Research Paper

Volume : 3 | Issue : 9 | September 2014 • ISSN No 2277 - 8179

Metrics for Measurement of Service Replicas in Grid Environment

Engineering

KEYWORDS : Agent replica, Data grid, Metrics, Mean time to failure, Quality of Service (QOS), Search rate.

*** P.SunilGavaskar**	Research Scholar, Sri Venkateswara University, Tirupathi, Andhra Pradesh, India * Corresponding Author
Ch.D.V.Subbarao	Professor, Department of Computer Science and Engineering, Sri Venkateswara University, Tirupathi, AndhraPradesh, India.

ABSTRACT

The important requirement in data grid environment is replicas utilization. In general replicas are used as proactive (i.e. failure considered before scheduling of a job) and post active handles the job failure after it has occurred. We used centralized and local replicas within agents. In this paper, we address the probabilistic models of Fault tolerance, communication, efficiency and reliability of simple job submission. We have shown an autonomous, self adaptive agents model to model replicas in distributed environment for long execution of jobs. The applicability of the proposed agent's architecture in data grid is able to setup some parameters of replicas such as centralized and local replicas search rate with mean time failure and number of objects used in it. The search values of centralized and local replicas are measured when computing happens. The Metrics used for search value are number of replicas, replication objects with different states and processing capability. The metrics we considered are to reduce complexity in search rate of replicas due to their huge computing and storage capabilities. On the basis of centralized and local replicas search value we predict the unreliable agents computing model as result of that we can increase computing reliability or we can avoid unreliable model into computing. Hence we showed an autonomous, self adaptive agent model in order to reduce service failures of replicas in large scale distributed environment.

I. INTRODUCTION

In general grid systems are important alternatives to provide decentralized access rate of data. Grid [2] is a basic infrastructure of national high performance computing and Information services, the grid achieves the integration and interconnection of many types of the high performance Computers, Data servers, large-scale storage systems which are Distributed and heterogeneous in nature. However, the vast complexity of these systems limits the efficiency of the availability, reliability focused on nodes. Aggregating the usage of resources between clusters nodes are important issues. In The encoding schemes each group has an encoding node to tolerate single failure per entire group. Some of the behavioural activities of such existing storage and computing node are discussed with efficiency. However, if there is more than one failure the encoding schemes not enough to handle and restart its execution [1,3].In scalable distributed replicas location method home node maintains all replicas and query other nodes for replica supported operations[4].

Paxos Consensus and Paxos Commit techniques uses single value from set of proposed values, in such kind of techniques a process may perform many communication operations during its protocol operations [6]. In order to ensure data availability when nodes join, leave or crash at any time, the Data consistency is enforced by performing all data operations on a majority of replicas [7]. Reliability and Availability are two qualitative measures in Grid computing, which depends on Mean time to Failure [8].

Replication in distributed environment is a practical and effective method to achieve efficient and fault tolerant data access in grids [8].Neighbour replication grid(NRG)requires lower communication cost for an operation, so it provides high system availability[9,10].By considering complexity in real time application in data grid environment, when the data or service is replicated when new type of attacks comes into system[13].Implementing workflows in loosely coupled systems are dynamically changes in order to achieve QOS dynamic changes are required in grid system.

In Grid environment data replication has several replica allocation policies are used based on system state. The policies could be optimal and it is very unbalanced when network connections are heavy loaded, others when the CPU usage is very unbalanced, etc. The dynamic selection of adequate replication and utilization for each state could improve the system performance strongly.

Replica utilization in fault tolerance strategy is challenging research in data grid. By placing multiple replicas at different locations, replica management process can reduces network delay and bandwidth consuming of remote data access [5]. Issues with Replication are widely studied in Grid. Our Proposed Agents replica model mainly shows that the optimized Agents model. This agent's model provides a deep knowledge about replicas usage in behavioural activities of replicas storage and computing requirements. In our model we considered the properties of replicas search policies using dynamic changes in agent's movement that help to estimate performance and reliability of the model.

The rest of the paper is organized as follows. In section 2, we present proposed model related to replication in grid environment, Section 3 is evaluation. Finally, we conclude our strategy in section 4.

II. SYSTEM MODEL

- **Agents replica architectural model**

In proposed model each agent is associated with group of objects. A group of tasks are associated with agents for execution with duration of time t units for execution in computing grid with agents as replicas, the jobs are divided into subtasks and submitted to agents for execution..The following fault tolerance Replica model performs major role in replicas search cost estimation, which helps to handle faults.

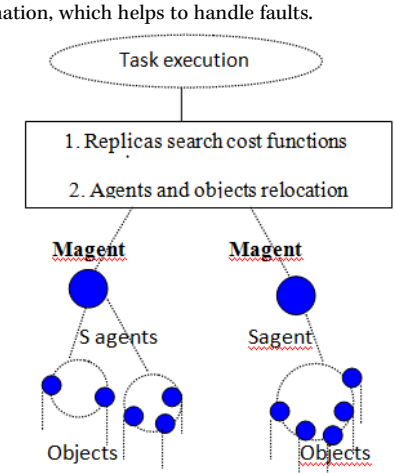

Figure.1. Replicas Initial Stage Figure.2.Replicas when S agent fails

The fault tolerance system proposed in this section has been to provide reliable model for replication in grid environments. If Agents become failed then adjacent agent performs its fault recovery operations with help of the status maintained by resource broker. Based on this information different intensity of replicas and objects are discovered while jobs execution on agents. To increase the performance of jobs execution within specified time we use metrics for search rate of replicas.

S agents:
The role of agents is storing local replicas of the currently registered distributed objects.

M agents:
They also act as replicas and contain local copies, there by subset of Agents becomes under the control of M agents, in the proposed model we consider that we have numbers of M agents are equivalent of classes. S agents are having length of class i.e. Number of objects supporting in it.

The agents are considered as replicas, it contains objects to identify particular replica or same replica with different request will be treated as object with different state clients. In order to identify particular object failed or not it requires log C steps, where C represents number of objects currently possessed by agents. The client initiates the update or search operations. (1). The M agent discover that particular S agent is not active in the system(i.e. agent failed), then the M agent which is in charge of its S agent performs object reallocation to adjacent S agent using inheritance technique.(2).If M agent fails, one of the M agent connected S agent converted into M agent.

In proposed execution model, value 'x' denotes centralized replicas are allocated to the application as Minimum and Maximum number of objects and agents. The search costs of replicas are as follows.

x= M*K **(1)**
The Local replicas used with agents and objects, where search costs are calculated using the following equation.

y=M*K*(4O+Log C) **(2)**
Where K is the object/agents keys in model, M is root agents or M agents, O is the number of objects matching search criteria, C is the objects currently posses by agents.

In dynamic grid environment, there is some fuzziness in the factor of Replicas used as centralized, value 'x' denotes the centralized replicas search cost, the service reliability of centralized replicas are

$$X(t) = 1-(1-e^{-\lambda t})^{x} \quad (3)$$

For Local replicas service reliability of maximum and minimum agents/object is

$$Y(t) = 1-(1-e^{-\lambda t})^{y} \quad (4)$$

Where Y(t) represents the service reliability of the agents groups. Moreover the 'y' denotes Local replicas search rate calculated from equation2, 't' is the time period of a task.

The reliability of the centralized/Local replication are performed using the above said metrics, Furthermore replicas cost of access rate applied in Distributed hash structure of the replicas, It will be as follows.

Average of replicas search (AS) =∑ Replicas search rate (Min or Max) ÷ task execution in an unit time (t) (5)

Search rate Deviation (σ) =AS- Min $(\bar{x})^2$ or Max $(\bar{x})^2$ (6)

Search rate Deviation (σ) =AS- Min $(\bar{y})^2$ or Max $(\bar{y})^2$ (7)

Variance factor of Replicas (VR) = $1/(SD)^2$

Reliability Metrics (RM) =1- $(1/(SD)^2)$

$$= (SD)^2 - 1 / SD^2 \quad (8)$$

Where σ is the search rate deviation of local and centralized replicas of cluster group .The centralized and local replicas search value are measures based on the metrics, number of objects ,key values used in search and task execution time per time unit(t)=0.5 presented in data grid environment. As shown in metrics the replicas weighted search cost from the Equation 8 is carried for all local and centralized strategies in order to benefit for Reliability factor of replicas, which gives better performance when we have negative signs in the access rate of replicas changes dynamically.

TABLE. I Reliability with mean failure rate=0.001

Agents size	Centralized (x)		Local (y)	
	Min.	Max.	Min.	Max.
2	0.4995	0.9685	0.9372	1.0000
3	0.7495	0.9990	0.9994	1.0000
5	0.9842	0.9999	1.0000	1.0000
7	0.9997	1.0000	1.0000	1.0000

TABLE II Number of REPLICATIONS SEARCH RATE

Agents size	Centralized(x)		Local(y)	
	Min.	Max.	Min.	Max.
2	1	5	4	103.49
3	2	10	16.602	205
5	6	30	74.86	1835.28
7	12	60	199.22	3678.06

The table1 shows that reliability of nodes in replication cost (read cost, update cost) and its scalable performance. The table2 shows that search rate factor of the replicas with constant rate of replication.Table3 and Table4 shows that dynamic changes in replicas location and its improved search rate with reliability in task execution in dynamic replication of replicas.

TABLEIII search rate deviation with task execution 1800sec (t=0.5)

Agent size	Centralized(x)		Local(y)	
	Min.(σ)	Max.(σ)	Min.(σ)	Max.(σ)
2	-25.56	-639	-89165.79	-2096955.7
3	-19.56	-489.06	-88646.53	-2034284.7
5	-44.43	1110.93	-77989.45	4618170.6
7	260.43	6150.9	-9820.57	51994166.75

TABLE IV Reliability WITH TASK execution 1800sec (t=0.5)

Agents size	Centralized (x)		Local (y)	
	Min.	Max.	Min.	Max.
2	0.9984	0.9999	0.9999	1.0000
3	0.9973	0.9999	0.9999	1.0000
5	0.9994	0.9999	0.9999	1.0000
7	0.9999	0.9999	0.9999	1.0000

From the observation of Table.1 and Table.4 the model with agents 2,3,5,7 are considered to express the performance of 25 tasks with an estimated execution time (t) duration of 1800sec and mean failure rate=0.001,we achieved reliability in table4.

Volume : 3 | Issue : 9 | September 2014 • ISSN No 2277 - 8179

- **Job Configuration Settings**

Let us consider the following configuration to point out the node reliability. From table1 the values with less than 0.4002, which notifies the failure nodes in the job execution, i.e those are not under execution threshhold.Further these nodes are to be improved using metrics. The enhanced replication reliability is shown in table4.

In our proposed scheme due to dynamic changes in the distances of replicas we obtained different cumbersome distances such kind of distances can be reduced using dependency and Independency of replicas further we obtained improved search rate.

We used database replication scheme into our proposed scheme. Our primary goal is to reduce the search rate of replicas in DHT systems by applying data access latency with following queries.

Select Node_Name, Reliable from MFR Where Replication='Centralized'Having(Reliability<=0.4002 AND Timeout=0.5); output all every 4 min.

Second request consumes performance metrics of computed model with specified metrics which point outs the node reliability with less than 0.8002 and then which notifies the Reliability of nodes to be improved.

Select Node_Name, Reliable from MFR Where Replication='Local' Having (Reliability<=0.8002 AND Timeout=0.5) output all every 4 min.

Third and fourth statements indicate that reliability attained with centralized and local replication schemes respectively. We would like to ensure that reliable nodes with replications are placed in the reliable table and gives satisfied metrics for computing using following requirements.

Create table Reliable (Ns Number, Cr_Min Number (2,6), Cr.Max Number (2,6) CHECK (Cr_Min=>0.9800 AND Cr_Max=>0.9800).

Create table Reliable (Ns Number, Lr_Min Number (2,6), Cr.Max Number (2,6) CHECK (Lr_Min=>0.9800 AND Cr_Max=>0.9800).

The changes in search factors are to be evaluated during overhead monitoring .Furthermore interactive replication can be improved using following metrics.

insert into AvgCr_Min, AvgCr_Max, select Node_Name, Timeout=0.5 as Cr_min,Cr_Max from ReplicasSerach.

Then the result of the above requests is applied for distance variance between replicas, finally we have updated replication cost withTable3 and Table4.

The following event trigger gives nodes replication metrics to enhance the performance of dynamic replication.

Create or replace trigger Per_improved_metrics After insert on AvgCr_Min, AvgCr_Max for each row

Begin
Loop : newAvgCr_Min or newAvgCr_Max:=1-VarianceFactor;
Exit When Reliability =1.0000 and Reliability =>0.9990
End Loop

III. EVALUATION

Data replication is one of the key estimation with respect to time of the replica location can reduced by Key of replicas which are stored in multiple nodes. Our analysis results shows that reliability service depends on the probability of all agents are operational and can properly communicate with each other in entire time interval (0, t). The model with nodes 2, 3, 5, 7 is considered to express the performance 40 to 1000 records of 25 tasks with job data size is from 30 to 180MB. The probabilistic values reliability comprises of the simple grid job submission with an estimation of mean failure rate 0.001 of agents and tasks duration. Reliability performance of agent's model calculated upon agents, weighted search cost of local and centralized replicas. Unlike centralized replicas where as in local replicas model almost we have achieved 1.0000 reliability irrespective of agents size as result of that number of agents failures are less and also scalability is effective when compare to centralized replicas. The Table.2 shows that number of replica used as centralized, local, minimum, and maximum and its access rates are formulated.Let us considerTable.2 and Table3 We have two different scenarios of values that represents search rate values growth rate is always depends on number object replica states and agents replicas. One of the useful and applicable items of statistics is deviation in search rate (σ) which is measured based on the centralized and local replication with their minimum and maximum replications.

While the metrics captured by Eq.5, Eq.6, Eq.7 are estimated for each cluster of agents group individually to test the search rate weighted average of each job execution per unit time=0.5. The table3 determines search deviation, that the negative sign values can always give low weighted search cost which constitutes fewer objects and positive values can give high reliability due to supporting objects are always high.

Our proposed model is extension of my previous work, in previous work [12] we had shown only the mean failure rate reliability with agents used in the data grid replication while the number of tasks submitted for execution. In the proposed architecture model we shown replicas search rate as weighted factor for estimation of job execution reliability with specified mean failure rate. In the metrics we applied time unit=0.5 into search rate of local and centralized replicas we may assign average rate of table2 values into specified model as result of that we achieved majority of search rate values which is specified in table3.It serves the objects among computing and data nodes of the specified architecture but reliability have no guarantee in table3 for all the nodes. As consequence of that we applied reliability, using search rate deviation values to predict acceptable search rate and to avoid failures. This mechanism provides proper reliability in large scale systems.

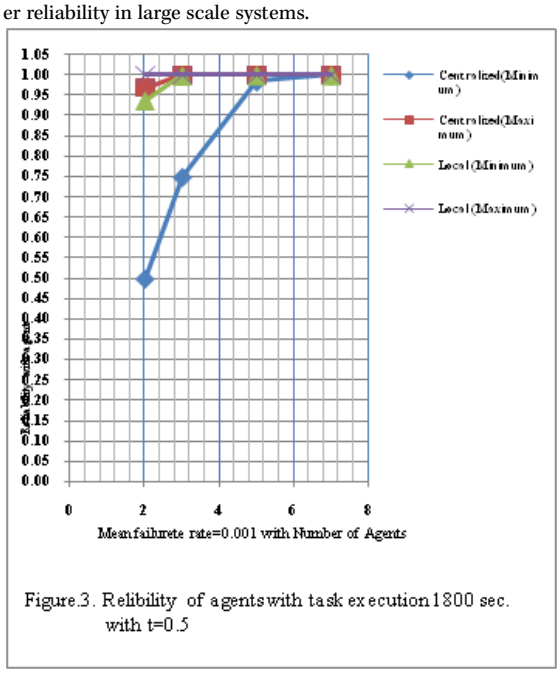

Figure.3. Relibility of agents with task execution 1800 sec. with t=0.5

In the above Graph representation, from figure 3 we have brought mean failure rate of 0.001 into agents model to attain reliability of 1.0000.In this section local (maximum) strategy alone have achieved the reliability of 1.0000 with task execution time 1800 second along with t=0. 5. The search rate of local (Maximum) can assesses the constant impact even we applied agents size=2, 3, 5, 7. It could be argued that this measure will provides connection between the structural representations of the model and job execution with replicas.

The next step is to apply job execution with search rate reliability into our model.

From the presentation of Figure4,we considered same mean failure rate which had been considered in previous graph but difference exists only in the search rate weighted factor of replicas, The replicas with centralized and local strategies shows that average search rate is enough to obtain proper reliability as result of that few agents are in failure. In general the negative sign of search rate factor presented in table3 indicates reduced access rate of the replications. We present table4 for reliability this is improved reliability when compare to table1, as for working group of agent's replica have been number consistent initiatives to integrate with fault tolerance model.

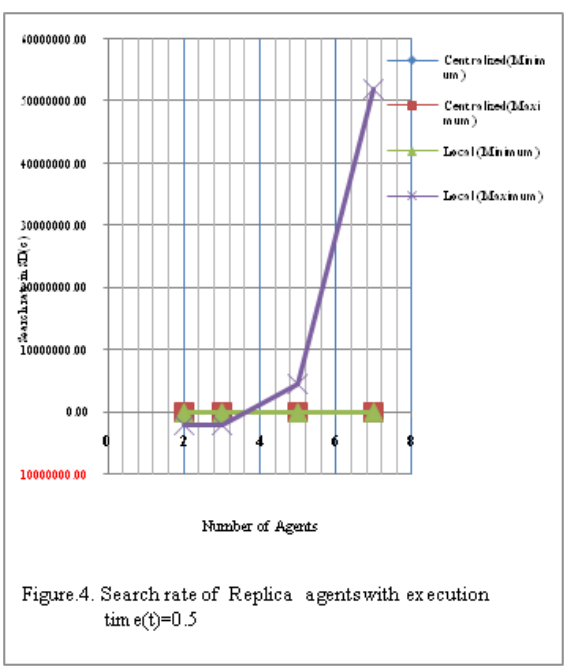

Figure.4. Search rate of Replica agents with execution time(t)=0.5

The replicas with centralized and local strategies shows that average search rate is enough to obtain proper reliability ,Therefore possibility of dynamic scheduling can be adopted in which information of resource usage changes dynamically. This kind of progressive technique can be helpful where mapping occurs on tasks to resources.

IV. CONCLUSIONS

In this work, we have proposed agent metrics for data grid replicas using localized and centralized agent objects search rate, which predicts reliable formation of agents and objects of scalable architecture techniques. The system complexity is reduced by maintaining more than one state of replicas with objects. The applicability of the designed model is shown with replicas reliable search rate, CPU processing capabilities. The achieved model has high impact on the fault recovery with available replicas where frequent failures occur, QOS in Job execution, resource utilization and Memory they use. This strategy will helpful when long run of jobs. Furthermore, we intend to apply our model to reduce the load on the grid systems and data access services. Both Centralized and local replicas are merely achieved reliability, hence we guaranteed that Replicas scalability is always high with this proposed model. In future, we plan to implement workflow model approach into our proposed strategies.

Acknowledgment

We deeply thank Dr. Ch D.V. Subbarao for his valuable help. The authors would like to thank the anonymous peer reviewers for their suggestions that aided in improving this paper.

REFERENCE

[1] NA Kofahi,S. Al-Bokhitan,A. Al-Nazer,On Disk-based and Diskless Checkopinting for parallel and distributed Systems:An Empirical Analysis-Information Technology Journal,2005. | [2] Foster I, Kesselman C.The grid: blue print for a new computing infrastructure [M]. San Francisco, USA: Morgan Kaufman Publishers, 1999. | [3] J.Bent,G.Gibson,G.Grider,B.McClelland,P.Nowoczynski,J.Nunez and M.Wingate,Plfs:A checkpoint filesystem for parallel applications in Super Computing Conference,Portland 2009. | [4] D.Li, C.Li, N.Xiao ,et al . Dynamic self-adaptive replica location method in data grids[J].Journal of Computer Research and Development,2003,40(12):1775-1780. | [5] H. Lamehamedi , B. Szymanski, Z. shentu, and E. Deelman , "Data replication strategies in grid environments, " in Proceedings of the Fifth International Conference on Algorithms and Architectures for Parallel Processing , 2002, pp.378-383. | [6] R.D.Prisco, B.W.Lampson, N.A.Lynch. Revisiting the PAXOS algorithm,Theor. Comput.Sci.,243(1-2):35-91.2000 | [7] Jose M. Perez, Felix Gracia-Carballeira, Jesus Carretero, Alejandro Calderona and Javier Fernandeza, "Branch replication scheme: A new model for data replication in large scale data grids", Future Generation Computer Systems, Vol.26, No.1,,pg12-20,Jan2010. | [8] Gao,M.Dahlin,A.Nayate,J.Zheng and A.Iyengar,"Improving Availability and Performance with Application-specific Data replication",IEEE Trans.Knowledge and Data Engineering,Vol.17,No.1,pp.106-200,2005.. | [9] A.Noraziah,M. MatDeris,N.A.Ahmed,R.Northayati,M.Y.Saman,M.A.Zeyad,"Preserving data Consistency through Neighbor replication on Grid Daemon",American Journal of Applied Sciences,Science Publications,Vol.4,No.10,pp.748-755,2007. | [10] Gao, M.Dahlin, A.Nayate, J.Zheng and A.Iyengar,"Improving Availability and Performance with Application-Specific Data Replication", IEEE Trans.Knowledge and Data Engineering, Vol.17, No.1, pp.106-200, 2005. | [11] M.Tang, B.S.Lee, X.Tang and C.K.Yeo"The impact on data replication on Job Scheduling Performance in the Data Grid" International Journal of Future Generation of Computer Systems, Elsevier (22), pp254-268, 2006. | [12] P. SunilGavaskar and Dr. Ch. D. V. SubbaRao " Fault tolerance Agents in Grid Environment ", International Journal of Advanced Research in Computer and Communication Engineering Volume2.Issue 10,October 2013,pp 3980-3982. | [13] Wenbing Zhao, "Design and implementation of a Byzantine Fault tolerance framework for web Web services", Journal of Systems and Software, Volume 82, Issue6, pp.1004-1015, June2009. | [14] C.Chen and W.Zhou,"A Solution for fault-Tolerant in Replicated Database systems", Springer-Verlang Berlin Heidelberg, pp.411-422, 2003. | [15] Zehui Maoa, Bin Jianga, and Peng Shi,"Observer based Fault tolerant control for a class of nonlinear networked control systems",Journal of the Franklin Institute, Article in Press, available online on doi:10.1016/j.jfranklin.2010.01.007. | [16] Dorian C. Arnold, Dieter Bachmann, and Jack Dongarra, "request Sequencing: Optimizing Communication for the Grid" A. Bode et al., Ed.berlin, Germany: Springer-verlag, 2000.See,alsoHttp://wwwhep.uta.edu/hep_notes/computing/computing_0017.pdf. |