

# A Prefetching Technique using Apriori and Logistic Methods for the DFS in Cloud

V. Thilaganga<sup>1</sup> and S. Selvam<sup>2</sup>

<sup>1</sup>Research Scholar, MCA Department

<sup>2</sup>HoD, BCA Department, NMSSVN College, Madurai, Tamil Nadu, India

E-mail: [gangamahi92@gmail.com](mailto:gangamahi92@gmail.com), [s.selvammphil@gmail.com](mailto:s.selvammphil@gmail.com)

(Received 6 January 2017; Revised 23 January 2017; Accepted 13 February 2017; Available online 20 February 2017)

**Abstract** - This paper presents an creativity data prefetching scheme on the loading servers in distributed file systems for cloud computing. The server will get and piggybacked the frequent data from the client system, after analyzing the fetched data is forward to the client machine from the server. To place this technique to work, the data about client nodes is piggybacked onto the real client I/O requests, and then forwarded to the relevant storage server. Next, dual prediction algorithms have been proposed to calculation future block access operations for directing what data should be fetched on storage servers in advance. Finally, the prefetching data can be pressed to the relevant client device from the storage server. Over a series of evaluation experiments with a group of application benchmarks, we have demonstrated that our presented initiative prefetching technique can benefit distributed file systems for cloud environments to achieve better I/O performance. In particular, configuration-limited client machines in the cloud are not answerable for predicting I/O access operations, which can certainly contribute to preferable system performance on them.

**Keywords:** DFS, Piggybacked, Prediction Algorithms

## I. INTRODUCTION

A cloud computing means, storing and accessing the data over the internet instead of client computers. Cloud computing and mobile computing is moderately new trend in Information Technology which are growing rapidly [1]. In the present period of time, cloud computing has been made use of for enabling end users to produce and utilize software without having to worry about the implementation of technical data at anytime from anywhere [2] [3]. DFS is a process of parallel process of sharing information between many clients that process of sharing over the internet is called as distributed file system then it continuously is persisted to be some backend depository system for providing I/O services to different types of information extensive trainings on the cloud computing situations [4], [5], [7], [6]. A major of the applications being executed on cloud need number files toward processing and computation [8]. In this paper signifies fetching I/O actions files from the client using dualistic prediction algorithms such as Apriori for access the frequent individuals data and logistic for if the accessed data was in server or piggybacked and frontward to significant machine. In previous researches are applied linear regression and chaotic time series algorithms

for predicted the data from client machine and response from the piggybacked data.

## II. CLOUD COMPUTING

The term clouds aimed at the Internet, cloud computing incomes at the Internet-based computing, somewhere different services are networks, servers, storage, applications, and services. Cloud computing is similar to grid computing, a type of computing where unused handling cycles of all computers in a network are harnesses to solve problems too intensive for any stand-alone machine. Cloud computing is a type of computing that mostly hinge on resource sharing as an alternative of handling applications by local servers or separate devices. By the internet enabled devices, cloud computing authorization the function of application software. Cloud computing, also known as the cloud, can be used as a alternative word for the Internet. Cloud computing can help a diverse range of tasks over the Internet similar storage and virtual servers; presentations and authorization for desktop applications. Through taking benefit of resource sharing, cloud computing is capable to achieve consistency and economies of scale. The types of cloud computing are categorized based on dual models. In a Cloud computing technology models are service models and cloud computing deployment models.

### A. Cloud Deployment Models

Around mainly four cloud deployment models, which are debated below, along with the scenarios a business could opt for each.

1. Public Cloud (The public cloud model represents true cloud hosting.)
2. Private Cloud (The cloud organization is operated uniquely for an organization.)
3. Community Cloud (The cloud organization is shared by numerous organizations and supports a specific community.)
4. Hybrid Cloud (The cloud infrastructure is a composition of two or more distinct cloud infrastructures are private, community and public.)

**B. Cloud Service Models**

Here are mostly three cloud service models, which are discussed below, beside with the scenarios in a business could opt for each.

1. Infrastructure as a Service (IaaS)
2. Platform as a Service (PaaS)
3. Software as a Service (SaaS)

**1. Infrastructure as a Service (IAAS)**

IaaS is the hardware and software that powers it all – servers, storage, networks, operating systems. Cloud substructure services, known as Infrastructure as a Service (IaaS), are self-service models for accessing, monitoring, and handling isolated datacenter infrastructures, such as compute (virtualized), storage, networking, and networking amenities such as firewalls.

**IaaS Examples:** Amazon Web Services (AWS), Cisco Metapod, Microsoft Azure, Google ComputeEngine(GCE), Joyent.

**2. Platform as a Service (PAAS)**

PaaS is the predictable tools also services planned to create coding and positioning individual’s applications swift and efficient cloud platform services, or else Platform as a Service (PaaS), are used on behalf of applications, and extra development, although as long as cloud mechanisms to software. Cloud Services offer an ideal means of supportive Commercial Process as a Service, a logical allowance or complement of Business Process Outsourcing, providing professional business functionality to several organizations. Enterprises profit from PaaS because it reduces the amount

of coding necessary, automates business policy, and helps drift apps to hybrid model. Designed for the needs of enterprises then further organizations.

**PaaS Examples :** Apprenda

**3. Software as a Service (SAAS)**

SaaS applications are considered for end-users, distributed over the web. Cloud application service area and Software as a Service (SaaS), signify the leading cloud market and are stagnant rising quickly. Its Benefits is Centralized Management of Data [3]. SaaS habits the web to distribute applications that are accomplished by a third-party retailer and whose interface is retrieved on the clients’ on the side. Supreme SaaS applications run openly from a mesh browser without any transfers or setting up required, even though some require plugins. Since of the network delivery model, SaaS removes the essential to set up and run applications on distinct computers. Through SaaS, it’s informal for enterprises to modernize their maintenance and support, since entirety can be managed by retailers: applications, runtime, information, middleware, OSES, virtualization, servers, storage and networking. Standard SaaS proposing varieties include email and collaboration, customer relationship administration, and healthcare-related applications. Certain large enterprises that are not usually supposed of as software retailers have started building SaaS as an additional source of income in order to achievement a competitive advantage.

**SaaS Examples:** Google Apps, Salesforce, Workday, Concur, Citrix GoToMeeting, Cisco WebEx

**TABLE I COMPARISON OF CLOUD SERVICES**

S.No	IaaS	PaaS	SaaS
1. Use.	Host (Or) Migrate	Build	Use (Or) Consume
2. Components.	Data Centers, Network, Storage, Security	Operating Systems, Databases, Languages.(Components & Services)	User Interfaces, Reports, Transactions.(Machine & Users)
3. Providers.	Amazon, At & T, Blue Cloud.	Google App Engine, Force.com, Azure.	Email, Facebook, Games, Virtual Desktop, CRM
4. Services.	Network Architecture.	Application Developer.	End Users.
5. Deployment Models.	Private Cloud, Hybrid Cloud.	Community Cloud, Public Cloud.	Community Cloud, Public Cloud, Hybrid Cloud.

**III. PROPOSED MECHANISMS**

**A. Apriori Algorithm**

Apriori is an algorithm for recurrent thing usual mining and association instruction learning above transactional databases. This one proceeds by classifying the regular individual objects in the database and extending them to larger and superior element sets as extended as those element sets seem sufficiently frequently in the database. The regular item sets determined by Apriori algorithm used to define association rules which highlight

general trends in the database: this consumes applications in domains such as arcade basket analysis. Apriori uses breadth-first examination and a Hash tree configuration to sum candidate item sets professionally. It produces applicant item sets of measurement from element sets of length. Formerly it prunes the applicants which have an uncommon sub pattern. Permitting to the downward closure lemma, the applicant set contains all frequent length item sets. Afterwards that, it scans the transaction database to define frequent item sets among the aspirants. It is a characteristic algorithm.

### ***B. Logistic Regression***

Logistic regression, or Logit reversion, or Logit model is a regression model somewhere the dependent variable (DV) is definite. The logistic regression circumstance are condition of a binary reliant on supple that is, somewhere it fire take merely two values, "0" and "1", which characterize products such as permit or flop, victory or lose, blooming or gone or strong or sick. Belongings where the dependent variable has more than two consequence classes might be analyzed in multinomial logistic regression, or if the several classes are ordered, in ordinal logistic regression. In the lexicon of economics, logistic regression is an example of a qualitative response or discrete choice model. Logistic regression is the suitable regression analysis to behavior at what time the dependent variable is binary. Similar all regression analyses, the logistic regression is a prognostic analysis. Logistic regression is used to define data and to clarify the relationship among one dependent binary variable and one or more insignificant, ordinal, break or ratio-level self-governing variables. Occasionally logistic regressions are problematic to interpret; the Intelligences Statistics tool effortlessly allows developer to conduct the analysis, and then in plain English interprets the output.

### ***C. Distributed File Systems (DFS)***

A file system is a subsystem of the operating system that performs file management activities such as organization, storing, retrieval, naming, sharing, and protection of files. A file system releases the programmer from anxieties around the details of space portion and plan of the subordinate storage device. A distributed file system is a file system the data is accessed and processed as if it was stored on the client machine. The DFS should perform or permit many clients have to access the data and processed the data.

## **IV. PIGGYBACKING CLIENT INFORMATION**

Most of the I/O tracing approaches proposed by other researchers focus on the logical I/O access events occurred on the client file systems, which might be useful for affirming application's I/O access patterns [9]. Nevertheless, without relevant information about physical I/O access, it is difficult to build the connection between the applications and the distributed file system for improving the I/O performance to a great extent. In this newly presented initiative prefetching approach, the data is prefetching by storage servers after analyzing disk I/O traces, and the data is then proactively pushed to the relevant client file system for satisfying potential application's requests. Thus, for the storage servers, it is necessary to understand the information about client file systems and applications. Towards in this conclusion, influence piggybacking mechanisms, happening the direction of transfer associated information from the client node to storage servers for contributing to modeling events of I/O access decorations and promoting the prefetching

data after distribution a logical I/O request to the storage server, the client file system piggybacks info approximately the client info systems and the application. In this way, the storage servers are able to record disk I/O events with associated client information, which plays a critical role for classifying access patterns and determining the destination client file system for the prefetching data. Scheduled the further side, the client info is piggybacked to the storage servers, consequently that the storage servers are conceivable to record the I/O operations complementary with the information about relevant logical I/O events.

## **V. PROCESS AND IMPLEMENTATION**

### ***A. Process of DFS with Prediction Algorithm***

We have proposed, implemented and evaluated a data prefetching approach on the storage servers for distributed file systems, which can be employed as a backend storage system in a cloud environment that may have certain resource-limited client machines. Towards be specific, the storage servers stay proficient of predicting future clients history of I/O details access to guide fetching data in advance after analyzing the existing logs, and then they proactively push the prefetching data to relevant client file systems for satisfying future applications' requests. For the purpose of effectively modeling disk I/O access patterns and accurately forwarding the prefetching data. Twofold prediction algorithms have been situated proposed to prognostication upcoming block access processes for guiding what data should be fetched on storage servers in next. Finally, the prefetching data can be pushed to the relevant client machine from the storage server. In this proposed mechanism, implemented and assessed an initiative data prefetching method on the storage servers for distributed file systems, which can be working as a backend storage system in a cloud situation that may have certain resource-limited client machines. Near remain the specific, the storage servers are talented of predicting future details of I/O access to groundbreaker fetching data in previously then analyzing the remaining logs, and then they proactively shove the prefetching data to applicable client file systems for sustaining upcoming applications' requests. For the purpose of effectively modeling disk I/O access patterns and accurately forwarding the prefetching data, the info approximately client details systems is piggybacked onto applicable I/O requests, and then transferred from client nodes to corresponding storage server nodes. Therefore, the client organizer systems running on the client nodes either process of neither I/O events nor behavior I/O access prediction; consequently, the thin client nodes attention on performing essential jobs with partial computing volume and liveliness endurance. Besides, the prefetching data will be proactively forwarded to the relevant client file system, and the latter does not need to issue a prefetching request. So that both network traffics and network latency can be

reduced to a certain extent, which have been demonstrated in our evaluation experiments.

### ***B. Implementation***

Dualistic prediction algorithms obligate been proposed to estimate upcoming block access operations for pointing what data should be fetched on storage servers in previously. To conclude, the perfected data should be pushed near the significant client machine from the storage server. Two prediction algorithms including the Apriori prediction algorithm and the logistic regression prediction algorithm have been proposed respectively. In this paper implementation is using Apriori and logistic regression algorithms are reduced the searching time in the piggybacked data process.

## **VI. CONCLUSION**

In this paper represents the basic knowledge of DFS, Apriori algorithm and Logistic algorithm. The Client details are accessed using prediction algorithms and forward to relevant piggybacked data, after request from the client and corresponding data will forward to the client machine proactively. Compare the linear and chaotic time series

algorithm it will avoided the searching timing for the relevant data on the piggybacked.

## **REFERENCES**

- [1] Minal Padwal, Prof. Manjushri Mahajan, 'Multimedia Storage System Providing QoS in Cloud Based Environment'.
- [2] N. Nieuwejaar and D. Kotz. The galley parallel file system. *Parallel Computing*, 23(4-5):447-476, 1997.
- [3] E. Shriver, C. Small, and K. A. Smith. Why does file system prefetching work? In *Proceedings of the USENIX Annual Technical Conference (ATC '99)*, USENIX Association, 1999.
- [4] D.Pratiba, Dr.G.Shobha and Vijaya Lakshmi.P.S "Efficient Data Retrieval From Cloud Storage Using Data Mining Technique" *International Journal on Cybernetics & Informatics (IJCI)* Vol. 4, No. 2, April 2015.
- [5] S.S and A. Basu, "Performance of eucalyptus and open stack clouds on future grid,"*International Journal of Computer Applications*, Vol. 80,No.13,pp.31-37, 2013.
- [6] J. Kunkel and T. Ludwig, Performance Evaluation of the PVFS2 Architecture, In *Proceedings of 15th EUROMICRO International Conference on Parallel, Distributed and Network-Based Processing, PDP '07*, 2007
- [7] D.Nurmi, R.Wolski, C.Grzegorzcyk, G.Obertelli,S.Soman, L.Youseff and D.Zagorodnov, "The eucalyptus open-source cloud- computing system," *CCGRID 2009.9th IEEE/ACM International Symposium*, 2009.
- [8] Vijay, Mrs.J.Jackulin Reeja, 'Receiving Files From System Server Using Data Perfecting Technique On Cloud'.
- [9] Jianwei Liao, Francois Trahay, Guoqiang Xiao, Li Li, Yutaka Ishikawa ,'Performing Initiative Data Prefetching in Distributed File Systems for Cloud Computing'.