SAND™ TECHNOLOGY

INTELLIGENT INFORMATION MANAGEMENT >>

WHITE PAPER

SAND™/DNA™ Access™

INCREMENTALLY SCALABLE NEARLINE SOLUTION

Sun microsystems Partner Advantage PRINCIPAL PARTNER

Contents

Visit SAND Technology on the Web at www.sand.com.

06/02/16

Introduction

An incrementally scalable nearline storage solution should offer an environment where Service Level Agreements can be satisfied simply by adding hardware to support higher throughput or larger data volumes. This document describes a solution based on SAND/DNA Access and a specific hardware architecture that enables just this kind of scalability.

Definition of Scalability

For the purposes of this discussion, scalability can be defined as the ability of a system to increase performance under an increased load when resources (typically hardware) are added.

A solution with perfectly linear scalability should be able to demonstrate the following:

- When processing power is doubled, the solution should perform the same amount of work in half the time, or provide double the throughput (that is, perform twice the work in the same time).
- When processing power remains the same, doubling the workload should double the service time.

Normally, linear scalability is demonstrated in terms of a range of functionality, including:

- Load Speed
- Single query execution
- Multi-query execution
- Database size.

SAND/DNA Access

SAND/DNA Access has been designed to provide linear scalability. Its architecture is based on a shared-nothing model that is able to take advantage of a distributed hardware architecture in which storage is shared. The smallest SAND/DNA Access processing component can be distributed among multiple servers; each of these components is fully multi-threaded, ensuring scalability within a distributed processing architecture based on small SMP nodes or servers.

The SAND scalable architecture is truly dynamic, in that addition or removal of processing capabilities (execution nodes) or storage capacity can be performed transparently to the users. This offers the potential for truly dynamic workload management.

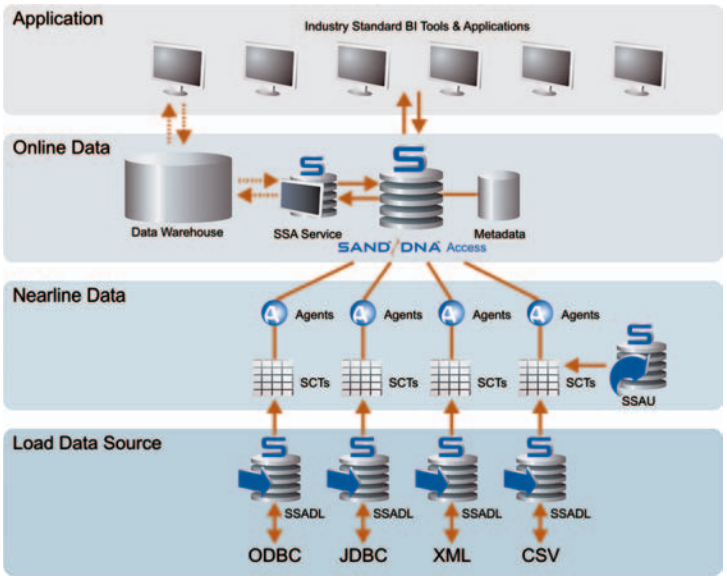

Figure 1: High-level view of the SAND/DNA Access architecture

Hardware Architecture

The SAND/DNA Access software solution is designed to scale in a Blade or Grid computing environment direct-attached or network-attached to multiple SANs or NAS units. The blade or grid computing design is perfect for delivering incremental scalability in small increments. The minimal configuration of a single node or blade server should be as follows:

- 2 x 64-bit processors
- 4 GB real memory
- Internal SCSI drive of at least 40 GB, for temporary storage
- 1024 Mbps BASE-T Ethernet port
- Solaris 10 or Linux operating system

The SAN or NAS configuration should be based on a minimal capacity of 512 GB, in a Raid1 or Raid5 configuration, and equipped with a cache of 1 GB.

A single node or blade server coupled with a basic SAN or NAS is the hardware "building block" of SAND's incrementally scalable nearline storage. A modeling approach can be used to identify the number of nodes or blade servers and SAN or NAS units required to achieve a specific performance level for a given workload. This hardware/software solution is able to deliver linear scalability, as demonstrated by the benchmark results presented in the next section.

Benchmark Tests

In October 2005, a benchmark was jointly conducted by SAND Technology and Sun Microsystems to demonstrate the scalability of SAND/DNA Access running in a Sun Grid computing environment.

The Sun Grid computer was configured as follows:

- 32 x V20z Nodes
  - 2 x 2.6 GHz Opteron processors
  - 8 GB of RAM
  - 1 x 73 GB internal Ultra320 SCSI Drive
  - 1 x 1000 Mbps BASE-T Ethernet Port
  - 1 x TIA/EIA-232-F asynchronous (DB9) port (fibre channel)
  - Solaris 10
- 4,427 GB of storage distributed among 6 Sun StorEdge 3510 FC Arrays

Version 2.1 of SAND/DNA Access was used for the benchmark.

The benchmark was conducted using 165 TB of customer data, with a total of 400 billion rows.

The benchmark measured the solution's scalability in terms of the following functionality:

1. Load speed
2. Single Query Execution
3. Multi-User Concurrent Query Execution
4. Database Size

1. Load Speed

Load speed testing identified the solution's ability to handle large volumes of data (gigabytes per hour) on different hardware configurations. The following graph presents the results:

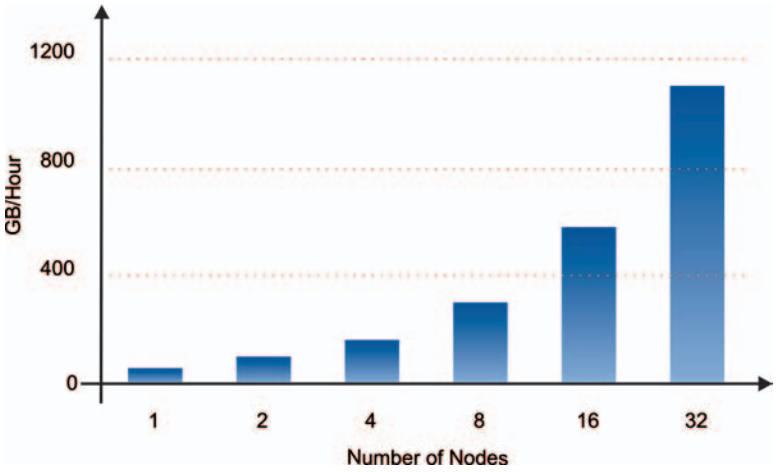

Figure 2: Load Scalability

The hardware/software solution demonstrated linear scalability, with load speeds reaching 1088 GB/ hour using a configuration with 32 nodes.

2. Single Query Execution

This test measured the solution's ability to improve performance for a single query using different hardware configurations. The following graph presents the results:

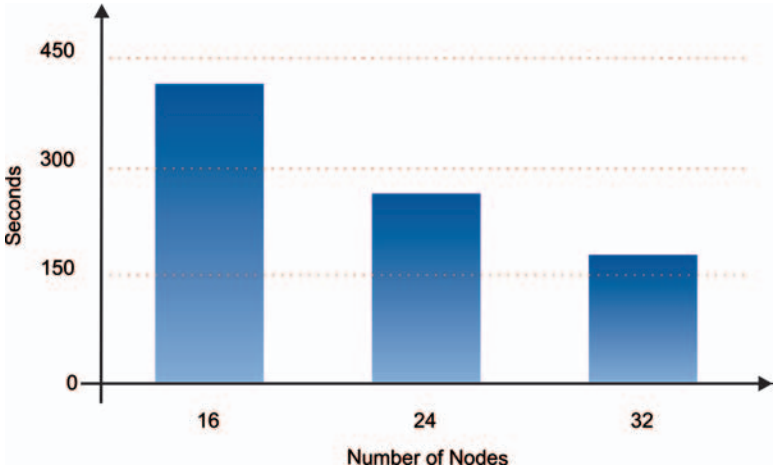

Figure 3: Single Query Scalability

The result demonstrated perfect linear scalability, with a peak performance of 70 million rows processed per second, equivalent to 2 TB of data scanned per minute.

3. Multi-User Execution

This test measured the solution's ability to support a larger number of users executing concurrent queries with different configurations of processing power. The following table presents the results:

Number of Users	Number of Nodes			
	6	12	24	30
6	19.1	11.2	7.9	8.4
12	31.6	17.1	9.7	9.1
24	62.8	30.1	12.3	10.4
48	152.6	58.9	17.8	12.3
60	178.5	76.1	21.4	13.9
96	299.4	110.8	29.8	16.7

Table 1: Average Response Times (Seconds)

The results demonstrated perfect linear scalability for average query response time as the number of concurrent users and processing nodes increased.

4. Database Size

This test measured the solution's ability to perform a single query executing a full table scan on tables of different sizes. The following graph presents the results:

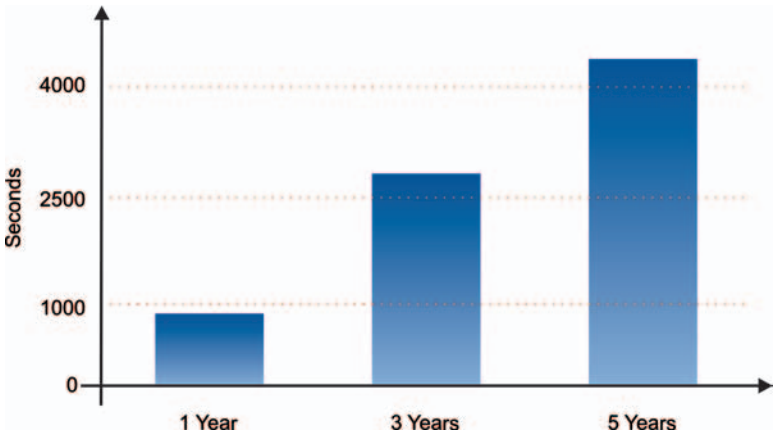

Figure 4: Database Size Scalability

Here the solution offers better than linear scalability: the time required to process five years of data was less than five times what was required to process the data for a single year. The query speed figures show that full table scans on over 382 billions rows were performed at a rate of 2 TB per minute.

Conclusion

The benchmark results described in this paper demonstrate linear or better scalability with very high performance. They show that the SAND incrementally scalable nearline storage solution coupled with a blade or grid computing architecture is the perfect way to achieve the level of "on-demand" performance that customers are currently looking for. The dynamic configuration of this solution enables processing power to be shared between applications, and demonstrates the potential for an environment in which processing power can be treated as a simple commodity.

About SAND Technology

SAND is an international provider of intelligent information management software. The SAND/DNA product suite scales to help any size enterprise cope with exploding data requirements, now and into the future. SAND/DNA Access allows for retaining all potentially relevant data in a tiny footprint while providing instant access to just what's required. SAND/DNA Analytics allows for complex what-if analysis to meet any planned and unplanned business need. Sharing SAND's patented "ask-anything" DNA, together they provide a just-in-time approach to data management with unparalleled productivity and cost-effectiveness.

SAND/DNA solutions include SAP-certified information management, CRM analytics, and specialized applications for government, healthcare, financial services, telecommunications, retail, transportation, and other business sectors. For more information, visit www.sand.com

SAND Technology has offices in the United States, Canada, the United Kingdom and Central Europe.

www.sand.com
+1 877--468-2538 +44 (0)1276 804604 +49 (0)40 32005 836