced interruption of operations and the ability to protect increasing amounts of data over time. This has the effect of maintaining or even improving Recovery Point Objectives in spite of increasing data growth rates. The ability to quickly restore Exadata-resident Oracle databases allows administrators to show improved Recovery Time Objectives for critical transaction-based applications, such as SAP, resulting in minimized operation impact when application recovery is required. We believe that the accelerated performance offered by the ZBA in backing up Exadata will result in increased IT productivity and a more efficient data protection environment that will support escalating influx of data.

About Evaluator Group

Evaluator Group Inc. is dedicated to helping IT professionals and vendors create and implement strategies that make the most of the value of their storage and digital information. Evaluator Group services deliver in-depth, unbiased analysis on storage architectures, infrastructures and management for IT professionals. Since 1997 Evaluator Group has provided services for thousands of end users and vendor professionals through product and market evaluations, competitive analysis and education. www.evaluatorgroup.com Follow us on Twitter @evaluator group

Copyright 2013 Evaluator Group, Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or stored in a database or retrieval system for any purpose without the express written consent of Evaluator Group Inc. The information contained in this document is subject to change without notice. Evaluator Group assumes no responsibility for errors or omissions. Evaluator Group makes no expressed or implied warranties in this document relating to the use or operation of the products described herein. In no event shall Evaluator Group be liable for any indirect, special, consequential or incidental damages arising out of or associated with any aspect of this publication, even if advised of the possibility of such damages. The Evaluator Series is a trademark of Evaluator Group, Inc. All other trademarks are the property of their respective companies.