ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

Predicting Resource Allocation in Distributed Environment by Using Online Predictive Approach: a Review

Prof. S. M. Tidke[1], Rucha Ravindra Galgali[2]

Assistant Professor, Computer Science and Engineering, Shreeyash Engineering college, Aurangabad, India[1]

Student, Computer Science and Engineering, Shreeyash Engineering college, Aurangabad, India[2]

Abstract: A distributed system consists of a collection of autonomous computers connected through network which enables computer to coordinate their activities and to share the resources of the system so that users perceive the system as a single, integrated computing facility. In distributed system there are many clients are connected, so it is very difficult to handle, analyze and process the data in distributed environment. To process such a large data many techniques like migration of data, replication of data, parallelism are available, but they have some disadvantages. These disadvantages have motivated this paper to implement the online predictive approach algorithm to predict the resource allotment for the process. The resource prediction is used to optimize the data access operations like read, write, uploading and downloading the file etc. Advantages of this strategy are the client can analyze the process behavior and also client will get application execution time.

Index terms: Time series, Application prediction, Distributed system, Application analysis, Replication

I. INTRODUCTION

A Distributed system is a software system in which components located on networked computers can communicate and coordinate their actions by passing messages. While handling large amount of data we will face some improper problems such as the time for executing the process will be more, the efficient data is not provided by the operation (Read/Write).

There are many scientific applications which produces large amount of data. It is very difficult to handle, analyze and process such data. There are many existing systems which causes processing, handling and analyzing such a large data. For example clusters and grids [1].

The aim of this paper is to improve scheduling decisions in large scale environments by avoiding historical processes. This paper uses time series to predict process resources by avoiding historical data. Here we are forming every process as time series. If we formed as time series means we can easily understand how much time taking for each process.

At a same time we can reduce application execution time, by using time series we can easily analyze the operations such as Read/Write.

The purpose of this system is to improve the performance of the distributed system. This can be achieved by predicting the resources and reducing the data access to database. There are many techniques available to optimize data access like data replication, migration, distribution and access parallelism but these techniques doesn't consider the dynamic behavior of process [1]. This paper supports a strategy which evaluates the efficient reducing system memory locations and predicts the application files and maintain every process that is read and write operations time series for each operations , so we can able to easily analyze the operation. This can also avoid historical data duplication.

This paper contains data optimization approach organizes application behaviors as time series and, then, analyzes and classifies those series according to their properties. By knowing properties, the approach selects modeling techniques to represent series and perform predictions, which are, later on, used to optimize data access operations [1].

II. STUDIES DONE BY DIFFERENT AUTHORS

This section gives the related works on the analyzing application behavior and then predicting application behavior to reduce data access. Following are the few papers studied for analyzing behavior of various processes and allocation of resources for their execution.

Renato and Mello [1] propose the strategy which supports the online prediction of application behavior in order to optimize data access operations on distributed systems without requiring any information on past execution. In this approach the first step is to monitories the process behavior. After that these processes are converted into time series, according to properties time series get analyzed and classified. Then modeling technique is selected to model these time series to get some future observations. In last step these predictions are used to optimize the data access. Rahman and Barker [2],propose a framework that predicts the sites transfer times which are hosting replicas, the data from various sources is used for that purpose. They also used the neural network to predict the transfer time of different sites that currently hold file replicas. Devarakonda and Iyer[3, 1] propose a statistical approach to predict the consumption of CPU, file system I/O, memory. This study is verifying the process behavior with automaton stored in database. When a new process arrives at the system it is verified with the automaton, if any automaton from the database is capable to represent that process then that automaton is used to estimate the resource requirements for the process. Kim and Chandra [4, 1] presented accessibility aware resource selection techniques to choose nodes which can efficiently access data from remote data sources. They showed that the accessibility of node is depending on the local data access observations collected from the nodes neighbors. They also proposed the heuristic to reduce execution time of data intensive applications. Vazhkudai and Foster[5] designed and implemented high level replica selection service which uses information regarding replica location and user preferences to guide selection from storage replica alternatives. They presented a dynamic information collection using Globus information service capabilities concerned to storage system properties and how this information can help to improve and optimize the selection process.

Faerman and Wolski [6, 1], propose the adaptive regression modeling (AdRM) to determine file transfer times for network bound distributed data intensive applications.AdRM method accurately predicts data transfer times in wide area multiuser environments. Oldfield and Kotz [7, 1] propose Armada framework to monitor, control and execute the applications. Armada represents process and data flows by using Graph structure. It also improves network throughput. Oldfield and Kotz [8], also described the design of the flexible parallel system that allows the application to control the behavior and functionality of file system aspects. Jeong and chan-hyun-youn [9] propose optimal file replication scheme (CO-RLS) that uses users cost and deadline requirements to minimize the total replication cost. This cost is calculated under two constraints one by using storage cost at remote sites and other is the transfer time from data source to multiple candidate sites. Senger [10, 1] propose an online approach to acquire, classify and extract process behavior. By using this approach one can simply monitor the user application without any need of recompilation or modification. This approach is capable of automatically modeling process behavior. Senger [11, 1] also propose an approach to predict execution times of parallel applications. This approach has improved scheduling decisions in large scale environments and it has also provided the knowledge of application to system scheduler which makes resource allocation.

III. APPLICATIONS

A. Amsat:

Amsat is a name for amateur radio satellite organizations worldwide, but in particular the Radio Amateur Satellite Corporation (AMSAT-NA) with headquarters at Silver Spring, Maryland, near Washington DC. AMSAT organizations design, build, arrange launches for, and then operate (command) satellites carrying amateur radio payloads, including the OSCAR series of satellites. Other informally affiliated national organizations exist, such as AMSAT Germany (AMSAT-DL) and AMSAT Japan (JAMSAT).

B. Commodity Futures Trading Commission:

The Commodity Exchange Act (CEA), 7 U.S.C. *et seq.*, prohibits fraudulent conduct in the trading of futures contracts. In 1974, Congress amended the Act to create a more comprehensive regulatory framework for the trading of futures contracts and created the Commodity Futures Trading Commission, replacing the Commodity Exchange Authority. The stated mission of the CFTC is to protect market users and the public from fraud, manipulation, and abusive practices related to the sale of commodity and financial futures and options, and to foster open, competitive, and financially sound futures and option markets

C. Online prediction market:

Predicting how people will vote is even trickier than predicting what and how much they'll buy or what they'll be willing to pay for stocks. There are comfortable quantitative fundamentals in the world of finance; there's guidance, there's supply and demand and rational expectations. Although stock market predictions are far from scientific, people buy more rationally and predictably than they vote, where the picture can be clouded by the most subtle and unpredictable of psychological nuances

IV. SYSTEM DESIGN

Renato and Mello propose the strategy which supports the online prediction of application behavior in order to optimize data access operations on distributed systems without requiring any information on past execution [1]. Our paper uses online predictive approach algorithm to develop the system. This paper implements the proposed algorithm that is online predictive approach algorithm. The system architecture for development of the system is shown in the following figure. In this figure the server is generating time series for every operation and monitoring the application.

A. System Architecture

System architecture diagram shows how the operation flow is happening first step the execution start from if the new client the register process will be first otherwise already register client means directly process with login process. Client will process with user name and password. After that the user name password will check with database if exist means the client page will open. Otherwise again it will process login page itself. After that client give the request to server and server will responds to client. Data flow diagram shows how the operation flow is flowing between client and server in this case the first step shows the client request and server will receive the client request then server will responds to client. After finishing all operations between client and server time series will calculate for each client request and server responds. Finally the graph will generate for each process. Then how many clients are running that will be monitor.

B. E
xplanation of online predictive approach algorithm

There are few steps which can be performed in this algorithm. These steps are given below [1]:

1) Interception: The first step, is application knowledge acquisition, is responsible for monitoring process behavior by using event interception. The interception mechanism is associated with the process under execution.

When a program calls a function, DLSym intercepts the call and injects any code instead.

2) Conversion: After extracting the application behavior, it transform the sequence of read-and-write events in a multidimensional time series.

3) Time Series : The third step evaluates the generation process of time series TR according to specific properties: stochasticity, linearity, and stationarity

4) Selection: Based on the evaluation of the time series generation process, we select an adequate modeling technique.

E.g. when the series is deterministic, a reconstruction is conducted by considering the Takens' immersion theorem which relates series observations over time.

5) Adaptive Sliding Window: In this step we consider the adaptive sliding window (ASW) mechanism proposed to estimate the number of time series observations to be predicted, based on process behavior changes.

6) Prediction: After the previous steps, the prediction is performed on the time series, which represents process behaviors.

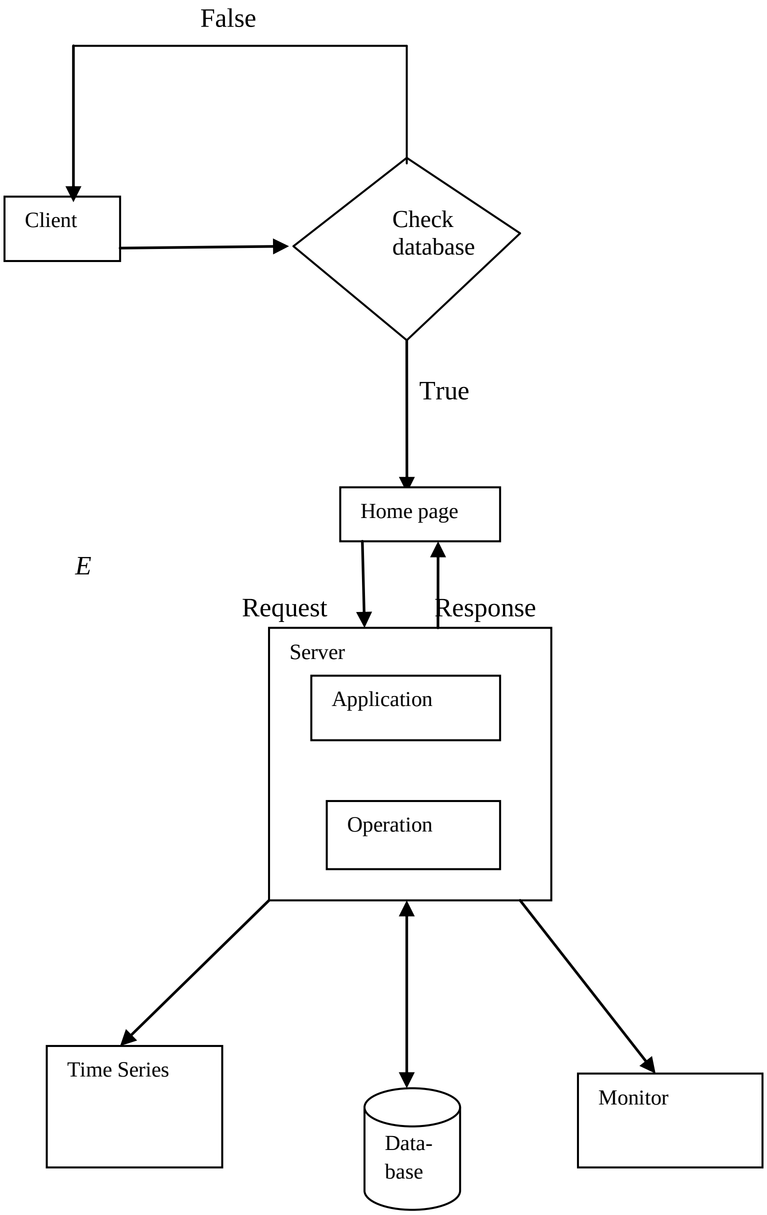

Fig. 1 System Architecture

V. CONCLUSION

A distributed system is a software system in which components located on networked computers communicate and coordinate their actions by passing messages. This system minimizes the application execution time by optimizing data accesses and then data access operations are transformed into time series We evaluated our approach to select modeling techniques for real systems data. This system uses time series by using which the client can analyze how much time the operation requires to execute. Time series is generated for every operation so that client will get performance chart. This is used to improve performance of distributed system.

REFERENCES

[1] Renato Porfirio Ishii and Rodrigo Fernades de Mello, "An Online Data Access Prediction and Optimization Approach for Distributed Systems", Parallel and Distributed Systems, *IEEE Trans.* vol.23, no. 06, June 2012.

[2] R. M. Rahman,, K. Barker and R. Alhajj, "A Predictive Technique for Replica Selection in Grid Environment ",*Proc. IEEE Seventh Int'l Symp.* Cluster Computing and Grid, pp. 163-170, May 2007.

[3] M. Devarakonda and R. Iyer, "Predictability of Process Resource Usage: A Measurement- Based Study on Unix," *IEEE Trans.* Software Eng., vol. 15, no. 2, pp. 1579-1586, http://dx.doi.org/ 10.1109/32.58769, Dec. 1989.

[4] Jinoh Kim, A. Chandra and J. B.Weissman, "Using Data Accessibility for Resource Selection in Large-Scale Distributed Systems ", Parallel and Distributed Systems, *IEEE Trans.* , vol. 20 , pp. 788 – 801, 2009.

[5] S.Vazhkudai. , S. Tuecke and I. Foster, "Replica selection in the Globus Data Grid ", Cluster Computing and the Grid, 2001. Proceedings. First *IEEE/ACM International Symp.*, pp. 106 - 113, 2001.

[6] M. Faerman, A. Su, R. Wolski, and F. Berman, "Adaptive Performance Prediction for Distributed Data-intensive Applications," Proc. ACM/IEEE Conf. Supercomputing (Supercomputing '99), p. 36, 1999.

[7] R. Oldfield and D. Kotz, "Improving Data Access for Computational Grid Applications," Cluster Computing, vol. 9, no. 1, pp. 79-99, Jan. 2006.

[8] R. Oldfield and D. Kotz, " Armada: a parallel file system for computational grids ", Cluster Computing and the Grid, 2001. Proceedings. First IEEE/ACM International Symp., pp. 194 - 201 , 2001.

[9] Sangjin Jeong , Chan-Hyun Youn and Hyoug-Jun Kim, " Optimal file replication scheme(CO-RLS) for data grids ", Advanced Communication Technology, 2004. The 6th International Conf., vol. 2, pp. 1055 - 1059, 2004.

[10] L. Senger, R.F. Mello, M.J. Santana, and R.H.C. Santana, "An On-Line Approach for Classifying and Extracting Application Behavior on Linux," High Performance Computing: Paradigm and Infrastructure, pp. 381-401, John Wiley and Sons Inc., 2005.

[11] L.J. Senger, M. Santana, and R. Santana, "An Instance-based Learning Approach for Predicting Parallel Applications Execution Times," Proc. Third Int'l Information and Telecomm. Technologies Symp., pp. 9-15, Dec. 2005.