SANMAN- Management Software for Hyperscale SAN based storage system

Urvashi Karnani Gaur
Computer Division
BARC
Mumbai, India
urvashi@barc.gov.in

D. D. Sonvane
Computer Division
BARC
Mumbai, India
sonvane@barc.gov.in

Vaibhav Kumar
Computer Division
BARC
Mumbai, India
kvaibhav@barc.gov.in

Rajesh Kalmady
Computer Division
BARC
Mumbai, India
rajesh@barc.gov.in

***Abstract*— SANMAN – Storage Area Network Management Software is an in-house developed Linux based software to manage, monitor and configure SAN based high capacity storage systems which are built over off the shelf commodity hardware. SANMAN is designed and developed using open source software components providing features of storage virtualization, volume management, access control and snapshot offered by all commercial proprietary SAN solutions. To ensure business continuity and disaster recovery in storage systems deployed on non-commercial hardware along with SANMAN, a remote replication module is developed and integrated with the software. SANMAN facilitates software defined orchestration and monitoring of storage resources. This paper discusses the implementation of remote replication module in detail which is developed using DRBD-9 (Distributed Replicated Block Device) software facilitating remote replication and automatic failover.**

Keywords*— *SAN, Storage virtualization, iSCSI, LVM, DRBD

I. INTRODUCTION

Storage technology has evolved manifolds to serve the ever-increasing requirements of various enterprises, industries and R&D institutes. To assuage the problem of storing variety and huge volume of data a lot of products, solutions and services are available. You can either "buy it"- a black box solution, rely on cloud-based services "lease it" or "build it". Various architectures for storing data have been developed and are being used in the industry, SAN based storage system is one of them. SAN based systems are usually manufactured as a commercial appliance consisting of proprietary hardware and management software. In Bhabha Atomic Research Centre (BARC), instead of purchasing these high-priced appliances, we build a complete in-house solution. The hardware is based on standard server platform using off-the-shelf commodity components and the storage management software named "SANMAN" is developed based on open source software components augmented by in-house software development. The goal is to build centralized, hyperscale mass storage systems for diversified applications; providing high availability, centralized backup, manageability and disaster recovery. The intended applications cater to the storage requirements for High Performance Computing (HPC), Private Cloud Services, Email services, Security Surveillance Systems, User Data Backup, and Remote Backup for Disaster Recovery.

A. Need to Develop SANMAN

SAN based storage systems are manufactured and supplied as a black box appliance combining hardware and software to manage the SAN hardware. Usually these high-end technologies are not available to BARC. Hence, we build complete in-house solution for high capacity and modular SAN Storage systems. Our earlier deployments consisted of off the shelf commodity hardware components, but the software was vendor specific. Proprietary software had many drawbacks like

- Interoperability: The commercial SAN Software are bound with specific SAN hardware and not interoperable, thus, making management of large systems very cumbersome.
- Vendor support is always required for debugging and troubleshooting operational issues
- Data Security: Getting vendor support required sending logs which are encrypted posing data security issues
- Licensing: Software license is based on capacity and not number of clients thereby increasing the price exorbitantly.

Hence to avoid any vendor lock in and dependency we have developed our own software to manage the SAN based system called as "SANMAN".

SANMAN is a storage virtualization and management software designed and developed for High capacity, Storage Area Network based storage systems. It provides a unified approach for the management of storage devices and servers in hyper scale storage infrastructure environments, making it easier for administrators to integrate scalable storage solutions at the data centre. The first version of SANMAN supported the

features associated with storage virtualization using Logical Volume Manager -LVM and RAID, access control and authorization of storage volumes, Storage Target Management via iSCSI, report generation and LV Snapshots and caching. To keep pace with the proprietary solutions to provide high availability, we have developed remote replication module using open source software -Distributed Replicated Block Device (DRBD) for centralized SAN based storage systems.

II. Storage Infrastructure

A Hyperscale storage environment should ensure scalability, facilitate key features like performance, fault tolerance, and avoid vendor lock-in and lower costs. The SAN based storage infrastructure developed by us is based on a modular architecture comprising of various off-the-shelf commodity components. Storage target subsystem consists of Storage Expansion Units – Just Bunch Of Disks (JBODs) encompassing storage devices like Hard Disk Drives and Solid State Drives. Storage Controllers are standard servers, having RAID Adapters with external Serial Attached SCSI (SAS) interfaces for backend storage connectivity to the JBODs. A set of file servers to access the storage from the Storage Target Subsystem over the storage network to provide NAS based services to clients. Interconnect comprising of full duplex Serial Attached SCSI to attach storage controller and JBODs. Storage Area Network is implemented by means of Infiniband and Ethernet network. For block level data access, iSCSI (Internet SCSI) over Ethernet and iSER (iSCSI Extensions for RDMA) / SRP (SCSI RDMA Protocol) over Infiniband network have been used. The storage system is managed by SANMAN software that provides a unified and simplified web-based interface to storage administrators for storage virtualization, configuring and managing distributed storage resources and SAN infrastructure required for various applications.

III. Storage Virtualization Technologies

Storage virtualization is a technique to abstract the physical storage hardware resources on storage area networks, pool them into what appears to be a single centrally manageable storage device having large storage capacity in range of Petabytes [1]. We have implemented a combination of RAID-Redundant array of Independent Disks and LVM- Logical Volume Manager technologies to provide beneficiary characteristics of virtualization such as flexible file system resizing, snapshots, scalability etc. RAID allows us to provide logical units (LUNs) as storage units and improve reliability of SAN based storage by providing protection against failures of physical drives. To achieve higher reliability, we implement RAID 6 on Physical RAID volumes on JBODs. LVM is a device mapper target that provides logical volume management of physical disks for Linux Kernels [2]. Volume groups are created over the RAID volumes and then logical volumes are made available to intended clients for storage. Using this approach, the addition and deletion of storage is done without affecting the client application, data migration is simplified and management of storage becomes easier.

IV. SANMAN Architecture

SANMAN - Storage Area Network Management Software is a Linux based storage management software used for managing centralized data storage servers for SAN based storage system. It provides a unified approach for the management of storage devices and servers in hyper scale infrastructure environments, making it easier for administrators to integrate scalable storage solutions at the data centre. SANMAN is the heart of the SAN based centralized storage system. It provides storage management features such as access control, volume management, snapshots (point-in-time copies), and remote replication. The SAN Storage Management Software developed by us provides iSCSI target functionality using Linux IO target subsystem [3], redundancy through RAID and replication feature through DRBD version 9 that is freely available as open source. SANMAN software consist of various independent and pluggable modules that helps storage system administrators to configure, manage and monitor SAN based storage systems using a web-based GUI [4]. Fig 1 illustrates the various modules that have been designed and developed. SANMAN is modular software consisting of various modules to implement specific features required by storage administrators. RAID manager is implemented using Command Line Interface MEGACLI [5] provided by the RAID Adapter manufacturer.

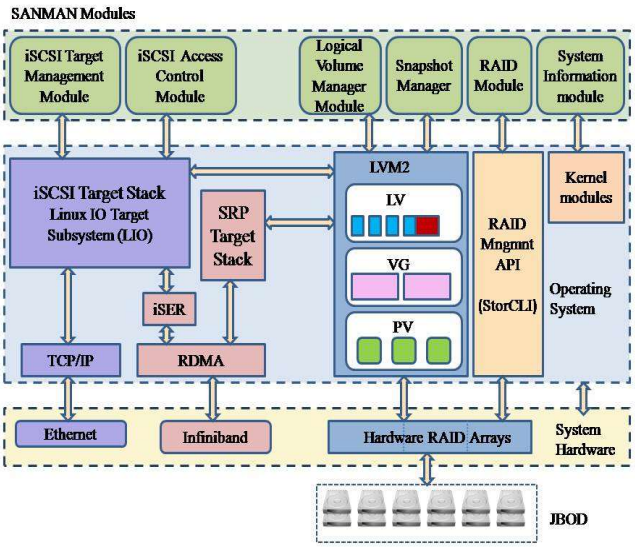

Fig. 1. SANMAN Architecture

The RAID volumes are accessed as LUNs and treated as physical storage devices. To provide block level access to storage devices, iSCSI protocol is used that carries SCSI commands over TCP/IP network. Linux IO target subsystem (LIO) is used to implement iSCSI target and initiator stack. Caching and Snapshot Manager Module uses Linux LVM 2 engine in the backend to implement the snapshot functionality. Caching module facilitates assignment of SSDs / NVMe as

cache for hard disk based logical volumes to improve performance. Snapshot module facilitates creation and scheduling of LV snapshots.

V. Remote Replication in SANMAN using DRBD

Fault tolerance is capability of a system to continue operating properly even in case of failure of some of its components to deliver uninterruptable service. To achieve fault tolerance by replication is a common practice to improve availability. Replication of data helps in removing a single point of failure using the concept of redundancy. Choosing a right data replication strategy is very tricky as a balance has to be maintained between I/O performance and data reliability. In SANMAN, storage space is provided as iSCSI Targets built over Logical Volumes (LVs). To ensure high availability of data present in our storage system, replication module has been designed and developed. Keeping in mind the infrastructure and storage service requirements, the module is built over DRBD (Distributed Replicated Block Device) software component that facilitates the replication of storage systems by networked mirroring.

A. Distributed Replicated Block Device

DRBD- Distributed Replicated Block Device is software based, replicated solution mirroring the content of block devices between hosts [6]. It is implemented as a kernel module that constitutes a virtual block device and comes with a set of administration tool to communicate with the kernel modules in order to configure and administer DRBD resources. Each DRBD device corresponds to a volume in a resource. With DRBD 9, each resource can be defined on multiple hosts. Each resource may be classified as Primary or Secondary. DRBD provides three replication modes – (i) *Protocol A*- Asynchronous replication mode which is often used in long distance replication scenario. (ii) *Protocol B*- Memory / Semi Synchronous replication mode and (iii) *Protocol C* – Synchronous replication mode where loss data due to failure of data node is minimum. The choice of replication protocol influences two factors of deployment: latency and throughput [6].

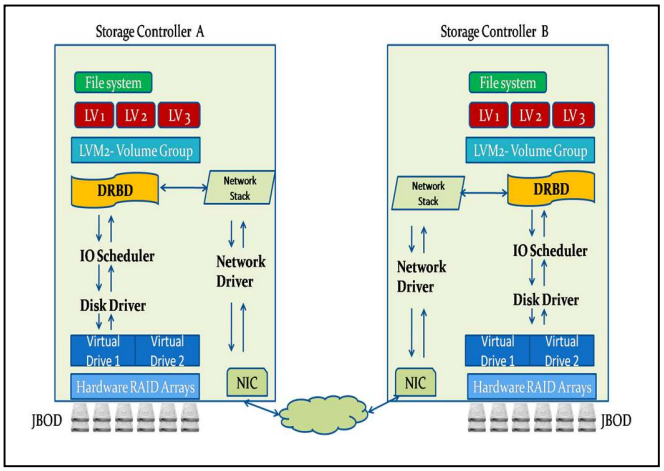

Fig. 2. Replication Module Architecture based on DRBD

B. Replication Module Implementation Details

SANMAN software runs on the storage controller servers having RAID cards connecting the storage target subsystem consisting of JBODs. Fig 2 illustrates the Architecture of replication module. DRBD kernel module installed in storage controller servers is responsible for mirroring the content of block devices (logical volumes in our case) between two servers. In DRBD, resource refers to a replicated data set having a resource name, volume, device name (/dev/drbd*X*), where *X* is the device minor number and a role. If a DRBD device is in primary role, it can be used unrestrictedly for read / write operations. A secondary device receives all updates from peer node devices.

In our implementation, DRBD runs in "single primary mode" where a resource, at a given point of time is in primary role only in one cluster member. While configuring replication of block device, Synchronous or Asynchronous replication protocol can be chosen. Synchronous replication guarantees that write operations are confirmed only after writing data to local and remote disk whereas in asynchronous replication, local write operations on primary node are completed as soon as local disk write has finished and the replication packet has been placed in local TCP send buffer [6][7]. To minimize the probability of data loss, in SANMAN synchronous replication mode is chosen by default. As DRBD supports multiple network transports TCP and RDMA our implementation also supports both the network transports for storage resources. Replication module of SANMAN enables the storage administrator to create resources (devices) for remote replication on servers. A typical example of resource configuration file created by SANMAN web interface is shown.

```
resource r1 {
    on controllerA.barc.gov.in {
        volume 0 {
                device      /dev/drbd2;
                disk        /dev/sde2;
                flexible-meta-disk   internal;
               }
          address     10.*.*.*:7791; (Private IP Address , Port)
                    }
    on controllerB.barc.gov.in {
        volume 0 {
                device      /dev/drbd2;
                disk        /dev/sdg2;
               flexible-meta-disk   internal;
               }
          address     10.*.*.*:7791; (Private IP Address, Port)
                    }
    }
```

It can be observed in the above configuration that a RAID volume attached as device “sde” at primary node controller A has been replicated at device “sdg” at secondary node controller B.

SANMAN internally uses the services of DRBD manager “drdbmanage” to configure and administer DRBD resources. It also provides a web-based frontend to DRBD-utils program suite “dsrbdsetup” and “drbdadm” to create, dump, restore and modify DRBD metadata structures directly. The Admin can perform various operations like creating a new resource, attach or detach devices to resource; monitor the status of primary and secondary resources etc.

VI. Conclusion

In traditional storage architectures, storage devices were managed and deployed manually, but with SAN based storage architectures and virtualization technologies, provisioning and configuration of storage has reduced the tedious task that burdened the storage administrators. We have developed a complete in–house solution for High capacity based centralized SAN based storage solutions providing all the vital features like iSCSI volume management, RAID management Snapshots, Caching etc given by any commercial appliance. Our solution is based on off the shelf commodity hardware and SANMAN is developed using open source software components hence, it is not bound with any storage hardware. This software can be used for storage virtualization use cases like RAID and LVM management. It also supports iSCSI-based storage target management that focus on how administrators need to allot storage space for various services in a data centre. Fault tolerance and high availability are the fundamental features expected by storage systems for which a solution based on latest freely available version 9 of DRBD has been implemented.

We have found that SANMAN software solution works on off the shelf commodity hardware and offers essential functionalities that simplify the way SAN based storage can be allocated, monitored, and managed by means of a web-based interface. Replication implemented using Distributed Replicated Block Device (DRBD) works in real time. The applications need not be modified and need not be aware that data is stored on multiple hosts. It was also observed that DRBD throughput is limited by the available network bandwidth. DRBD also provides various optimizations and tuning parameters like max-buffer size, send buffer size which may be tweaked depending upon the applications and hardware setup.

References

[1] Frank Bunn, Nik Simpson, Robert Peglar, Gene Nagle : SNIA Technical Tutorial Booklet

[2] Chapter LVM https://access.redhat.com/documentation.

[3] John L. Hufferd : Book Title ISCSI: The Universal Storage Connection

[4] D Sonvane, Urvashi Karnani et al.: Design and Development of High Capacity Modular SAN Storage System, DAE-VIE 2016

[5] Introduction to LSI MegaCLI Utility https://www.cisco.com

[6] DRBD 9 User guide : https://docs.lindit.com/docs/users-guide9.0

[7] Pedro Pla : Drbd in a heartbeat, Linux Journal, Vol 2006, No. 149.

[8] Internet small computer system interface (iSCSI) protocol (consolidated) in Internet Requests for Comments RFC Editor RFC 7143, April 2014.