ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 3, November 2017

Data Recovery and Reconstruction Using Business Intelligence Methodology Erasure Code Implementation Using Cloud Computing

B.Muthukumar, R. Dhanalakshmi, R. Aroul canessane,

Faculty of Computing, Sathyabama University, Chennai, India

Faculty of Computer Science and Engineering, Jeppiaar Eng College, Chennai, India

Faculty of Computing, Sathyabama University, Chennai, India

ABSTRACT : Cloud Computing and data storage have become very important for everyone. From an employee to executive managers, everyone uses cloud storage to store their data safely. But sometimes, the data stored on the cloud is not safe from factors like memory format, data corruption and loss. So, a new type of reconstruction has been introduced to decrease the risks of data loss. The PUSH model uses a large number of its storage nodes to share the workload equally to avoid any additional pressure. For downloading the PULL model is used. Whenever the files are needed to be downloaded, it will download from the replication server. So if main file is deleted or corrupted it downloads the file from replication server and regains the data needed. It has security alert feature, i.e., when the file is corrupted or deleted, it will send a mail alert. During the validation of file too, it will alert the mail for verification.

KEYWORDS: Data Recovery, PUSH model, Cloud Computing, Business Intelligence, Reconstruction, Nodes, Memory.

I. INTRODUCTION

Data recovery is the process of getting back all the data that has been lost due to memory format, damage or memory corruption. Recovery is usually done from all types of storage media, internal as well as external. Cloud Computing is vastly used in the present world to store data securely. Leading IT companies such as Google, Microsoft, IBM and Yahoo! have made cloud storage their forte. Cloud Computing lets the user store, access and manage data in a remote server hosted by the network.

Reconstruction is a process in which pieces of data are recovered and then put back together in order to try and construct back the old data. There is an existing PULL-type (PULL-Rep and PULL-Sur) reconstruction algorithm that is used to reconstruct the data but in this paper, a PUSH-type (PUSH-Rep and PUSH-Sur) reconstruction algorithm using Business Intelligence is being proposed. Business Intelligence is the unique feature that has been added to the model. It refers to a set of applications which are used to analyse raw data.

Erasure Code is a process in which the data is broken up into pieces, encoded with surplus data and then stored in different locations on the network. Introduction glides through the basics of data recovery, cloud computing, erasure codes, reconstruction and business intelligence. Existing System describes the methods that are still in use and why they are recommended to be replaced. Proposed System describes the model, the uniqueness of the model and why it should replace the existing model(s). Algorithm explains the working of the PUSH-based models, going through the process and its advantages. Application(s) tries to explain the use of this model in the real world. Methodology explains the working of our model along with small code strips to improve the understanding. Conclusion describes why our model needs to be used.

II. RELATED WORK

The existing PULL-type reconstruction technique has a master node, which starts the reconstruction system by sending requests to all the worker nodes that are dedicated to the reconstruction of data. But due to the existence of rebuilding

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 3, November 2017

nodes, the PULL model faces a transmission relay problem. The following elements lead to the proposal of the PUSH model for erasure-coded clustered storage:

1. Erasure-coded storages are highly cost-effective.
2. PULL-based models are not reliable because the recovery time gets impacted due to the process.
3. PULL-based reconstruction I/Os are very rarely available.

Since in this model, the data requires to be broken up into several parts, the PULL-based technique does not offer enough assurance to reconstruct the data as the original one. That is the reason PUSH-based techniques have been proposed.

III. PROPOSED SYSTEM

In this model, a PUSH-based reconstruction technique is proposed, which consists of two schemes, PUSH-Rep and PUSH-Sur, which can exploit high parallelism and sequential access patterns. The models introduced are used to point out bottlenecks in the process. This process is used because the model requires the application to create replicas of the original data, which later will be broken up and stored in several replica servers. By using the PUSH-based reconstruction technique, all the data from the replica servers can be recovered and reconstructed.

Another thing that is introduced is Business Intelligence. It is a set of applications, products and technologies which are used of analysis of raw data. It can also collect, review and process the information. Business Intelligence is used by the model to review the data that is stored in the cloud server, before the data is broken up and stored in the replica server

3.1 METHODOLOGY

User in Cloud: The user can upload or download the data from the cloud server. To access the data present in the cloud server user has to be registered with the cloud server. Once the registration is done the user will be alerted

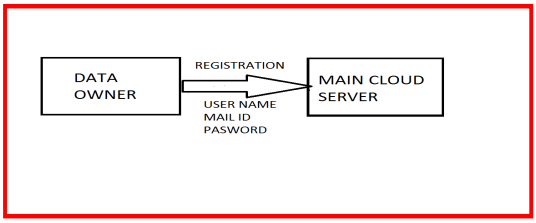

Fig: 1 Registration process.

UPLOAD PROCESS: When the data is [2] uploaded into the main cloud server it will split the data into many parts and store the data in separate servers located in that cloud. Hacking can be avoided by splitting the data, encrypting it and storing it in corresponding servers.

[3] Encryption is done based on the encryption keys that are stored in key server. By this, the security of the cloud server can be tightened up. The user has to provide entire keys to retrieve the data stored in key server.

3.2 ARCHITECTURE DIAGRAM

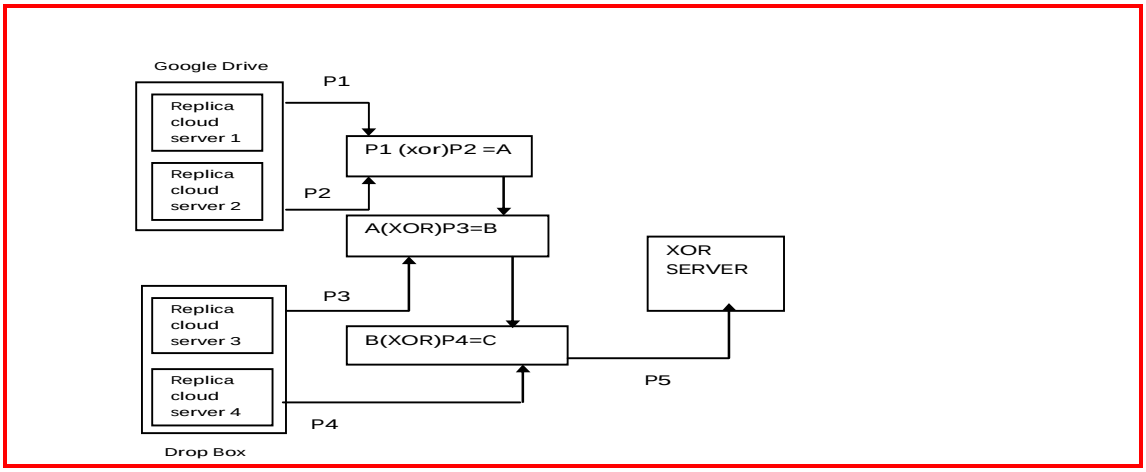

Fig :2 p1-binary format 1,p2- binary format 2,p3- binary format 3, p4-binary format 4

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 3, November 2017

XOR Reconstruction

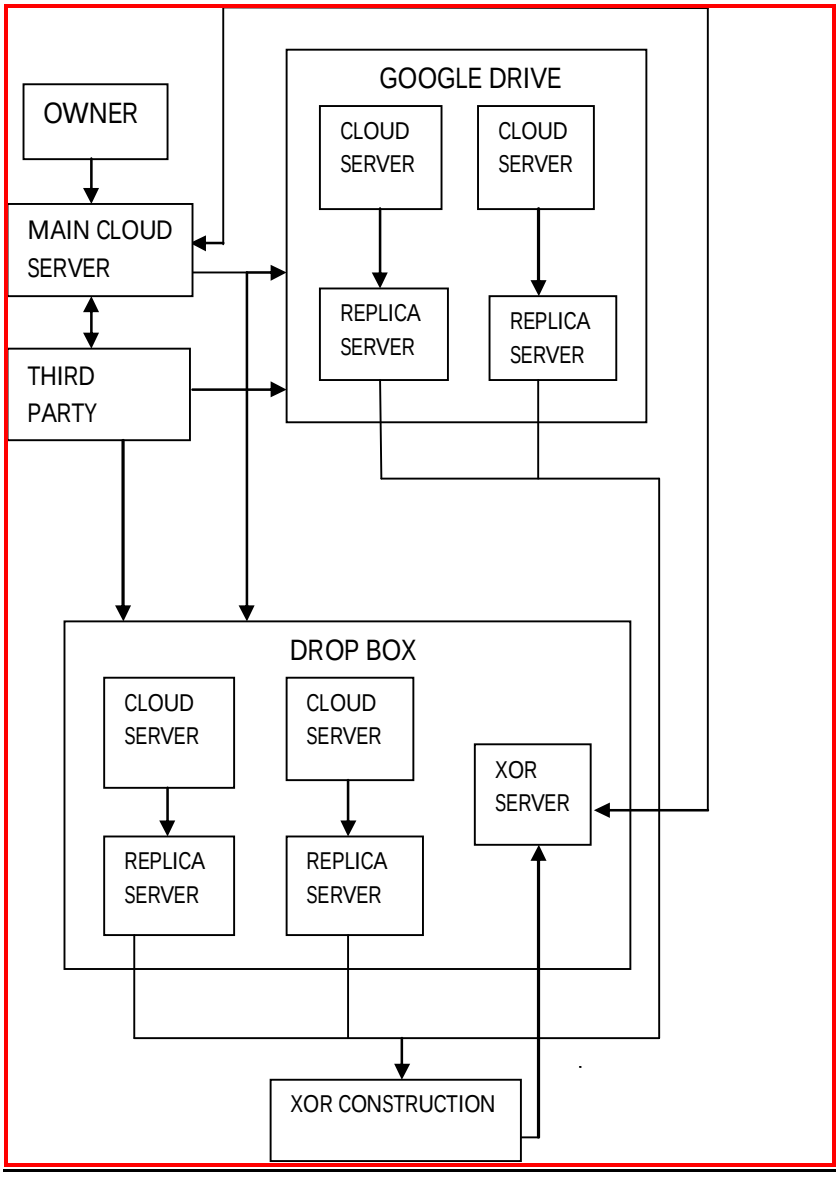

Fig 3 : Architecture Diagram.

XOR CONSTRUCTION: Once the data is stored in data server the parity bits are added to the data, then the data will be changed. [2] Data will be converted into binary data while performing XOR operation. If main cloud server retrieves the lost data in case of cloud server replica server failure it fetches the XOR-red data XOR-server and reconstructs the lost data using PUSH-Rep technique.

3.2 RECONSTRUCTION:

This scheme is used to reconstruct the lost data, rebuilt blocks are sequentially written to the disks of replacement nodes. [1]To maintain data integrity TPA checks are performed on the reconstructed data after reconstruction of lost data.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 3, November 2017

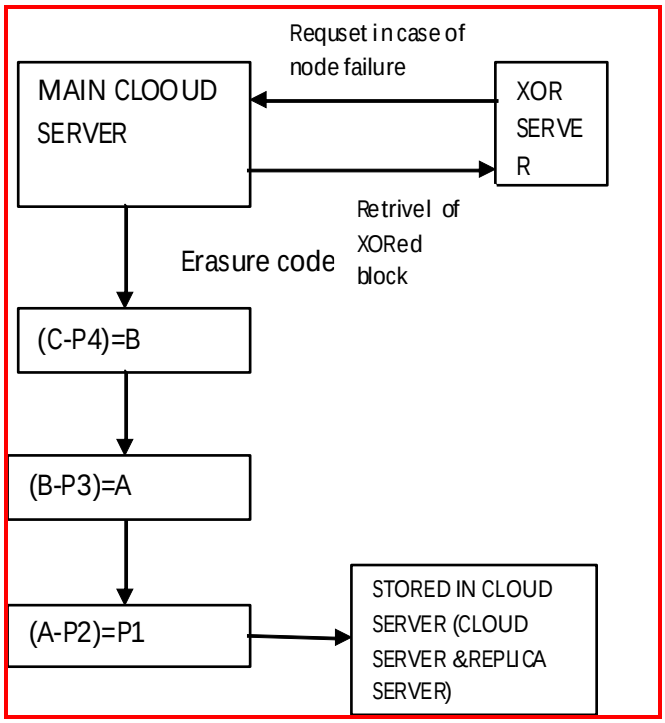

Fig : 4 Reconstruction system.

Erasure codes help a great deal in the reconstruction process, because they offer good fault tolerance which saves a lot of storage, thus allowing the user to store more data in relatively lesser storage space. Even though a large number of nodes are involved in the process, sufficient capability is offered by the storage nodes to process the data effectively. The nodes quickly share the workload, which eases the pressure on the other nodes, thus allowing them to work freely. This is completely unlike the PULL model, where a node first sends requests to the other nodes and then repairs a failed block using the requested ones. It is like a normal Master-Worker model, where the master first passes the orders and then waits for the tasks to be finished by the workers. The complication does not end here because the master node could be designated which takes in surviving nodes and reconstructs **or** it could be played by each of the surviving nodes.

- In the PUSH model, the reconstruction load is distributed among the nodes.
- Every surviving node combines a local block with a received block to produce a small part of a block, which is delivered to another node. This helps the nodes to spend all their resourceson the reconstruction.
- This is very important; because the data stored by the user is very delicate and even a small loss can have a long-term effect on the reconstruction.

3.4 ALGORITHM

All the nodes that are involved form a reconstruction chain. The PUSH step of an average surviving node would consist of the following steps:

1. Read a surviving block from memory.

//public static final String APP_KEY = "drnud8ob0u7f5sg";
public static final String APP_SECRET = "ndfbbwcht3dbwak";//

2. Receive an intermediate block.
3. Compute a combination of the blocks.
4. Deliver the resulting block to the corresponding node.
   Since PUSH involves a lot of storage nodes, failed blocks can only be reconstructed if every node pushes the intermediate blocks to the corresponding destination. So, the reconstruction might stop due to the reading or receiving process. To avoid this, a memory in every surviving node is pre-allocated to store both local and

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 3, November 2017

intermediate blocks. If two blocks are in a pair, then the node which has stored both blocks will be called to perform a linear-combination computation. Surviving nodes shouldput the rebuilt blocks in to the disks where they belong. Another rebuilt block is written into another space in the same disk. In this way, slowly, all the reconstructed data is put back in its place with the help of PUSH technique. Once the data is reconstructed, it can be requested and accessed by the user.

IV. PERFORMANCE EVALUATION

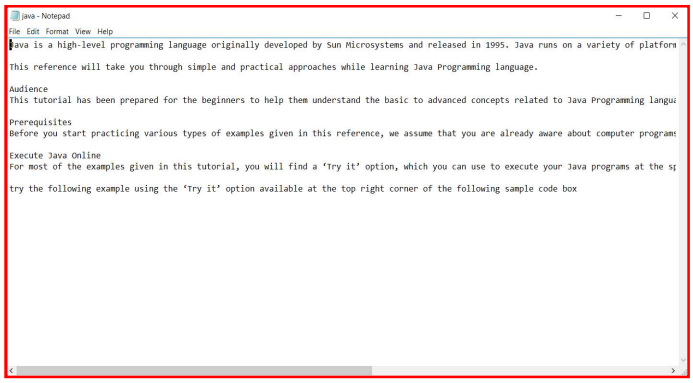

Fig 5.output file after download

The fig 5 is downloaded file , the data that is present in file will be stored in binary format. If the data is corrupted then it will downloads the corrupted data in binary format and replace the data.

Data recovery and reconstruction can be used in a number of real life scenarios such [4]as recovering all lost or damaged data by step-by-step reconstruction. This property is used to break up the data into pieces and then store them in several replica servers. Later, when the user requests for[5] the data, the module can reconstruct all the data and present it to the user. Business Intelligence is used to analyse the data before breaking it up so that sensitive information does not get damaged. All the data is securely stored in multiple replica servers, thus reducing its vulnerability to data loss.

V. CONCLUSION

Nowadays it is a grand challenge for storage clusters to efficiently migrate the data replicas to create an erasure code archive. So PUSH based reconstruction is extended to reconstruct lost data due to the failure of more than one node in erasure storage cluster.

REFERENCES

1. A.ArunDevi and D.Durai Kumar "data recovery usin bi methodology erasure code", International Journal of Science,Engineering and Technology Research , Vol 5, No .6,pp.1891-1895,June 2016.
2. P.Praveen Kumar, K.Madhan "cloud based data recovery and reconstruction system" volume 1,issue 1,2015//International Journal of Advanced Reasearch In Computer Science and Engineering Technologies .
3. Pradeep subedi "exploration of erasure code storage systems for high performance,reliability,andinter operability" 2016.Virginia Coomon Wealth University.
4. Prof.P.Manikandan , Santhosh kumar "cloud based data recovery and reconstruction system using bi methodology erasure code implementation" volume 2,issue 4.aprial 2016Intenational Journall of Recent Trends in Engineering & Research.
5. Lakshmi.A,Prof.C.Prakash Narayanan "cloud based data recovery and reconstruction system using bi methodology erasure code implementation" volume 18,issue 2- November 2015.International Journal of Emerging Technology inComputer Science & Electronics.
6. K. Rao, J. Hafner, and R. Golding, ―Reliability for networked storage nodes,‖ IEEE Trans. Dependable Secure Comput., vol. 8, no. 3, pp. 404–418, May 2011
7. L. Xiang, Y. Xu, J. Lui, and Q. Chang, Optimal recovery of single disk failure in rdp code storage systems,‖ ACM SIGMETRICS Per- form. Eval.Rev., vol. 38, no. 1, pp. 119–130, 2010.
8. O. Khan, R. Burns, J. Plank, W. Pierce, and C. Huang, Rethinking erasure codes for cloud file systems: Minimizing I/O for recovery and degraded reads,‖ in Proc. 10th USENIX Conf. File Storage Technol., 2012, pp. 251–264.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 3, November 2017

9. Z. Zhang, A. Deshpande, X. Ma, and E. Thereska, "Does erasure coding have a role to play in my data center?" Microsoft research MSR-TR-2010, vol. 52, 2010.
10. C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li, and S. Yekhanin, "Erasure coding in windows azure storage," in Proc. USENIX Annu. Tech. Conf., 2012, p. 2.
11. C. Ungureanu, B. Atkin, A. Aranya, S. Gokhale, S. Rago, G. Calkowski, C. Dubnicki, and A. Bohra, "Hydrafs: A highthroughputfile system for the hydrastor content-addressable storage system," in Proc. 8th USENIX Conf. File Storage Technol., 2010, pp. 225–238.
12. M. Aguilera, R. Janakiraman, and L. Xu, "Using erasure codes efficiently for storage in a distributed system," in Proc. Int. Conf. Dependable Syst. Netw., 2005, pp. 336–345.
13. Y. Hu, Y. Xu, X. Wang, C. Zhan, and P. Li, "Cooperative recovery of distributed storage systems from multiple losses with network coding," IEEE J. Select. Areas Commun., vol. 28, no. 2, pp. 268–276, Feb. 2010.
14. A. Kermarrec, N. Le Scouarnec, and G. Straub, "Repairing multiple failures with coordinated and adaptive regenerating codes," in Proc. Int. Symp.Netw. Coding, 11, pp. 1–6.
15. F. Zhang, J. Huang, and C. Xie, "Two efficient partial-updating schemes for erasure-coded storage clusters," in Proc. IEEE 7th Int. Conf. Netw., Archit. Storage, 2012, pp. 21–30.
16. Muthukumar B and Jabezs J. (2015),"Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach", ELSEVIER, Procedia in Computer Science , Vol 48, pp. 338-346.
17. Muthukumar B and Praveen Kumar R.(2015)," Intelligent Intrusion Detection System for Private Cloud Environment", Springer ,Communications in computer and Information Science Series(CCIS) , ISSN : 1865:0929 , Vol 356, pp 54-65.
18. Muthukumar B, Praveen Kumar R, and Nagarajan G.(2015),"Hybrid Intrusion Detection System for Private Cloud: A Systematic Approach", ELSEVIER, Procedia in Computer Science, Vol 48, pp. 325-329,
19. Muthukumar B and Jabezs J. (2015),"Anomaly-Detection in Diabetes using SVM", International Journal of Applied Engineering Research (IJAER),Vol 10, No 5, pp. 12627-12636, ISSN 0973-4562,