Technical Report

NetApp Storage Optimization with Clustered Data ONTAP 8.3 for Synopsys SiliconSmart

Standard and Custom Cell Characterization Tool

Bikash Roy Choudhury, NetApp

Sachin Bhatia, Synopsys

Ragini Suram, Synopsys

March, 2015 | TR-4390

Abstract

Standard cell library characterization is an important component in the chip design and manufacturing workflow. Standard cell libraries are developed and distributed either by a library vendor who uses the process and device characteristics and electrical information provided by the foundries or by the foundries themselves. Large fabless semiconductor companies may choose to design and characterize their own libraries for business and competitive reasons as well. The characterization step provides timing, power, and noise behavior across a wide range of voltage, temperature, and process variations that make up the electrical behavior and specifications of the device that is being designed.

The ever-increasing demand for low-power SoC in the current market dynamics increases the complexity of characterizing cells in an accurate manner. Synopsys's SiliconSmart is a comprehensive tool that is used by many semiconductor design companies and foundries for standard, custom, and complex cell characterization.

The NetApp® scale-out architecture with the NetApp clustered Data ONTAP® 8.3 operating system provides all the storage efficiencies for the SiliconSmart files in the characterization process. It also improves job completion times by up to 7% with NFSv3 and with adequate storage optimization and best practices. This report highlights some performance tuning on the NetApp storage and Linux® clients in the compute farm to improve the efficiency of the application license costs and time to market. These optimizations do not change the cell characterization workflow, and the changes to the infrastructure are dynamic does not have any impact on the application workflow.

TABLE OF CONTENTS

LIST OF FIGURES

1 Introduction

Standard cells consist of preconfigured and laid-out functional block elements that provide a particular operation (such as the output of a two-input AND function being true only if both of the two inputs it contains are true). These cells can be simple or complex digital gates, sequential cells, and I/O cells. They are named standard cells because they typically must conform to a standard height (although their width can vary) to simplify the task of placing them on the chip, connecting them automatically by the place and route software, and to provide the most efficient packing density. These cells are designed for a specific function and usually come in libraries. Each of these libraries consists of a few hundred to a few thousand different cells.

Standard cell libraries are required for the following reasons:

- Design complexity keeps increasing because of area and yield optimization. Standardizing cell libraries by foundries reduces manufacturing costs and enhances yields.
- Complete custom chip design is no longer possible. Standard cell libraries expedite the chip design process through logical cell layout models.
- Standardization allows efficient automated software routing and connectivity among these cells for maximum packing density and improved performance. The optimized wire lengths prevent performance slowdown caused by poor connections or routing.

The characterization step provides timing, power, and noise behavior across a wide range of voltage, temperature, and process variations that make up the electrical behavior and specifications of the device being designed. Standard cell library characterization is an important component in the chip design cycle for the following reasons:

- Functionality behavior; electrical characteristics extraction; and simulation of gates, sequential cells, and so on in a chip must be fast and simple.
- The variability of performance across a wide range of processes, voltage, and temperatures must be analyzed and accounted for to guarantee performance specifications.
- Accurate timing, power, and noise are essential to guarantee device behavior under different operating conditions.

The Synopsys SiliconSmart tool is designed to generate library models with accurate timing, power, and signal integrity of the individual gates, which allows precise static timing and power analysis at the chip level. This requires many input files to simulate the library layout models and generates many output files, depending on the number of cell libraries characterized for various functions. The SiliconSmart tool uses a network of compute nodes to access files of various sizes from a shared storage infrastructure such as NetApp over a file-based protocol such as Network File System (NFS).

The NetApp clustered file system in Data ONTAP 8.3 offers scale-up and scale-out storage architectures that provide storage for the large and complex cell and memory library characterizations that are required by chip designers. It addresses the growing storage needs of these customers while efficiently handling the different workloads that are generated during the entire cell library characterization workflow. NetApp clustered Data ONTAP 8.3 provides the following key drivers to shorten the chip design process with a faster time to market and improved return on investment (ROI):

- Performance
- High availability and reliability
- Capacity
- Storage efficiency
- Agile infrastructure
- Data protection
- Manageability
- Low cost

2 Target Audience and Objectives

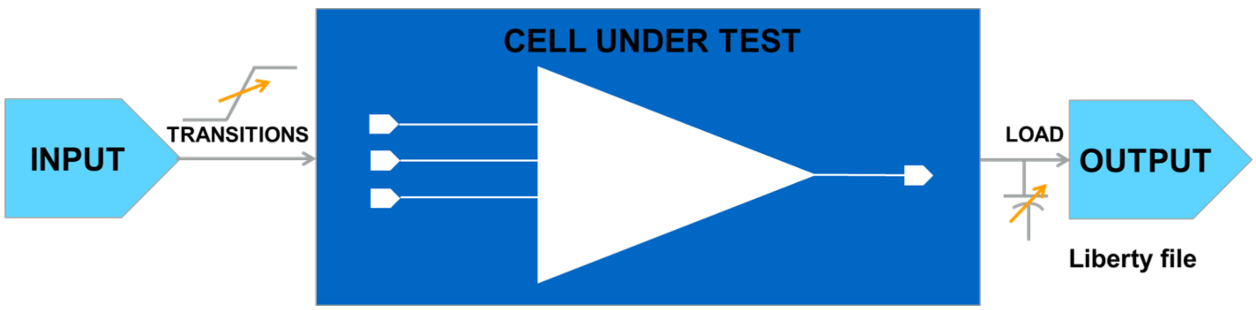
The SiliconSmart tool is one of the most popular cell and memory characterization tools used by chip manufacturers. This technical report is intended for cell library engineers, storage administrators, and architects. This paper provides the following information:

- Best practices and sizing required with clustered Data ONTAP 8.3 to support the performance, capacity, availability, and manageability requirements for SiliconSmart workloads
- How to use the NetApp scale-out clustered file system solution for the SiliconSmart application and for validating the performance improvements during the cell library characterization process

3 SiliconSmart Cell Library Characterization

SiliconSmart cell library characterization tools are critical in the chip design process because they provide library views for signal integrity, timing, and power analysis for cell layout on a chip. During this characterization process, netlists that contain connectivity information about a cell are provided as input to each of the cells in the library. The input transitions and output loads are varied, and the cell netlists are simulated with simulation program with integrated circuit emphasis (SPICE) engines to generate signal integrity (noise), timing, and power models.

Figure 1) Cell library characterization overview.



3.1 SiliconSmart Cell Library Characterization Workflow

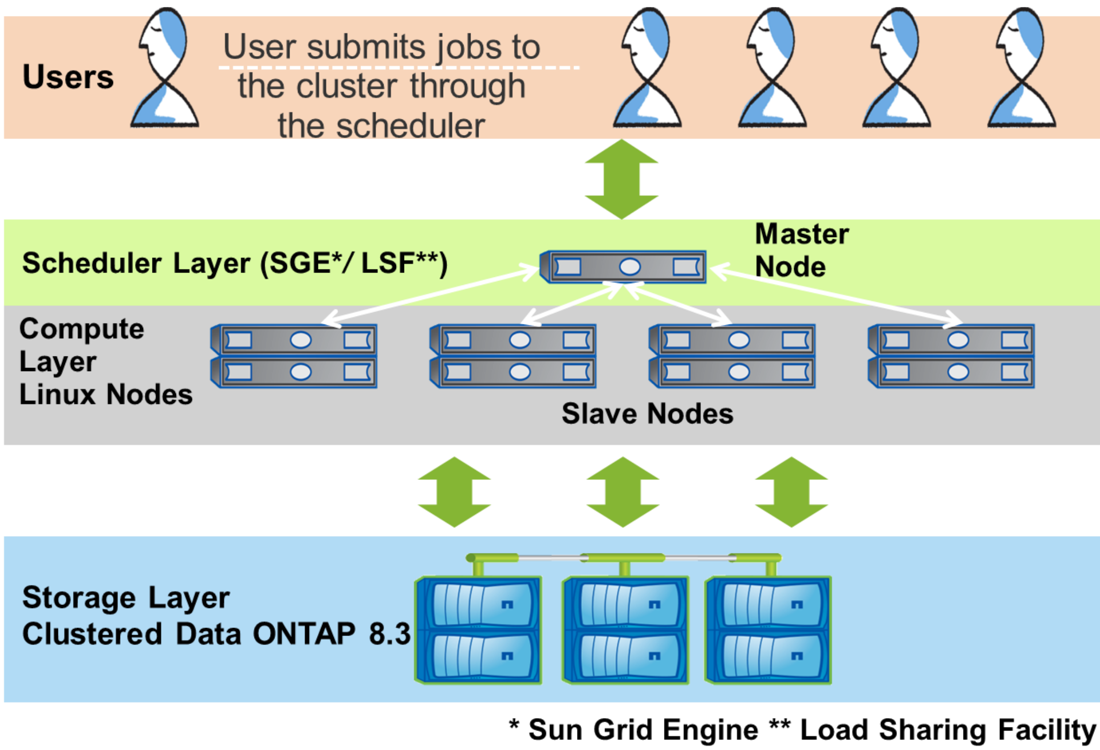
SiliconSmart performs the cell characterization in parallel across heterogeneous compute nodes that scale. SiliconSmart uses SPICE simulators such as HSPICE and FineSim (standalone as well as embedded) and generates a final liberty file with timing, power, and noise models. The application functions in four different phases:

- Phase 1: During this phase, the input SPICE netlist files per cell are read. It generates one text file per cell. Many text files will be generated depending on the number of cells.
- Phase 2: Each cell is analyzed and configured for the next phase. All the possible timing/noise/power arcs for a cell are configured. This phase creates hundreds of templates for each cell.
- Phase 3: This phase takes the inputs from the text files submitted in the configure phase, SPICE netlist, and process modules for simulation. SPICE simulations are run, and many simulation and SiliconSmart log files are generated.
- Phase 4: During the final phase, all the log files from the previous phase are read, and a single large liberty file is created. All the temporary files are deleted.

3.2 SiliconSmart Tool in Clustered Data ONTAP 8.3 Environment

It is not uncommon for tools to run on a high number of compute nodes—in the magnitude of thousands of cores—with job schedulers such as Load Sharing Facility (LSF) or Sun® Grid Engine (SGE) in cell library characterization environments. Apart from all the optimization that happens at the SiliconSmart application layer, it is imperative to also optimize and tune the compute nodes, the network, and the storage layers to complement the faster job completion times. Depending on the size and type of standard and custom cells, many files that need storage along with efficiency and performance are generated. Figure 2 shows a typical storage integration with the different cell library characterization workflows in a data center.

Figure 2) Cell library characterization with clustered Data ONTAP in a data center.



NetApp clustered Data ONTAP provides advanced technologies for software-defined storage that abstracts data from the underlying hardware by virtualizing the storage infrastructure with storage virtual machines (SVMs) to allow an efficient, scalable, nondisruptive environment. Some of these virtualization capabilities may be similar to past NetApp vFiler® unit functionality. Others go beyond anything else available today. Clustered Data ONTAP is built on the same trusted hardware that NetApp has been selling for years. We bring together the different hardware platforms, connect them, and give them the intelligence to communicate with each other in a clustered environment. Refer to TR-4238 for more detailed information on how clustered Data ONTAP can provide improvements to SiliconSmart environments.

3.3 Storage Cluster Node Sizing and Optimization

After the right hardware is configured and clustered Data ONTAP 8.3 is installed on the cluster setup, the following sizing efforts must be made to the RAID disk subsystem for optimal performance. All the cluster nodes are multipath in a storage failover configuration. In the event of hardware failure, the surviving partner node takes over the disks from the failing node to provide continuous accessibility to the data volumes.

How to configure clustered Data ONTAP for creating aggregates, volumes, and NFS protocol is not within the scope of this paper. Refer to TR-4067, "Clustered Data ONTAP NFS Best Practice and Implementation Guide," for clustered Data ONTAP 8.3 and NFS configuration details.

Aggregate and RAID Group Sizing Best Practices

- Disk spindles help to improve write performance by reducing CPU utilization and the time it takes to look for free space, while writing full stripes of data on the storage controller as the file system ages from constant deletions and insertions. For optimal sizing:
  - A RAID group should have 28 SAS disks.
  - A minimum of 6 RAID groups should be present in a single aggregate.
- Spread the volumes across all the high-end platform (FAS8080) cluster nodes that are part of a single project that has a high-performance requirement. All of these volumes are connected by the dynamic cluster namespace junctions:
  - Spreading out the volumes across the high-end platform prevents putting all your eggs (volumes) in one basket (aggregate). In this way, all the project volumes can meet the IOPS and bandwidth requirements from all the controller nodes of which these volumes are part and not saturate just a single controller.
  - Spreading out the volumes also allows moving the volumes nondisruptively in the cluster namespace for workload balancing to meet the high-performance SLO.
  - During the verification phase, much transient data is generated that has large writes and reads accompanied by many deletions. Isolating the volume that writes the transient data on a different cluster node from other volumes that are part of the project helps to prevent a single controller from being the bottleneck.

4 Synopsys SiliconSmart Validation with NetApp Clustered Data ONTAP 8.3

The various chip design houses constantly run cell library characterization on cells or circuits during the chip design cycle based on standard guidelines developed by the foundries. The complex cell library characterization process leads to high-performance demands, with low latency and faster job completion times.

The SiliconSmart application performance was validated at the Synopsys lab in Mountain View, California. For this performance validation, NetApp, through its partnership with Synopsys, ran the tool on real-production 28nm standard 519-cell data from a large chip design customer. The test results from the Synopsys lab demonstrated consistent improvement with clustered Data ONTAP 8.3 optimization, network, and compute nodes.

4.1 Performance Validation Objectives

The primary objectives of the SiliconSmart performance validation with clustered Data ONTAP 8.3 were the following:

- Validate that the SiliconSmart job completion time (wall-clock time) on clustered Data ONTAP 8.3 is on par with or better than the baseline for jobs performed on Data ONTAP 8.2.3 operating over NFSv3.

- Explore new technology such as pNFS for cell library characterization workloads. Validate that the SiliconSmart job completion time with clustered Data ONTAP 8.2.3 over pNFS is comparable to or better than that of clustered Data ONTAP 8.2.3 over NFSv3.
- Enable reduction in the job completion time, allowing users to move on to other cell library characterizations and make better use of their license costs. This improves the ROI of the SiliconSmart tool.

4.2 SiliconSmart Test Lab Details

The Synopsys lab consists of 16 compute nodes. Each compute node has 8 cores and 64GB of physical memory, with a 1GbE connection to the storage through a 10GB switch plane. The tests were able to scale up to 128 cores. All of these compute nodes ran on Red Hat Enterprise Linux (RHEL) 5.7 for the baseline and storage optimization tests. Then all the compute nodes were reimaged to RHEL 6.5 kernel 2.6.32-504.8.1.el6.x86_64 for validating NFSv4.1/pNFS test results. LSF was used as the scheduler to submit the jobs in the compute farm. SiliconSmart 2014.09-SP2 was used for this test.

On the storage side, we had a two-node FAS8080 cluster with 10GB data ports. Each cluster node had two 10GB data ports, aggregated to provide 20GB of network bandwidth that can provide up to 2.4GB/sec throughput. Each of the cluster nodes had 3 shelves of 10,000-RPM 600GB SAS disks. Each cluster node had 1TB of a PCIe-based Flash Cache™ card. The NetApp cluster was running on clustered Data ONTAP 8.2.3.

Both the baseline and the optimizations were performed on clustered Data ONTAP 8.2.3 and 8.3 and had two RAID groups, with 24 disks on each of the aggregates.

4.3 Test Plan

The cell libraries from a major chip design company were used to run the various tests on NetApp storage to generate the workload with the SiliconSmart tool. The tool ran on 519 cells to put stress on the storage. Four different sets of tests were performed, including:

- Cell library characterization tests over NFSv3 with clustered Data ONTAP 8.2.3. This is a typical Synopsys environment that it uses in production. This test was considered to be the baseline for the rest of the tests that followed.
- Cell library characterization tests over NFSv3 with clustered Data ONTAP 8.2.3 with optimization.
- Cell library characterization tests over NFSv4.1/pNFS with clustered Data ONTAP 8.2.3 fully optimized for Red Hat clients and storage.
- Cell library characterization tests over NFSv3 with clustered Data ONTAP 8.3 with optimization.

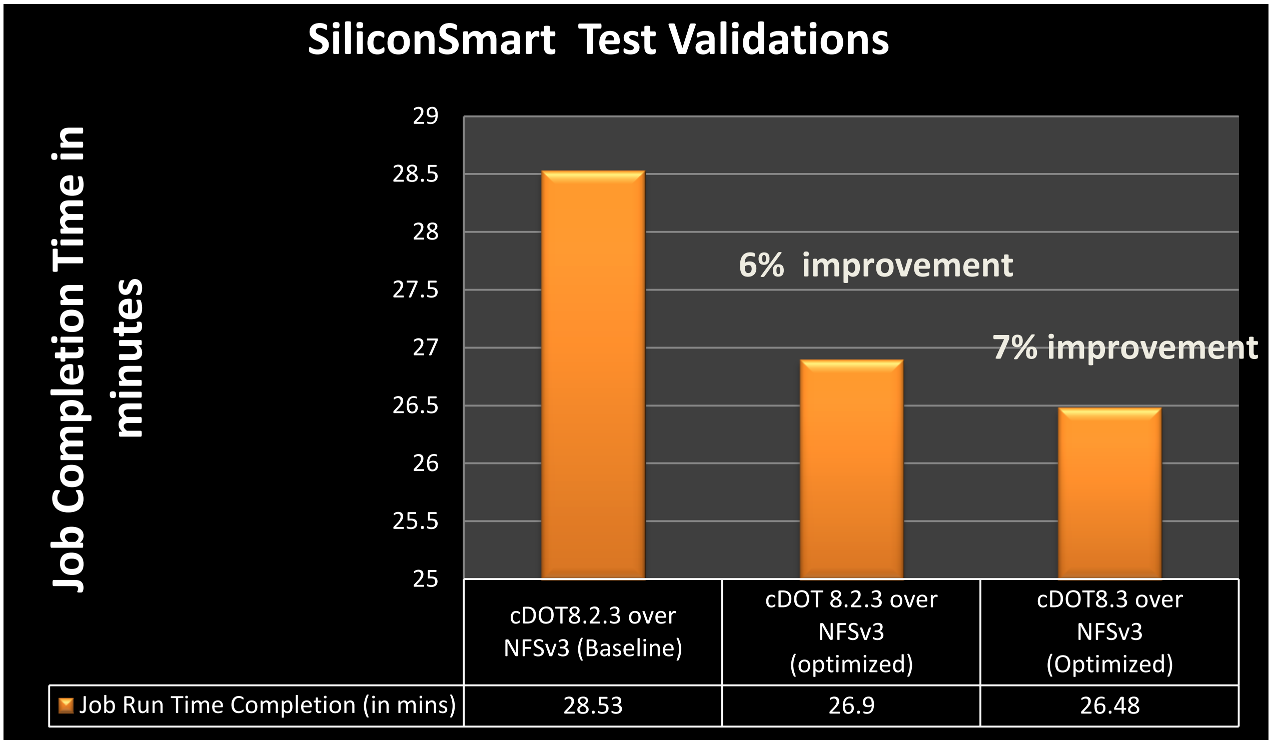
4.4 SiliconSmart Test Results

The SiliconSmart test results indicated up to 7% improvement in the job completion time in optimized environments compared to the baseline tests. Each test was run for less than 30 minutes with a subset of the large cell library for more control over managing the test run times. For multisocket machines, NUMA can affect the job completion times. The environment for the baseline tests that were performed used the following:

- High-end FAS8080 NetApp storage platform
- 1TB Flash Cache with the recommended settings
- 20Gbe network connection
- The recommended RAID group size

Note: The baseline test is already optimized to a greater degree and run on a high-end NetApp storage platform compared to what our customers use in their production environments.

Figure 3) SiliconSmart performance test results.



4.5 SiliconSmart Performance Validations and Observation

As mentioned in section 4.4, the test environment for baseline is greatly optimized on a high-end platform. The test results showed a progressive improvement in the job completion time over NFSv3 in an optimized environment compared to baseline. The optimization was done on the RHEL nodes and on the storage. The optimized tests were run on two versions of clustered Data ONTAP: 8.2.3 and 8.3. Clustered Data ONTAP 8.3 provided the best results in the optimized environment compared to the baseline. RHEL6.5, FAS8080 with Flash Cache, and clustered Data ONTAP 8.3 with all the best practice recommendations provided up to 7% faster job completion time compared to RHEL5.7, FAS8080 with Flash Cache and clustered Data ONTAP 8.2.3.

NFSv4.1/pNFS tests did not deliver any improvements over the baseline. The SiliconSmart workload is primarily random reads and writes. The current implementation of pNFS in RHEL6.5 is not matured to handle random workloads. Joint engineering work between NetApp and Red Hat is in progress to improve random workload performance. RHEL7.1 has some preliminary improvements for random workloads.

However, tests over NFSv4.1 with pNFS disabled on the nodes and the storage were performed. Many customers are looking forward to move some of the critical workloads to NFSv4.1 for security reasons. NFSv4.1 provides file-level and directory-level security with access control lists (ACLs) and user security with Kerberos. The tests indicated that the NFSv4.1 run times were better than the baseline tests. Customers who are planning to move some part of the workload over to NFSv4.1 should not notice a performance degradation, provided the recommended RHEL and NetApp clustered Data ONTAP versions are used for NFSv4.1.

5 Best Practices for SiliconSmart Tool with Clustered Data ONTAP 8.3

The SiliconSmart tool is one of the most common tools used for cell library characterization during chip design cycles. An increasing number of customers are deploying clustered Data ONTAP 8.3 for storage, supporting the characterization phase. Scale-out clustered FAS storage has to be properly architected to handle the SiliconSmart workload. The aggregates and volumes that store the cell libraries and inputs such as the netlists and so on must be optimally laid out in the cluster nodes.

The best practice recommendations in this section provide guidance to optimize clustered Data ONTAP, the network layer, and the compute nodes for SiliconSmart workloads. It is imperative to also validate some of the key clustered Data ONTAP 8.3 features and functions to improve the overall efficiency of the SiliconSmart application.

5.1 Storage Cluster Node Architecture

NetApp highly recommends implementing the right storage platform in a clustered Data ONTAP setup, along with adequate storage sizing and configuration to accommodate cell library characterization workloads for standard and custom cells that have different SLOs. If the workload is performance driven and has the highest SLO, NetApp recommends storage controllers with multiple cores and a large memory footprint. Faster serial-attached SCSI (SAS) disks should always be used for designs that require a faster response time.

Choosing the Right Hardware for SiliconSmart Workloads in a Clustered Scale-Out Architecture
The SiliconSmart cluster setup can provide different SLOs for standard and custom cells and other dependencies that can coexist in the same or different SVMs. The choice of hardware can be different based on the price-to-capacity (GB) and price-to-performance ratios for various SLOs: • If the SiliconSmart workload requires that performance be at the highest level, NetApp strongly recommends FAS8080 controllers with a minimum of 900GB 10k RPM second-generation SAS disks and a minimum of 6GB backplane. • If the cluster setup is designed to accommodate library database (.ldb) files for archiving, NetApp recommends a minimum of FAS3270 controllers with SATA disks. • NetApp recommends having a minimum of 1TB PCIe-based Flash Cache 2 cards on each controller. • A four-node, eight-node, or larger cluster with different types of disks (SSD, SAS, and SATA) can be configured based on the SLOs for different workloads. • NetApp highly recommends engaging with the appropriate NetApp sales account team to evaluate your business requirements before architecting the cluster scale-out setup in your environment.

5.2 File System Optimization

After the volumes have been created in the SVM, NetApp recommends certain best practice configurations on the aggregate and volumes to address the following issues in a SiliconSmart storage environment:

- Fragmentation after constant writes and deletions to the file system during the assembly phase
- Fragmentation of free space for writes to complete a full stripe

The file system can be kept healthy at all times with the help of some maintenance and housekeeping activities on the storage as it ages and grows in size, including:

- **Defragmenting the file system.** Reallocate is a low-priority process that helps constantly defragment the file system, and it can run in the background. However, NetApp recommends implementing measures to keep the aggregate utilization under 80%. If the aggregate runs close to 90% capacity, the following considerations apply:
  - Some amount of free space is required to temporarily move the data blocks to free space and to rewrite to those blocks in full and complete stripes in contiguous locations on the disk, thereby optimizing the reads that follow.
  - Insufficient space in the aggregate allows reallocate to run in the background, but defragmentation of the file system is never completed.
  - An NDO to move the production chip design volume to another controller that is part of the cluster setup for capacity balancing must occur.
  - New shelves must be added to the original controller to provide more space to the aggregate that is running low on space. Perform `reallocate start -vserver vs1_eda_lib -path /vol/VOL06 -force true` for all the volumes in that aggregate.
  - This forces all the existing volumes to spread out on the new disk spindles that were added to the aggregate. Otherwise, the new writes coming into that aggregate go only to the new disks.
- **Defragmenting free space.** Continuous segment cleaning, which was introduced in clustered Data ONTAP 8.1.1 and further optimized in clustered Data ONTAP 8.3, helps coalesce the deleted blocks in the free pool to use for subsequent writes.
- **Thin provisioning.** The volumes in the cluster namespace can be thin provisioned by disabling the space guarantee. This provides flexibility to provision space for chip design or different project volumes that can autogrow in increments of 1GB.

NetApp recommends enabling the following storage options to optimize the entire life of the file system.

File System Optimization Best Practices

- The following settings cannot be put into place from the cluster shell. They can be made only at the CLI mode:

```
bumblebee::*> vol modify -vserver vs1_eda_lib -volume vol6 -min-readahead false
  (volume modify)
Volume modify successful on volume: VOL06
bumblebee::*> aggr modify -aggregate aggr1_fas6280c_svl09_1 -free-space-realloc on
bumblebee::*> vol modify -vserver vs1_eda_lib -volume CMUSERVOL -read-realloc space-optimized
Volume modify successful on volume VOL06 of Vserver vs1_eda_vcs.
bumblebee::*> vol modify -vserver vs1_eda_lib -volume VOL06 -atime-update false
Volume modify successful on volume VOL06 of Vserver vs1_eda_vcs.
```

- NetApp recommends always setting up an alarm that triggers as soon as the aggregate reaches 80% capacity. The critical chip design volumes that need more space can automatically use WFA or manually be moved to another aggregate on a different controller.
- NetApp recommends thin provisioning the volumes. This can be done when the volumes are created, or they can be modified later. It can also be implemented by using OnCommand® System Manager 3.0 from a GUI:

```
bumblebee::*> vol modify -vserver vs1_eda_lib -volume VOL06 -space-guarantee none
  (volume modify)
Volume modify successful on volume: VOL06
```

- Adequate sizing is required for the number of files in each directory and the path name lengths. Longer path names lead to a higher number of NFS LOOKUP operations.
- Default quotas cannot be implemented for users and groups. Include an explicit quota entry for users and groups.

5.3 Storage Network Optimization

After you create the aggregates and volumes based on the recommended sizes to support the cell library workload, you must then configure the network. At that time, the cluster, management, and data ports are all physically connected and configured on all the cluster nodes to the cluster switches. Configuring the network includes:

- **Data port aggregation.** Before the LIFs and routing tables are configured for each SVM, it is very important to aggregate at least two 10GbE data ports for handling the cell library workloads. Depending on the number of chip design and tool volumes that each controller has, NetApp recommends aggregating a larger number of data ports than required to achieve the desired SLO.
- **LIF failover.** In clustered Data ONTAP, LIF IP addresses are no longer tied to physical network ports. They are part of the SVM. When LIF IP addresses are created, NetApp recommends configuring a

failover path in case the home port goes offline. If a data port failure occurs, the LIF can fail over nondisruptively to another controller. This allows the application to continue accessing the volume even though the LIF moved to a different controller in the SVM.

Storage Network Optimization Best Practices

Aggregate at least two 10GbE data ports on each cluster node that interfaces with the compute farm:

```
bumblebee::*> network port ifgrp create -node fas6280c-svl07 -ifgrp
e7e -distr-func ip -mode multimode
```

```
bumblebee::*> network port ifgrp add-port -node fas6280c-svl07 -ifgrp
e7e -port e0d
```

```
bumblebee::*> network port ifgrp add-port -node fas6280c-svl07 -ifgrp
e7e -port e0f
```

Use the following option to configure the LIF failover for any LIF configured in the SVM clusterwide:

```
bumblebee::*> net int modify -vserver vs1_eda_lib -failover-group
clusterwide -lif vs1_eda_lib_data3 -home-node fas6280c-svl09 -home-
port e9e -address 172.31.22.172 -netmask 255.255.255.0 -routing-group
d172.31.22.0/24
```

- Always follow a ratio of one volume to one LIF. That means that every volume has its own LIF. If the volume moves to a different controller, the LIF should move along with it.

5.4 Flash Cache Optimization

A caching tier on the storage supplements the number of I/O requirements for the cell library workload. Flash Cache enables the read workload: metadata and random reads. Random data access is a function of the disks. A higher number of disk spindles helps to generate a greater amount of read I/O. Storage platforms with PCIe-based Flash Cache improve read performance.

Flash Cache serves additional I/O requests from the flash-based cache, while disk spindles help improve the performance of write-intensive workloads.

All the movement of the data between the base memories (DRAM), Flash Cache, and disks happens transparently to the application.

NetApp recommends enabling `options flexscale.random_write_through_blocks` in Flash Cache settings in clustered Data ONTAP 8.3. The SiliconSmart tool tends to read random data from the storage as soon as it is written. This option allows the I/O requests to be served from Flash Cache instead of from the disks, which improves application performance. To cache random read workloads, `flexscale.normal_data_blocks` should be enabled.

Flash Cache Optimization Best Practices

The following settings cannot be made from the cluster shell. They can be made only at the node shell CLI of each controller. These commands allow you to get from the cluster shell to the node shell on each controller:

```
bumblebee::*> system node run -node fas6280c-svl09

Type 'exit' or 'Ctrl-D' to return to the CLI

fas6280c-svl09> priv set diag

Warning: These diagnostic commands are for use by NetApp

         Personnel only.

fas6280c-svl09*>

Enable Flash Cache:

options flexscale.enable on

Enable caching of metadata and random read data:

options flexscale.normal_data_blocks on

Enable caching of sequential read data:

options flexscale.random_write_through_blocks on (Available only in
cDOT8.3 and later)

Type exit in the node shell to get back to the cluster shell:

fas6280c-svl09*> exit

logout

bumblebee::*>
```

5.5 Network File System (NFSv3) Optimization

Almost all of the cell library characterization workload accesses the file system from the back-end storage controllers over the NFSv3 protocol.

NFSv3 is a stateless protocol and is geared primarily toward performance-driven workloads such as the SiliconSmart environment with asynchronous writes. Communication between the NFSv3 client and the storage happens over remote procedure calls.

RHEL 5.x is the most common Linux vendor-supported version that is used by most of the semiconductor companies in SiliconSmart compute farm environments.

NFS runs in the kernel space of the network stack in the clustered Data ONTAP code. Minimal tuning is required for NFS running on the network stack.

As one of the benefits of clustered Data ONTAP 8.2, a fast path for the local data path is available for NFSv3.

With the large number of compute nodes accessing files from a single controller, the TCP receives window size, or the receive buffer may quickly become exhausted. The storage will not accept any further TCP windows over the wire until the receive buffer is freed up. NetApp therefore recommends increasing the TCP receive buffer value.

NetApp recommends enabling NFS failover groups to provide another layer of protection at the protocol level.

NFSv3 Optimization Best Practices

- The command `force-spinnp-readdir` enables making effective readdir calls from the data stack; increasing the TCP buffers also optimizes performance. The buffer size also must be increased:

```
nfs modify -vserver vs1_eda_lib -force-spinnp-readdir true -tcp-max-xfer-size 65536
```

- The following steps must be followed to configure the NFS failover groups. The example shows how the LIFs vs1_eda_lib_data3 and vs1_eda_lib_data4, which are assigned to an NFS failover group, move the NFS traffic over port e7e on node fas6280-svl07.

```
bumblebee::*> network interface failover-groups create -failover-group lib_failover_group -node fas6280c-svl07 -port e7e

bumblebee::*> network interface failover-groups show -failover-group lib_failover_group -instance

Failover Group Name: lib_failover_group
               Node: fas6280c-svl07
               Port: e7e

1 entries were displayed.

bumblebee::*> network interface modify -vserver vs1_eda_lib -lif vs1_eda_lib_data3,vs1_eda_lib_data4 -failover-group lib_failover_group
2 entries were modified.
```

5.6 Storage QoS

Storage QoS provides another level of storage efficiency in which IOPS and bandwidth limits can be set for workloads that are not critical or when setting up SLOs on different workloads. In EDA environments, storage QoS plays an important role:

- Rogue workloads can be isolated with proper IOPS and bandwidth limits set in a different QoS policy group for users who generate these kinds of workloads in a production environment. This can be done at an SVM, volume, or specific file level.
- In an IT-managed cloud infrastructure, storage QoS helps to run multiple tenants with different service-level offerings. New tenants can be added to the existing one as long as the storage platform has the headroom to handle all the workload requirements. Different workloads, such as builds, verifications, cell library characterization, and other EDA tools, can coexist on the same storage controller. Each of the individual workloads can have different performance SLOs assigned to it.

Storage QoS Configuration

- A QoS policy group must be created for different SVMs in the cluster. In the following example, two QoS policy groups are created; business_critical and non_critical have different IOPS and bandwidth settings:

```
bumblebee::*> qos policy-group create -policy-group business_critical -vserver vs1_eda_lib -max-throughput 1.2GB/sec

bumblebee::*> qos policy-group create -policy-group non_critical -vserver vs1_eda_lib -max-throughput 2000IOP

bumblebee::*> qos policy-group show

Name             Vserver     Class        Wklds Throughput

---------------- ----------- ------------ ----- ------------

business_critical

                 vs1_eda_lib user-defined -     0-1.20GB/S

non_critical     vs1_eda_lib user-defined -     0-2000IOPS

2 entries were displayed.
```

- Volume vol06 is then set with the QoS policy group non_critical:

```
bumblebee::*> vol modify -vserver vs1_eda_lib -volume CMSGE -qos-policy-group non_critical

  (volume modify)

Volume modify successful on volume: CMSGE
```

- The file writerandom.2g.88.log has been set to a non_critical QoS policy group. You cannot set a QoS policy group on a file when the volume that holds that file already has a QoS policy group set on it. The QoS policy group on the volume must be removed before the policy can be set on a particular file in that volume:

```
bumblebee::*> file modify -vserver vs1_eda_lib -volume VOL06 -file //OpenSPARCT1/Cloud_free_trial_demo/OpenSparc-T1/model_dir/farm_cpu_test/writerandom.2g.88.log -qos-policy-group non_critical

bumblebee::*> qos workload show

Workload       Wid   Policy Group Vserver  Volume    LUN    Qtree  File

-------------- ----- ------------ -------- --------- ------ ------ -------------

CMSGE-wid12296 12296 non_critical vs1_eda_lib

                                           CMSGE     -      -      -

file-writerandom-wid11328

               11328 non_critical vs1_eda_lib

                                           VOL06

                                                     -      -

/OpenSPARCT1/Cloud_free_trial_demo/OpenSparc-T1/model_dir/farm_cpu_test/writerandom.2g.88.log

2 entries were displayed.
```


5.7 Nondisruptive Operation (NDO)

NDO completely changes the way that clustered Data ONTAP keeps data alive and available to the applications and the users who access the data. Disruptive scenarios were tested in the Synopsys lab under SiliconSmart workloads to determine whether there was disruption to users at the application layer:

- When a data port was taken offline, the LIF IP address instantly failed over to another node in the cluster. This did not cause any outage for the user accessing the data under the load.
- The chip library volume was moved to a different cluster node under the active SiliconSmart load for capacity- and workload-balancing reasons. The volume and the LIF were moved to the new location in the cluster namespace without disrupting the user’s running jobs on the chip design volume.

Nondisruptive Operation with Volume Move

- In this example, the volume VOL06 is moved from an aggregate in FAS8080-svl02 to an aggregate in FAS8080-svl01 while the SiliconSmart workload is in progress. There is no disruption to the application when the volumes are moved on the storage.

```
bumblebee::*> vol move start -vserver vs1_eda_lib -volume VOL06 -
destination-aggregate aggr1_fas8080c_svl01_1

  (volume move start)

[Job 17268] Job is queued: Move "VOL06" in Vserver "vs1_eda_lib" to
aggregate "aggr1_fas8080c_svl01_1". Use the "volume move show -vserver
vs1_eda_lib -volume VOL06" command to view the status of this
operation.

job show <job_id> can be used to check the status of the “vol move.”

bumblebee::*> job show 17268

                            Owning

Job ID Name                 Vserver    Node           State

------ -------------------- ---------- -------------- ----------

17268  Volume Move          bumblebee  fas8080c-svl01 Success

       Description: Move "VOL06" in Vserver "vs1_eda_lib" to aggregate
"aggr1_fas8080c_svl01_1"
```

- NDO can also be performed during hardware technical refreshes when all the volumes on an entire node can be evacuated to another cluster node and moved back nondisruptively to the new controllers after the refresh process.
- Nondisruptive upgrades (NDUs) can also be performed on clustered Data ONTAP versions and the shelf and disk firmware without causing any outage to the application.

6 Compute Farm Optimization

The engineering compute farms in a chip design environment consist of tens of thousands of cores, which translates to hundreds to thousands of physical compute nodes. Virtualization is usually not deployed in these farms. Linux is the most commonly used operating system in the compute farm. Linux clients in the compute farm provide the number of cores that are required to process the number of jobs submitted.

For better client-side performance with clustered Data ONTAP 8.3, the cell library characterization application and the schedulers, such as Sun Grid Engine (SGE) or Load Sharing Facility (LSF), must be run on RHEL 5.8 and later or RHEL 6.5 and later. Synopsys has validated and certified RHEL 6.5 for cell library characterization tools.

6.1 Best Practices for Compute Nodes

Considering the high volume of nodes in the compute farm, it is unrealistic to make significant changes dynamically on each of the clients. Based on the SiliconSmart workload evaluation, the following recommendations on the Linux clients make a big contribution to improving the job completion times for various chip design activities.

Compute Node Optimization for NFSv3 Mounts

- Turn off hyperthreading on the BIOS setting of each of the Linux nodes if the node is multisocket. This is not required if the compute nodes are single socket.
- Use the recommended mount options while mounting over NFSv3 on the Linux compute nodes: `vers=3,rw,bg,hard,rsize=65536,wsize=65536,proto=tcp,intr,timeo=600.`
- Set `sunrpc.tcp_slot_table_entries = 128`; this improves TCP window size. This option is fine for pre-RHEL 6.4 kernels that mount over NFSv3. RHEL 6.5, however, includes changes to the TCP slot table entries. Therefore, the following lines must be included when mounting file systems on an RHEL 6.5 kernel over NFSv3. The following lines are not required when mounting over NFSv4.1; however. NetApp storage may have its network buffers depleted by a flood of RPC requests from Linux clients over NFSv3:
  - `Create a new file: /etc/modprobe.d/sunrpc-local.conf`
  - `Add the following entry: options sunrpc tcp_max_slot_table_entries=128`
- If the compute nodes are using 10GbE connections, the following tuning options are required. The following changes DO NOT apply for clients that use 1GbE connections:
  - Disable irqbalance on the nodes:

```
[root@ibmx3650-svl51 ~]# service irqbalance stop
Stopping irqbalance:                                           [  OK  ]
[root@ibmx3650-svl51 ~]# chkconfig irqbalance off
```

- Set net.core.netdev_max_backlog = 300000**;** avoid dropped packets on a 10GbE connection.

7 Summary

Cell or circuit design is getting more complicated with respect to size and yield optimization on the silicon layers. As more and more <20nm chips are designed and manufactured for different consumer products, characterization of standard cell libraries becomes critical to profile the different characteristics and behavior of functions of a chip design across a broad range of operating conditions.

It is imperative to expedite the characterization process of the standard cell libraries to improve the overall chip design process time. Storing, accessing, and managing all the cell libraries in a shared storage infrastructure require low latency, high reliability, efficiency, and a single pane of manageability of the cell library data. All the validations and best practices listed in this report clearly indicate that NetApp clustered Data ONTAP 8.3, with the recommended storage optimizations and sizing, can accelerate the cell library characterization process.

The main objective for integrating Synopsys's SiliconSmart tool with the NetApp clustered file system is to improve the job completion time at the application layer. Another very important factor that is conducive to the chip design process is the fact that NetApp clustered Data ONTAP 8.3, with adequate storage optimization and sizing, can improve the performance of workloads generated from the various design and characterization tools that coexist in the chip design and manufacturing process. With QoS, workloads can be tied to different SLOs for tools running on the same node or in different cluster nodes in a scale-out architecture.

NetApp clustered Data ONTAP 8.3 allows load balancing of the cell library volume in the cluster namespace by moving it to different controller nodes in the cluster that is under load without any disruption to the application.

8 Conclusion

There is a high level of complexity in the cell library characterization process where foundry process, voltage, and temperature are validated and modeled for smaller silicon surface areas that are designed to perform different functions. As Synopsys keeps optimizing the SiliconSmart tool in every new release, it is important that the compute nodes, the storage, and the protocol all contribute to the overall job completion time.

NetApp originally set out to improve the job completion time for the SiliconSmart application with clustered Data ONTAP 8.3. With all the validations and optimizations at the compute nodes, NFS protocol, and storage layers, we ended up achieving improvement of up to 7% for cell library characterization with NFSv3. These improvements will provide a huge benefit compared to the application running unoptimized and nonconducive storage platforms.

This result leads to the conclusion that sizing adequately, optimizing the storage along with regular file system maintenance, and choosing the right protocol can improve cell library characterization performance significantly. This translates into two important factors that drive business in the EDA industry:

- Improved ROI with optimized license costs
- Faster time to market

Refer to the [Interoperability Matrix Tool (IMT)] on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

TR-4390-0315

NetApp®