Proceedings of International Conference on Scientific Research in Natural and Social Sciences
Hosted online from Toronto, Canada.
Date: 5th January, 2024
ISSN: 2835-5326 Website: econferenceseries.com

ANALYSIS OF CAUSES OF FAILURES IN INFORMATION STORAGE SYSTEMS OF DISTRIBUTED RELIGIOUS ORGANIZATIONS

Dadamuhamedov Alimjon Irgashevich
International Islamic Academy of Uzbekistan,
Senior Teacher of The "Department of Modern Ict".

Anvarjon Abdujabborovich Maxkamov
International Islamic Academy of Uzbekistan,
Senior teacher of the "Department of Modern ICT"

ABSTRACT
The article is devoted to the development of a probabilistic approach to assessing the fault tolerance of distributed data storage models on the servers of major religious organizations. A review of existing software tools for monitoring the health of data storage systems in religious organizations was carried out, the approaches used for data collection, processing and storage were identified, the tools used to identify failures were described, and the features ro is listed.

Key words: network protocols, multipath redundancy, anomaly detection, machine learning, fault detection, data storage, fault localization.

INTRODUCTION
There are four alternative data placement strategies in the system: centralized, isolated (fragmented), placement with full replication, and selective replication. Since the first computers were created, the problem of data storage has been a very important issue in computer technology. With the development of computing technology, more capacious, efficient and reliable methods of data storage were required, and research in this field has never stopped.

The servers of the religious organization work on the basis of modern operating systems and software. Methods of building fault-tolerant storage systems based on multiple disk drives appeared about 30 years ago. These were the so-called RAID arrays - a number of disks combined into one memory [1, 2]. Increasing the reliability of such storage was achieved by storing redundant data (checksums) on one or more additional disks. If one of the drives failed, the data could be recovered using this redundant data. The problem with increasing the number of disks is that the probability of failure of one of the many disks per unit of time is directly proportional to their number. If recovery from a previous failure is not completed by this time, data may be lost.

To achieve optimal reliability in a data warehouse that combines many disks, completely different technical solutions are required. This solution is to partition the stored data. Here, each fragment is stored as a set of disk blocks on different disks in the storage system with the necessary redundancy to ensure reliability. In order to increase reliability, a RAID array is achieved by distributing the damaged parts across all the disks in such a system if one of the disks fails. This means that recovery time decreases as the number of disks increases, effectively compensating for the possibility of any one disk failure as the number increases. The first industrial system built on this principle was the Google cluster file system [3]. Designed for internal use by Google applications. Similar ideas formed the basis of the Azure storage system developed by Microsoft [4]. Also open source CEPH system [5].

METOD

Methods such as scientific description, comparative, and component analysis were used to illuminate the research topic.

MAIN PART

Reliable storage of large amounts of information is one of the main requirements of modern information infrastructure. The volume of data produced by mankind doubles every two years. An increasing amount of this data is stored in "cloud" storage rather than on user devices. The rapid growth of computer computing power surpasses the needs of individual users and creates the necessary conditions for the development of cloud services. Contributes to the integration of information infrastructure [6].

Provides higher reliability than dedicated servers due to the ability to migrate data in case of systems and equipment failure. Data storage systems currently operate with one or more host computers, disk, network infrastructure, and complex software and hardware products in addition to the storage media itself.

The interaction of disparate software and hardware components under high load conditions can cause a storage system to fail, even if no individual components fail. For example, in [6, 7], the interaction of the components occurs before the failure of the process. Failures in the storage system make up 9% - 11% of the total number.

Distributed data storage is a system with a large number of storage devices connected via a network and the infrastructure that supports their operation. The increase in the size of modern distributed storage systems, which includes thousands of data storage devices, leads to an increase in the failure rate of its components [12]. 78% of all server failures are related to hard drives, 5% to RAID controllers (Rapid Array of Expensive Disk - RAID), 3% to memory, and the remaining 14% to other factors. Because distributed systems are built around servers and hard drives, they are prone to failures that can break applications that depend on them [11].

The developers of the Google distributed data storage system call the frequent failure of system components one of the most difficult problems. For example, in 2009, Facebook temporarily lost more than 10% of its stored photos due to hard drive failures. If the disk fails, it will not be possible to access the stored data, which is unacceptable in modern conditions (Table 1).

Table 1 - Causes of server failure in distributed systems

№	Exit status	%
1	Failure status of hard drives	78%
2	A memory	3%
3	Raid is a state of failure of controllers	5%
4	Other components	14%

Outages in religious organizations can cause temporary or permanent data unavailability. Data unavailability due to network outages, node failures, power outages, or the automatic recovery process is temporary and does not result in permanent data loss. Data unavailability due to hard drive failure or data corruption

ISSN: 2835-5326 Website: econferenceseries.com

results in permanent data loss. Increased reliability reduces the impact of failures that cause persistent data unavailability [9].

Cloud data centers use various mechanisms to increase the reliability of the storage system. The impact of hardware failures is reduced by RAID arrays, swappable drives, and Error Correction Code Memory (ECC RAM). RAID arrays are logical units consisting of multiple disks that store data in a linear, mirrored, and parity fashion [8].

Replaceable drivers allow administrators to replace drivers that have failed or are expected to fail while the system is running. ECC RAM is used to detect and correct single-bit errors by associating a parity bit with each binary code. In case of mains outages and power outages, a backup network and an additional power source are used, respectively.

Temporary unavailability of data on servers can be a significant risk. These failures require extensive changes to ensure servers and data are properly supported, operational, and secure. In the event of a failure that causes the data to be temporarily unavailable, the unavailable data may be considered lost. Later, in order to restore access to them, you can apply methods to deal with failures that lead to permanent data inaccessibility.

The consequences of any failure, regardless of the problems that caused them, can be eliminated using various methods of introducing data redundancy. The implementation of forward and reverse data conversion blocks depends on the methods chosen to ensure fault tolerance. Erasure codes, replication, and Resilient Distributed Dataset (RDD) are the most important ways to ensure fault tolerance in distributed storage systems [7].

The most common methods of introducing redundancy are replication and deletion codes. Replication is a simple data backup mechanism. The same data is copied and stored in several locations of the storage system. If the requested data is not available on one disk, it is provided from the next available disk.

Deletion codes are a more sophisticated mechanism for introducing data redundancy. Parity data is created and stored along with the original data, so if the requested data does not exist, it can be reconstructed from the parity data.

ISSN: 2835-5326 Website: econferenceseries.com

Storage costs for erasure codes are much lower than for replication, so it reduces the need for data storage hardware and provides significant cost and energy savings in data centers. However, data recovery after a failure is subject to high reconstruction costs and network traffic.

CONCLUSION

In the process of analyzing current trends, diagnostics of the data storage system, enterprise-level data, and the most common software management and monitoring systems are considered. Storage, data analysis and software are general purpose tools. Almost all of the current software reviewed offer some automatic functionality, allowing for automated or administrator-controlled troubleshooting of varying complexity. In addition, we can conclude that general-purpose systems that provide unified functionality for data collection, processing, and visualization are becoming more and more widespread. The advantage of such systems is that they use a large number of different anomaly detection algorithms, including those included in the field of cloud computing, which makes it possible to save computing resources of storage systems.

REFERENCES

1. D. A. Patterson, G. A. Gibson, R. H. Katz, A case for redundant arrays of inexpensive disks (RAID). ACM SIGMOD Record. 1988. Vol 17(3). pp. 109-116.
2. P. M. Chen, E. K. Lee, G. A. Gibson, R. H. Katz, D. A. Patterson. RAID: HighPerformance, Reliable Secondary Storage. ACM Computing Surveys. 1994. Vol 26(2). pp. 145-185.
3. S. Ghemawat, H. Gobioff, S.-T. Leung. The Google File System. ACM SOSP. 2003. 15 p.
4. B. Calder, J. Wang, A. Ogus, N. Nilakantan, A. Skjolsvold, S. McKelvie, Y. Xu, S. Srivastav, J. Wu, H. Simitci, J. Haridas, C. Uddaraju, H. Khatri, A. Edwards, V. Bedekar, 114 S. Mainali, R. Abbasi, A. Agarwal, M. F. ul Haq, M. I. ul Haq, D. Bhardwaj, S. Dayanand, A. Adusumilli, M. McNett, S. Sankaran, K. Manivannan, L. Rigas. Windows Azure Storage: A Highly Available Cloud Storage Service with Strong Consistency. Proceeding SOSP. 2011. pp. 143-157.

5. S. A. Weil, S. A. Brandt, E. L. Miller, D. E. Long, C. Maltzahn. Ceph: a scalable, highperformance distributed file system. Proceedings OSDI. 2006. pp. 307-320

6. Hewlett Packard Enterprise Development LP. Can Machine Learning Prevent Application Downtime? // URL: https://cdm-cdn.nimblestorage.com/2017/08/23090218/Can-Machine-LearningWhite-Paper-1708-FINAL-print.pdf).

7. Wang D. Artificial intelligence makes flash storage predictive // URL: https://www.hpe.com /us/en/insights/articles/artificial-intelligence-makesflash-storage-predict

8. Higley L. Storage analytics: Can we put any more lipstick on that pig? Available: https://cloud. kapostcontent.net/pub/3da21605-fc17-4712-991a-1c49dc77b871/mfx131e-pc-mon-130-higleyl.pdf?kui=xxP HjAO870Nzv0HTEjftEw (Accessed: 10.04.2019).

9. Gopisetty S. Evolution of storage management: Transforming raw data into information. IBM Journal of Research and Development, 2008, Vol. 52, No 4.5, Pp. 341—352.

10. Успенский М.Б. Обзор подходов к обнаружению сбоев в системах хранения данных // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникац

11. Дадамухамедов, А. И. (2017). Развитие национальной сети и корпоративной сети (на примере сети IX). Актуальные научные исследования в современном мире, (3-2), 133-137.

12. Alimjon D, Al-Dabbagh KA, Shnishil AT, Altememy HA, Jawad IA, Jabbar AM, et al. Investigation of the Effects of Cognitive Rehabilitation Training on the Memory Deficits of Adults with Attention Deficit Hyperactivity Disorder: Effects of cognitive rehabilitation training on the memory deficits of ADHD. Int J Body Mind Cult. 10(4).