

# Survey on Synchronizing file operations along with storage scalable mechanism

Pradnya More<sup>1</sup>, Neha Prabhune<sup>2</sup>, Nidhi Vyas<sup>3</sup>, Amruta Gadekar<sup>4</sup>

<sup>1,2,3</sup> B.E Student, Dept. of Computer Engineering, D.Y.Patil Institute of Engineering & Technology, Ambi, Pune University, Maharashtra, India

<sup>4</sup>Assistant Professor, Dept. of Computer Engineering, D.Y.Patil Institute of Engineering & Technology, Ambi, Pune University, Maharashtra, India

\*\*\*

**Abstract** - As of late, there has been a blast of enthusiasm for mining time series databases. Similarly as with most software engineering issues, portrayal of the information is the way to productive and viable arrangements. A standout amongst the most usually utilized portrayal is piecewise direct estimate. This depiction has been used by various investigators to help bunching, arrangement, ordering and information sharing methodology of time arrangement information. An assortment of calculations have been proposed to get this portrayal, with a few calculations having been freely rediscovered a few times. In this venture, we attempt the primary broad audit and experimental examination of all proposed methods. We demonstrate that every one of these calculations have lethal defects from an information sharing point of view. We present a novel calculation that we experimentally show to be better than all others in the writing.

**Key Words:** insertion, deletion, splitting, concatenation, GridFTP, security

## 1. INTRODUCTION

In this work, we deal with efficient load balancing between the different resource nodes that process the client tasks, in a secure way as well as the elimination of possible single point of failure in a semi centralized load balancing architecture. To ensure that the two fundamentals i.e. co-ordination(the right things) and synchronization(the ideal time) of the processes will be executed we use synchronization algorithms. With such synchronization algorithms security will be provided to the data while transmission. This leads to less time consumption as the tasks are been executed concurrently.

Our System is a mixture of distribution model for P2P network. Data Sharing System, which has attracted the largest number of users, is the main application scheme for P2P file sharing. In broadcasting network, a single file is shared by numerous clients. The global data(files) to be transmitted is divided into Chunks(i.e. breaking the files into pieces) using chunking mechanism. The chunks can be of fixed size or variable size. All the parts connects to a central node called tracker to get a list of parts. Once all the

distributed pieces are obtained at single location then whole data is successfully broadcasted to destination path.

## 1.1: MOTIVATION

Privacy is most important factor which is supposed to be obtained nowadays . Also what we would like to provide is global data sharing scheme in distributed network with an efficient increase in data transfer. Security can be provided during the transmission phase instead of being processed after the data transfer is completed. Accuracy of the data being inserted or deleted can be obtained using security algorithm. Synchronization of hardware with software being another aspect can be obtained.

## 2. RELATED WORK

The Source Resource Broker (SRB) [4] is a middle ware framework that gives a uniform, UNIX-style record I/O interface for getting to heterogeneous capacity assets circulated over wide zone systems. Utilizing its Metadata Catalog (MCAT), SRB gives accumulation construct access to information based with respect to abnormal state traits as opposed to on physical filenames. SRB additionally bolsters programmed replication of records on capacity frameworks controlled by SRB. Rather than the layered Globus architecture with coordinate client and application control over replication, SRB utilizes an incorporated design, with all entrance of information by means of the SRB interface and MCAT and with SRB control over replication and copy choice. The Active Data Repository [6] and Data Cutter systems support examination of sweeping informational indexes, besides, could be utilized to accomplish elite operations. Antonio Carzaniga et al [3] utilized customary communicate convention joined with a particular substance based convention and utilized a push pull system for the spread of routing data, evaluated the tradition transport messages to focuses, avoid superfluous message advancement and make a sensible and stable measure of control traffic Hiroyuki Ohsaki and Makoto Imase [12] discussed that GridFTP is a data transfer protocol in grid computing which is widely used for transferring a large volume of data efficiently. Wen-Zhan Song and Xiang-Yang Li

[13] proposed an innovative peer-to-peer system architecture called CBR Brain to implement the content based routing services over the backbone routers rather than of at the terminal hosts to improve the efficiency and security.

The author Nader Mohamed et al [14] used Java Object Router (JOR), it is an application level router that dispenses with the limitation of content based routers utilize XML extensible content based object router that goes beyond message to routing entire java objects. Additionally XML limits the protest sent to 3 sorts of data, document and messages. Geoff Coulson et al [15] gives Grid kit approach which integrates middleware and overlay network functionality and eliminating exiting grid middleware is poorly equipped to support next generation grid applications Piotr Grzybowski's [16] Data Management System (DMS) main task is to store and manage computational data. Sasu Tarkoma inspected the cost of distributor and supporter mobility utilizing 3 mobility mechanisms and topologies such as generic mobility support, acyclic graphs and rendezvous based topologies. He also discussed the impact of completeness and incompleteness of the distributor and supporter topology on the cost of mobility and distinguished nonexclusive mechanism has a high cost for mobility, and the other two are considerably at lower cost [17].

In a previous work [19] Stefano Castelli, Paolo Costa observed that it is able to tolerate high churn with low overhead by evaluating the author proposed HYPERCBR approach in two applicable CBR contents , for example, content based quests in P2P networks and content based distribute and subscribe. Roberto Baldoni et al [20] explored a new approach to content based routing in mobile Ad Hoc Networks. In it messages are broadcast efficiently to all neighboring nodes and differs the nodes in their choice to forward the message in view of an estimation of their distance from potential subscriber of the message. The simulation demonstrates that the messages can be conveyed with high likelihood to the the interested subscribers at a low cost. Phillip M. Dickens, and Illino [21] discussed about a high performance file transfer mechanism for grid registering to accomplish incredible execution in a computational network and he utilized the idea of Lightweight Object-Based (LOB) file transfer System. He demonstrated that it can accomplish information exchange rates of 35 MB per second on one set of endpoints connected by the Abilene backbone network. Hari Subramonietal [24] designed a Advanced Data Transfer Service (ADTS), to provide the low-level InfiniBand which supports to the Globus X10 layer, and introduced the concept of I/O arranging in Globus X10 ADTS driver (composed in light of TCP/IP, UDP) to accomplish efficient disk based data transfer in both LAN and WAN. Wantao Liu et al [25] framework design were meet the data transfer requirements

to move large number of relatively small files reliably and rapidly to remote locations and building on GridFTP the system utilizes combination of automatic concurrency adaptation and restart mechanisms to move large volumes of data with high performance. Likewise his planning policies support the specification of dependencies between transfers and multiple network paths. Their outcomes demonstrate data transfer system can essentially enhance data transfer execution and can recover well failures.

The request manager depends specifically on two parts of the Globus Toolkit the replica inventory for copy area and GridFTP for secure, productive transfer. What's more, it utilizes Message Digest(MDS)- 2 for access to Network Weather Service (NWS) , and Grid Security Infrastructure for authentication. This employments of Globus Toolkit components are consistent with proposals for an data grid architecture [7, 8] which includes four levels: texture, connectivity, resource and collective At the lowest fabric level of the architecture is the basic components that make up the Grid, including storage systems, networks and computational systems. Moreover, it incorporates two catalogs: a metadata index and replica catalogs are the place enlisted replica catalog's data is stored. At the connectivity layer there are different standard protocols for communication and security. At the resource level there are services associated with managing individual resources GridFTP is placed here, which is used for the MDS information service and the Grid Resource Allocation. Management (GRAM) protocol is used for resource management.

Finally, at the collective layer there is higher level services that manages multiple underlying resources, involves services given by the request manager such as reliable file exchange, replica selection and information services that provide resource discovery or performance.

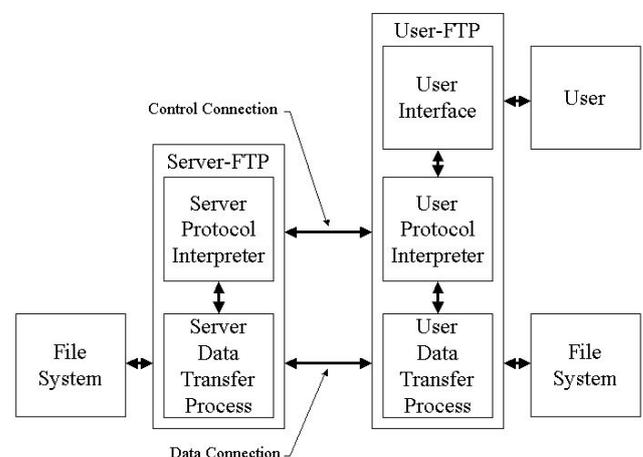
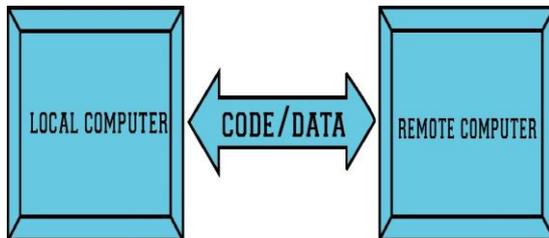


Fig -1: Normal Way-Data Transfer System

## 2.1 CONCEPT OF DATA TRANSFER

Data is transferred by some applications such as electronic mail, file transfer, web documents, so bandwidth and timing are important things for data transfer. Figure 2 refers data transfer;



**Fig -2:** Data Transfer

If you want to transmit small data, you need small rate bandwidth such as the application of internet telephony encodes voice at 32 kbps. However, if you have huge files and want to transmit them, you need more bandwidth. This is more advantages than small rate bandwidth. Timing is important when you transmit the data. Applications should provide quick data transferring to save time. For example, real-time applications of internet telephony, virtual environments, multiplayer games or etc.

## CONCLUSIONS

In this paper, we consider the issue for integrity checking of data sharing approaches with a remote server and will propose an efficient data securely sharing which will be specifically designed to handle a number of deletions linear in the length of the file for different operations where space utilization, security mechanism, splitting and concatenation operations are performed on file information. Our System also consist of verification methodology for integrity for the files stored on remote server, and reduces the storage costs and computation costs of the data. The presented scheme design is based on new lightweight hybrid data structure to support dynamic operations on blocks which incurs minimum computation costs by decreasing the number of node shifting. Using our new data structure, the data owner can perform insert, modify or delete operations on file blocks with high efficiency.

## ACKNOWLEDGEMENT

The authors would like to thank the publishers, researchers for making their resources available and teachers for their guidance. We thank the college authority for providing the required infrastructure and technical support. Finally, we extend our heartfelt gratitude to friends and family members.

## REFERENCES

- [1] Bart Jacob, (June 2003), TSO Redbooks Project Leader, IBM, "Grid computing: What are the key components? - Taking advantage of Grid Computing for Application Enablement".
- [2] Ann Chervenak, Ewa Deelman, Carl Kesselman, Bill Allcock, Ian Foster, Veronika Nefedova, Jason Lee, Alex Sim, Arie Shoshani, Bob Drach, Dean Williams, Don Middleton, (2001), "High-Performance Remote Access to Climate Simulation Data: A Challenge Problem for Data Grid Technologies" – A technical document, Supercomputing Conference-SC.
- [3] Antonio Carzaniga, Matthew J. Rutherford, Alexander L. Wolf, (2004), "A Routing Scheme for Content- Based Networking", Software Engineering Research Laboratory Department of Computer Science University of Colorado, Boulder, Colorado, USA, Technical Report CU-CS-953-03, June 2003 and IEEE INFOCOM.
- [4] Baru, C., Moore, R., Rajasekar, A. and Wan, M., (1998), "The SDSC Storage Resource Broker. 8th Annual IBM Centers for Advanced Studies Conference", Toronto, Canada.
- [5] Beynon, M., Kurc, T., Catalyurek, U., Chang, C., Sussman, A. and Saltz, J. (2001), "Distributed Processing of Very Large Datasets with DataCutter". *Parallel Computing*, 27 (11). 1457-1478. *International Journal of Grid Computing & Applications (IJGCA) Vol.2, No.4, December 2011* 61
- [6] Kurc, T., Catalyurek, U., Chang, C., Sussman, A. and Saltz, J. (2001), "Exploration and Visualization of Very Large Datasets with the Active Data Repository". *IEEE Computer Graphics & Applications*, 21 (4). 24-33.
- [7] Chervenak, A., Foster, I., Kesselman, C., Salisbury, C. and Tuecke, S. (2001), "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Data Sets", *J. Network and Computer Applications* (23). 187-200.
- [8] Foster, I. and Kesselman, C. (2001), "A Data Grid Reference Architecture", Technical Report GriPhyN-2001-12.
- [9] Stockinger, H., Samar, A., Allcock, B., Foster, I., Holtman, K. and Tierney, B. (2002), "File and Object Replication in Data Grids". *Journal of Cluster Computing*, 5 (3). 305-314.
- [10] Vazhkudai, S., Tuecke, S. and Foster, I., (2001), "Replica Selection in the Globus Data Grid. International Workshop on Data Models and Databases on Clusters And the Grid (DataGrid 2001)", IEEE Press.

[11] Foster, I., Kesselman, C., Tsudik, G. and Tuecke, S., (1998), "A Security Architecture for Computational Grids", 5th ACM Conference on Computer and Communications Security, 83-91.

[12] Hiroyuki Ohsaki, and Makoto Imase, (2009), "Performance Evaluation of Data Transfer Protocol GridFTP For Grid", International Journal of Applied Mathematics And Computer Sciences Vol. 3, No.1,.

[13] Wen-Zhan Song and Xiang-Yang Li, (2004), "CBRRBrain: Provide Content Based Routing Service over Internet Backbone", In Proceedings of ICCCN'2004. pp.101-106

[14] Nader Mohameda,, Xin Liub, Amy Davisb, Byrav Ramamurthy, (2005) , "JOR: a content-based object router", Computer Communications 28 ,654-663.

[15] Geoff Coulson, Paul Grace, Gordon Blair, Wei Cai, Chris Cooper, David, Duce Laurent Mathy, Wai Kit Yeung, Barry Porter, Musbah Sagar, Jason L, ( July 2006) , "A Component-based Middleware Framework for Configurable and Reconfigurable Grid Computing", Concurrency and Computation: Practice and Experience, Special Issue: First International Workshop on Emerging Technologies for Next-generation GRID (ETNGRID 2004), Vol.18. Issue 8, Pages 865-874.