International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Design of Intelligent Storage Systems

Vishwajit Vithoba Gouda[1], Dr. Minal Moharir[2]

[1]Dept. of Computer Science and Engineering, RVCE, Bengaluru, India
[2]Assosiatie professor, Dept. of Computer Science and Engineering, RVCE, Bengaluru, India

---

Abstract - *In the age of increasing amount of data to be stored for medium, large scale and enterprise level storage, it is important to maintain, monitor and manage the data stored. These can be done without human intervention by employing appropriate methods for making the systems intelligent. One can make the system smart by modifying hardware, firmware and software. Hardware and firmware methods use modifications in network systems to make the system smart whereas, software methods use efficient algorithms and machine learning. This is a study on design approach for such a storage at enterprise level.*

Key Words: Intelligent Storage, Adaptive Storage, Intelligent Network Disk, Storage Network Architecture, Software approach, Hardware Approach

1. INTRODUCTION

In this decade, the demand for intelligent storage has seen an increase. This requires the development of storage systems or arrays that can provide storage-space in millions of GBs for mission critical and non-critical storage. Major features to be included are security, backup and restore, fault tolerance and recovery, manageability of large volume of data, high speed and easy access of data and fast response to customers about the status of their data and array allocated for that customer.

There are several challenges in achieving a high degree of intelligence in the storage system. This paper discusses those challenges and recommends solution for the same, hence, gives design approaches for intelligent storage systems. This paper is structured into sections. Section 2 describes implementation of large amount of storage. Section 3 gives a brief about intelligent storage and its challenges. Section 4 and 5 mentions hardware-firmware approach and software based solution for the design. Section 6 concludes the paper.

2. LARGE SCALE STORAGE

Large scale and enterprise level storage deals with million TBs of data stored in traditional devices like disks, hard drives or tapes, which are here on, referred to as disks. When such a large number of disks are to be organized, their topology should be decided first as they consume large physical space and for ease of accessibility. In order to achieve this, network is required. Considering every disk as a node, network has a server that carries computations, operations and execution of algorithms, these are connected with each other like nodes in LAN. This leads to employment of Storage Network Architecture in storage. There are 3 types, Direct Attached Storage (DAS), Network Attached Storage (NAS) and Storage Area Network (SAN).

DAS is an architecture for which the storage is privately attached to the servers i.e. cannot be shared, scaling is difficult, expensive to implement and managing of data and system is complex. This type of storage best suits for small storage applications like maintaining records in schools where computer can be connected to a server.

NAS is a file-level storage of computer data, coupled with IP network to provide data access to a client group. NAS eliminates the responsibility of file serving from other servers on the network. They provide access to files using network file sharing protocols. They allow for easy appliance, provide clustered file-system. This type of storage does not suit for enterprise level storage due to file level storage which is expensive in terms of CPU usage and has lower disk performance.

SAN is an autonomous storage network, free from rest of the computer network. It is a dedicated network that provides access to consolidated, block level data storage. It provides I/O connectivity between hosts and storage devices. SAN devices which are hubs, switches, servers and storage devices constitute a storage resource environment. SAN can be a Fiber Channel network that uses a network of Fiber Channel connectivity devices such as FC Switches and Directors, FCP (serial SCSI-3 over Fiber Channel) for transport or an IP network that uses standard LAN infrastructure with Ethernet switches, for transport, iSCSI (serial SCSI-3 over IP).

Choice of network architecture depends on scale, budget, I/O requirements and availability. SAN being best suited for enterprise storages, modifications in SAN will make the system smart. The choice of disks also play an important role in enhancing performance and accessibility. Types of disks can be logical drives, Redundant Array of Independent Disk (RAID) with virtual levels and Solid State Drives. SAN, when used, they have three components: Adapters, Fabric and Storage; firmware, hardware and software modifications can be made to achieve intelligent storage.

There are other modes of storage which supports implementation of intelligence into the systems. Some of them are: Object Based Storage, Storage Grid, Active Storage, Adaptive Storage. [1]

3. INTELLIGENT STORAGE

It is a storage system that is smart enough to carry out below-listed operations without human intervention. Following are the operations:

a. File system- It used for storing the data, which depends on type of data to be stored, accessibility and the OS used for the system. They can be armed with attributes to make access easier.
b. Fault tolerance and repair- It is crucial to ensure no data loss or corruption. In case of any such events, repairs shall be carried out autonomously using redundant data, snapshots etc.
c. Backup and restore- It is required for ensuring zero customer data loss and restoring the data in case of any hazards to the array. Backup algorithms shall be such that they extract small amount of data from a disk and restore the data in the disk with that backed up data. Rate of backup and restore shall be higher.
d. Context awareness- It is a requirement for the storage array to allocate or re-allocate the data as per the type of data and operations like data distribution, request accesses shall be done autonomously.
e. Security of data, compatibility of hardware, firmware and software must be ensured. Records related to the storage about access, updates, requests have to be logged autonomously in case of any such events.
f. Along with these operations, data retrieval, de-duplication can also be included. An intelligent system need not have all of these operations automated. These features shall not compromise with the system performance by events like inconsistent power supply, overheating of disks, controllers. Storage scheduling also plays an important role in making the systems smart. These methods include: queue-based, table-based, object-based, storage agent-based, role based, time policy, space planning. [1]

To achieve compatibility for implementation methods, high level architecture of storage system can be represented using disks, adapters, network channels, server and UI. Figure 1 shows the direction of control flow during storage of data, opposite of which is followed during data retrieval. Hardware modifications in adapters or controllers and channels, and software changes in Servers can make the system intelligent.

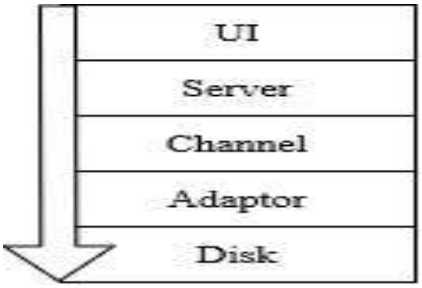

Fig – 1: Control Flow during data storage

4. HARDWARE APPROACH

Changes at adapters or controllers are the hardware changes that include component upgradation and network modifications; firmware changes, which employ efficient algorithms to achieve intelligence.

Component upgradation at disk level is using ATA, SAS or SATA disks. These disks allow virtualization of storage, hence, making access easy. Selection of OS plays a vital role in selection of compatible file systems, a few of which allow Virtual File System (VFS), which provide interface between upper layer and disk level. File system is an approach where there is global file system throughout the network, where a cluster of servers are connected P2P and every server controls its fabric. [2] This requires modifications in network LAN of every server. This approach provides data backup on a different network and availability of storage arrays connected to different servers. If adapters are provided with intelligence to route the access controls, requests, data flow, backup and high availability can be achieved in the same fabric, which reduces latency, hence, access time.

Further, by a series of algorithms and routines such as access, fault tolerance, migration, load-balancing, concurrency control, distribution management and monitoring can be implemented on adapters to carry smart routines. [3][5] The above mentioned routines can be implemented at network level to deal with big data. One such efficient approach is to use Intelligent Network Disks (IND). Here, a disk, controller, memory, cache and Network Interface Card are used to represent one complete IND. [5] Figure 2 represents the logic structure diagram of an IND.

INDs work like a PC connected to internet, e.g. produce automated response at request hence, it is also called active storage. INDs contain RAM, flash, I/O adapters and processors to act as a PC. Data storage, transmission and reception related computations are done within the IND. They improve intelligent storage management by realizing the storage virtualization of heterogeneous storage pool, automatic storage management measures and life cycles of data online, near line and off-line management non-disruptive migration, data protection and continuous backup. INDs have their own OS, MMUs and file systems to

manage the data and flow of data. One method in INDs to include new users is by fork operations based on array to be used, which proves that they need high computation power at the controller level. [5]

All these modifications at hardware level are expensive in terms of cost, complexity and maintenance of the storage systems. Giving the adapters, freedom to only route the data and taking all the computation at server level reduces the hardware issues, making the hardware and firmware 'lighter' than before.

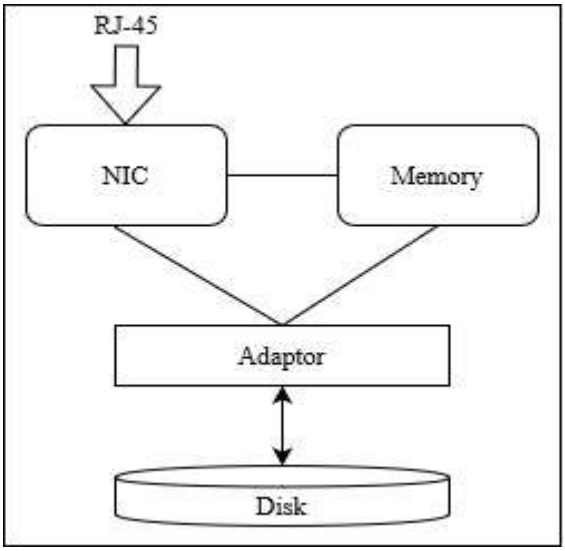

Fig – 2: Logic Structure Diagram of IND

5. SOFTWARE APPROACH

This lighter approach is via software changes. Keeping all the hardware components as it is, writing algorithms for the system to be intelligent requires higher RAM and clock speed within the server. These algorithms can be a few lines of code or several software running at the server to achieve high degree of intelligence in the system. [4] Along with software, corresponding databases may be required in order to manage temporary data if generated in large amount.

Traditional methods use manual checking of data and disk statuses when a request is obtained. Algorithms can be written to make the checking process automatic when requests are obtained and scheduled. To ensure security, backup and restore, algorithms can work on the data without tampering them, and obtain the required hashes, snapshots etc. Algorithms can also represent and replicate the flow of data and control in the arrays, file system and associated operations are the examples for the same. This allows the system to generalize the type of data, hence, be context-aware and help in non-disruptive migrations.

All the operations can automated using software. There is still a possibility to reduce the computation time and resources by eliminating redundant routines on a large scale and long term. One such approach is using machine learning. Time on operations like backup, restore and data security cannot be reduced as there is no pattern in such data. Patterns like mapping between type of data and allocated memory, type of messages and images can be stored in a database. When a new file, message or data arrives algorithms can directly check for the presence of patterns from previous arrivals which are patterns stored in the database. If there is a pattern match, same actions are taken; else, add the new pattern to the database along with new actions to be taken. On a long run, machine learning can help improve the performance along with being intelligent.

6. CONCLUSION

Intelligent storage systems can be designed using hardware or software approaches. Hardware approach demands upgradation of components and underlying firmware with same computation power in the server. On the other hand, software approach demands high computation at the server without any changes required at component level. Machine learning can make the system more intelligent with above-normal time and space complexities.

REFERENCES

[1] Zeng, W., Zhao, Y., Wang, W. and Song, W., 2009, March. Intelligent storage architecture and key technologies research. In *2009 WRI World Congress on Computer Science and Information Engineering* (Vol. 5, pp. 778-782). IEEE.
[2] Du, D.H., 2005, August. Intelligent storage for information retrieval. In *International Conference on Next Generation Web Services Practices (NWeSP'05)* (pp. 7-pp). IEEE.
[3] Yang, X. and Zhao, Y., 2011, July. Design of intelligent network disk storage system. In *2011 International Conference on Multimedia Technology* (pp. 4905-4908). IEEE.
[4] Tsagaris, A., 2014, September. Optimal control of storage system based on intelligent techniques. In *2014 19th International Conference on Methods and Models in Automation and Robotics (MMAR)* (pp. 264-269). IEEE.
[5] Li, H., Lu, C. and Sun, S., 2014. Intelligent network storage on the big data. In *MFI* (pp. 1-6).
[6] Meyer, H., Bruder, I., Finger, A. and Heuer, A., 2015, January. Building digital archives: Design decisions: A best practice example. In *2015 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services* (pp. 59-64). IEEE.