2015 The 5th International Workshop on Computer Science and Engineering
doi: 10.18178/wcse.2015.04.122

Erasure Correcting Code to Recover Failure Stripe Write Data technique in RAID Level 6

Ghulam Muhammad Shaikh[1,a], Quanxin Zhang[1,b], Yu-an Tan[1,c]

[1]Beijing Engineering Research Center of Massive Language Information Processing and Cloud Computing Application, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081, China

[a]gm.shaikh1@gmail.com; [b]zhangqx@bit.edu.cn; [c]victortan@yeah.net

Keywords: RAID, Erasure Correcting Code, Recover Failure Data, Reliability, Performance

Abstract. A Data storage system has grown to the point where failures are expected, where the data redundancy schemes and codes to protect against loss of data on a single or large storage systems. In a Redundant Array of Independent Disk system there are many methods to protecting data but in this paper we use erasure correcting code technique for recover failure data when fault happens. Many techniques are used from decades but here the value of data in a system and parity values are striped across disk drives. Redundant Array of Independent Disk systems are typically used to protect data which is stored in multiple hard disks. In this paper, we tell you how to recover data in a Redundant Array of Independent Disk in Level 6 when any disaster occurs. Moreover, erasure correcting code includes a multiple methods and coding techniques to recover data failure but here we use technique to recover a failed data from stripe write data in Redundant Array of Independent Disk Level 6.

Introduction

Redundancy array of independent disk (RAID) is arrangement of more than one hard disk drives where the system tolerates the failure of disks. From decades many RAID levels are failed to recover a failure data when it is lost on any disaster or disk failure, after research on multiple RAID levels the researchers found a Raid 5 which has some features and kind of techniques to recover failure data but not fully, might be partially or sometimes failed permanently. Recently scholars found a way to recover loss of data from failure hard disk by using different coding techniques, algorithms, and methods[1]. Raid 6 uses a double parity which can hold huge amount of data in a storage system, if failure happens by any cause like power failure or damage, but it can be recover by using erasure correcting code which we will define in next section where this level can allow for a data can be retrieved by any cost.

Redundant array of independent disk (RAID) is an efficient approach to provide a high reliability and performance on data storage with low and high cost according to implement the plan of RAID level[2].

When the data fail in a storage system, if rely on erasure correcting code which adds redundancy to the raid system and tolerate failures when happen, there are many levels used in RAID with different approach and research methodology, such as level-1 of Raid is uses byte of data to be stored on two disks and so on, but we do not have strong authentication to recover data when fault happen[3].

Although multiple methods and codes to use on performing, efficiency and tolerate data failures to the best knowledge, no prior work has focused on the recover stripe data during data written in storage system. Our aim is to focus on data recover when fault happens during stripe write in RAID system. In this paper we use technique to recover failed data in RAID storage system.

This paper organized as follows. In Section II, we summarize RAID, define erasure correcting code in Section III, Section IV present Technique to recover data, define reliability and performance in Section V, Section VI we conclude our paper.

RAID

Redundant array of Independent disk known as RAID, has many levels to store data which is introduced in the late 1980s by a group of U.C. Berkeley[4]. When a plan to construct a large system in organization where we keep the capability of securing data and which is to be execute in unique failures on the Raid architecture. The configuration offers very high fault and driven-failure tolerance. It is used for environment that needs long data rotation periods, such as archiving. One disadvantage in using RAID 6 is that each set of parities must be calculated separately. Which shows write performance implementing RAID 6 is also more expensive because of two extra drives required for parity.

Many storage organizations are grappling with the question of how to use protecting data in RAID 6. Where the stripe write data is calculated in the parity hard disk drives twice with the results stored in different block on the disks.

If RAID 6 array contains the minimum number of disks can hold half the total disk capacity for data, as well. The difference comes when you add disks in the disk array, the percentage of usable capacity will increase. If you double the number of disks from four to eight, the stripe write performance will be low as to comparison of RAID 5 which is depending on size of write and Raid system where the extra parity information that has to transfer on the back-end buses to disk results where the bandwidth of lower system.

Recovery Technique of Failure Data

There are many algorithm to use in raid technique where the worth of parity to tolerate the loss of data up to number of disk sectors at the same logic block address, that is extremely efficient with the calculation of parity and requiring XOR logic[5]. It doesn't require any additional hardware or recursive computation as with other implementations. In Raid 6 the performance is fast because it distributes parity data among all the disk drives and it has two dedicated parity known as P and Q. For a (P-2)-2-row-coloumn RDP code matrix, to retrieve P-2 lost elements, it would result in (P-2)*(P-2) reads the P-2 writes , or (P-1)*(P-2) I/Os per failed element. However, for a (P/2)-row0coloumn P-code matrix, to retrieve P/2 lost elements; it would result in $-\frac{P}{2} * (P-1) - C_{p/2}^{2}$ or $\frac{3}{8} * P^{2} - P/4$ I/O at least. This is to say, its recovery efficiency is ¾*P-1/2 per failed element. Total double failure recovery workflow of Erasure code to dedicated drives which are Row and Diagonal Parity where the data for a block location within one block stripe as similar raid level 5[6]. Diagonal parity in raid 6 is unique where the data diagonally across bits in a block stripe both parity disks are complexity independent of one another. Fig1 shows the parity is stored in RAID 6.

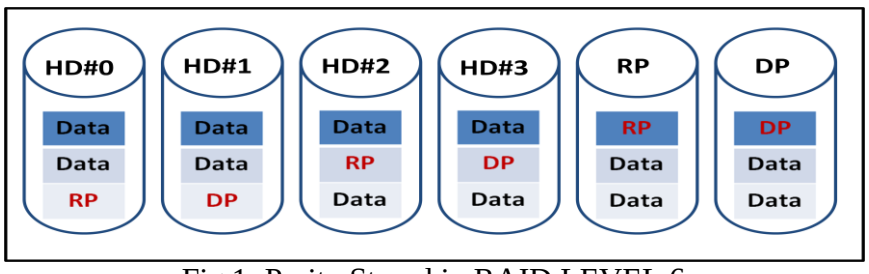

Fig.1: Parity Stored in RAID LEVEL 6

The original data must be restored from failure of any two of the K+2 devices. There are many criteria to store data when selecting erasure correcting technique, in above figure the system row and diagonal parity is rotated on different block of disks.

Fig.2 : Data is Stored in Form of Bits

In fig.2 Data is representing in bits because it is not possible to put all values in one block. One block contains many bits. Above figure we have to calculate the row parity which is to be even for each row of data, the row parity calculate from the above raid is $0 \oplus 0 \oplus 1 \oplus 1=0$. Row parity can be used to tolerate the loss of data from any single drive in a raid group.

Form the above figure we can calculate the diagonal parity which is like more difficult to visualize but it may be even or odd depending on implemented raid system. Above example, $P = 1 \oplus 0 \oplus 1=1$, so diagonal parity is odd in this stripe, the example is $1 \oplus 0 \oplus 0 \oplus 1=0$, that 0 is the stored in corresponding on another field.

In raid 6 system when two disk drives are physically inaccessible for reading a hard media error occur and when a second drive failure occurs during a single drive rebuild, the hard media error occurs from two drives on the same logical block address.

Fig.3 : Two Empty Disk Drives are Present in RAID System

In fig.3 two disk drives are empty and unreadable it means no data on these drives, if it can be constructed using the independent row and diagonal parity by using erasure correcting technique. In this system to recover the data which is mathematically proved in EVENODD research papers[7, 8]. Above fig we use XOR technique for all of diagonal and row parity bits which is odd in block stripe, therefore P is odd (P=1), so 'P' is also the parity of the disk 1, disk 2, and disk 3, when disk 2 can be calculated.

Fig.4: To Recover a First Block of HD#2 by Using Row Parity

In fig.4 using the row parity take the data from HD#1 and HD#3 and the 'unknown' data of HD#1 for e.g $P=1 = 1 \oplus \text{unknown} \oplus 1$, therefore the unknown bit must be 0. The piece of data is recovered, it is straight forward to use the row parity and recover the missing data for the empty hard disk.

Fig. 5 Recovered Data by Using Row Parity

Fig.6: Missing Data is Recovered by Using Row Parity

When the data is recovered by using row parity after that we use diagonal parity for all the missing block of hard disk in raid system. Where the row exactly one diagonal mirroring data element. We take the bits of HD#0(which is recovered bit), HD#1(second block of HD), than "unknown" from HD#2, after that P=1 from Diagonal Party (DP) which is calculate $1 \oplus 0 \oplus \text{unknown} \oplus P=1$ is shown in fig.7.

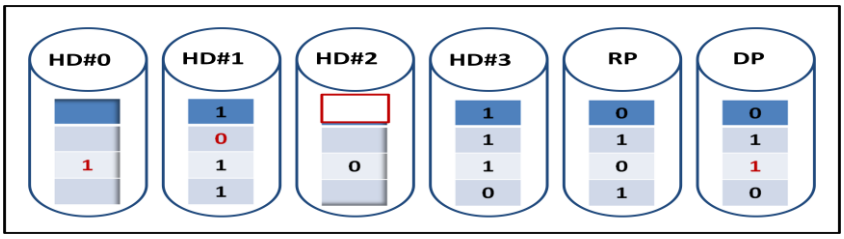

Fig.7: Recovering Data by Using Diagonal Technique

When the bock of HD#2 is recovered by using diagonal technique, there is exactly one row with one missing data element. This pattern continues until all data is recovered. The order of recovery is indicated by the catering in fig.8.

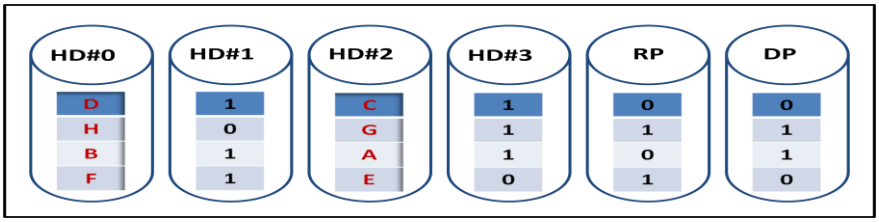

Fig. 8: The Pattern Continues until All Data Recovered

If the parity is lost then there are different techniques to use for data recovery which we will discuss in our next paper.

Reliability

Raid technology the reliability is the key criteria of selecting Raid level. Other Raid levels have good reliability but Raid level 6 has best. Reliability can be modeled through probabilistic calculations which are useful for profiling a new technology[9, 10]. However it is also important to look at the larger systems of how RAID fit is in one of many data protection mechanisms.

In this paper we compute the reliability of RAID 6 is theoretically by using some techniques and proved it, when the mean time to data loss (MTTDL) due to multiple disks failing during a rebuild. This is based on the predicted mean time to failure (MTTF) for disk drives. Different hard disk venders quote MTTF value more than 300,000 and less than 1,000,000 hours failing of multiple disks when rebuild which is depend upon the MTTF of the individual disk drives size of RAID system and the mean time for a rebuild to complete. During rebuild disk drives when loosing data which is encountering uncorrectable errors, which is depend upon hard media error at the drive level which is correctable by RAID unless two drives are rebuilding in RAID 6 level.
By using the MTTDL techniques we have to comparing the calculation for different impacts to get theoretical models which is produced by the some results like when MTTDL due to an uncorrectable error during a rebuild which is much less than the MTTDL due to multiple disk failures during a rebuild. The MTTDL due to an uncorrectable error during a rebuild which is depend on MTTF. This means an uncorrectable error during a rebuild causes data loss where often than multiple disk failures.

$$MTTDL\frac{(MTTF)^2 * (MTTF - 1)}{N\,(G-1) * (G-2)(MTTR)} \qquad (1)$$

When capacity of disk is increase the MTTDL uncorrectable error decrease which means that RAID system are larger drives more likely to encode an uncorrectable error then other RAID smaller systems. If the number of disks are increase it means more bits to read during the rebuild of the RAID system.

$$MTTF = \int_0^{\infty} e^{-9\lambda t}\ dt\frac{1}{n\lambda} \qquad (2)$$

Furthermore when the rebuild time increase, the MTTDL decrease, for computing different RAID technologies which theoretical models are useful.

Performance

When to calculate the performance of any system like RAID technology which is normally sized and designed on capacity, performance and Reliability.

For capacity purpose RAID 6 uses more disks than other RAID levels, when many applications runs on RAID 6 system there is an effect on performance and chances of fault happens, so it is very difficult to design a strong system performance to increase the workload. In RAID system the performance of disks which is depend on another effect such as controller, catching, data layout, and application or file system effects. The effect of controllers and caching to focus on the performance of the single disks RAID technology. The fully configured system of each RAID type of includes factors such as caching and controllers.
In this system the performance is much better than other raid levels with the increase of number of disks which are rotated. Read and write performance is different from other levels, because RAID 6 is similar to RAID 5 only difference is to add an extra parity hard drive, where the bandwidth and IO operations are limited.

Write performance is less than the comparison of RAID 5 which is depending on sizes of write and RAID system where the extra parity information that has to transfer on the back end buses to disk results in lower system bandwidth of the RAID system.

Conclusions

In this paper we define theoretically that how to recover a data from RAID 6 system by using error correcting code with the help of XOR technique moreover we show the performance and capacity of RAID 6 that how to calculate and implement. Furthermore, we define reliability of RAID

protection that other RAID technology where this can tolerate two failed disks. It may come at the cost of performance when compared because it has many ways to protect data when it implement backup and replication strategies to ensure that your most important data is protected in all situation.

Acknowledgements

This research work is supported by 863 program of China (No.2013AA01A212), Natural Science Foundation of China under and Shanghai Aerospace Science and Technology Fund (SAST201341).

References

[1] Guruswami,V., & Sudan,M. Improved decoding of Reed-Solomon and algebraic-geometric codes. In Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on IEEE,pp. 28-37,1998.

[2] Blaum,M., Brady,J., Bruck,J., & Menon,J. EVENODD: An efficient scheme for tolerating double disk failures in RAID architectures. Computers, IEEE Transactions on, 44(2), 192-202,1995.

[3] Niset,J., Andersen,U., & Cerf, N. Experimentally feasible quantum erasure-correcting code for continuous variables. Physical review letters, 101(13), 130503,2008.

[4] Patterson,D.A., Gibson,G., & Katz,R.H. A case for redundant arrays of inexpensive disks (RAID) (Vol. 17, No. 3, pp. 109-116). ACM,1988

[5] Wu,C., Wan,S., He,X., Cao,Q., & Xie,C. H-Code: A hybrid MDS array code to optimize partial stripe writes in RAID-6. In Parallel & Distributed Processing Symposium (IPDPS), 2011 IEEE International (pp. 782-793). IEEE,2011.

[6] Corbett,P., English,B., Goel,A., Grcanac,T., Kleiman,S., Leong,J., & Sankar,S. Row-diagonal parity for double disk failure correction. In Proceedings of the 3rd USENIX Conference on File and Storage Technologies (pp. 1-14),2004.

[7] Cassuto,Y., & Bruck,J. Cyclic lowest density MDS array codes.Information Theory, IEEE Transactions on, 55(4), 1721-1729,2009

[8] Cao,Q., Wan,S., Wu,C., & Zhan,S. An evaluation of two typical raid-6 codes on online single disk failure recovery. In Networking, Architecture and Storage (NAS), 2010 IEEE Fifth International Conference on (pp. 135-142). IEEE.2010

[9] Wu,S., Mao,B., Feng,D., & Chen,J. Availability-Aware Cache Management with Improved RAID Reconstruction Performance. InComputational Science and Engineering (CSE), 2010 IEEE 13th International Conference on IEEE,pp. 229-236,2010.

[10] Chung,C., & Hsu,H. Partial Parity Cache and Data Cache Management Method to Improve the Performance of an SSD-Based RAID.Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, 22(7), 1470-1480,2014.