

{tag}

{/tag}

IJCA Proceedings on International Conference
on Distributed Computing and Internet Technology 2014

© 2013 by IJCA Journal

ICDCIT 2014

Year of Publication: 2013

Authors:

Shailaja Pandey

Ashu A

K. Hemant K Reddy

{bibtex}icdcit1305.bib{/bibtex}

Abstract

Now a day's replication is an effective approach to improve the efficacy of distributed system, where large amount of data (terabytes or peta-bytes) is handled. An efficient replica technique is more effective than a shared distributed system (network attached storage, object based storage and storage area network) and common access point. In a distributed system, data access time depends on unreliable network bandwidth especially in desktop grid. The data

transfer is a major bottleneck in data intensive distributed grid environment due to high latency and low and unreliable bandwidth. In such an environment, an effective scheduling and effective replica technique can reduce the amount of data transfer across the internet by dispatching a job to a node where the required data are present for its operation. As the computing scale and the amount of data involved in grid applications is increasing exponentially, which causes grid resources to wait for long time period for data transfer when the involved data is saved in the remote nodes. This degrades the overall system performance. Using the file sharing mechanism in a distributed file system with a replica technique or by using a nature inspired meta-heuristic optimization technique system performance can be improved. In case of file sharing mechanism with replication techniques data can be processed in parallel. In this paper we proposed a novel combined model for data replication and job scheduling for the desktop grid environment. A reliability based replica management technique is proposed for the distributed grid environment in such way that overall data transfer is minimized. An adaptive technique is proposed for job scheduling which considers the parameters like node efficiency value, past execution history from execution log and node locality value (is a weighted parameter, depending upon the availability of replica).

Refer

ences

- M. Tang, B. -S. Lee, X. Tang, C. -K. Yeo, The impact of data replication on job scheduling performance in the data grid, *Future Generation Computer Systems* 22 (2006) 254–268
- J. Frey, T. Tannenbaum, M. Livny, I. Foster, S. Tuecke, Condor-G: A computation Management agent for multi-institutional grids, *Cluster Computing* 5 (2002) 237–246.
- P. Andreetto, S. Borgia, A. Dorigo, A. Gianelle, M. Mordacchini, M. Sgaravatto, L. Zangr, S. Andreozzi, V. Ciaschini, C. Di Giusto, F. Giacomini, V. Medici, E. Ronchieri, V. Venturi, Practical approaches to grid workload & resource management in the EGEE project, in: *Proceedings of the Conference on Computing in High Energy and Nuclear Physics, CHEP*’04, 2004, pp. 899–902.
- H. Jin, X. Shi, W. Qiang, D. Zou, An adaptive meta-scheduler for data-intensive applications, *International Journal of Grid and Utility Computing* 1 (2005) 32–37.
- I. Foster, K. Ranganathan, Design and evaluation of dynamic replication strategies for high performance data grids, in: *Proceedings of International Conference on Computing in High Energy and Nuclear Physics, Beijing, China, September 2001*.
- I. Foster, K. Ranganathan, Identifying dynamic replication strategies for high performance data grids, in: *Proceedings of 3rd IEEE/ACM International Workshop on Grid Computing*, in: *Lecture Notes on Computer Science*, vol. 2242, Denver, USA, 2002, pp. 75–86.
- I. Foster, K. Ranganathan, Decoupling computation and data scheduling in distributed data-intensive applications, in: *Proceedings of the 11th IEEE International Symposium on High Performance Distributed Computing, HPDC-11*, IEEE, CS Press, Edinburgh, UK, 2002, pp. 352–358.
- J. Basney, M. Livny, P. Mazzanti, Utilizing widely distributed computational resources efficiently with execution domains, *Computer Physics Communications* 140 (2001) 246–252.

- E. Deelman, H. Lamahemedi, B. Szymanski, S. Zujun, Data replication strategies in grid environments, in: Proceedings of 5th International Conference on Algorithms and Architecture for Parallel Processing, ICA3PP'2002, IEEE Computer Science Press, Beijing, China, 2002, pp. 378–383.
- H. H. Mohamed, D. H. J. Epema, An evaluation of the close-to-files processor and data co-allocation policy in multiclusters, in: 2004 IEEE International Conference on Cluster Computing, IEEE Society Press, San Diego, California, USA, 2004, pp. 287–298
- A. Anjum, R. McClatchey, A. Ali, I. Willers, Bulk scheduling with the DIANA scheduler, IEEE Transactions on Nuclear Science 53 (2006) 3818–3829.
- R. McClatchey, A. Anjum, H. Stockinger, A. Ali, I. Willers, M. Thomas, Data intensive and network aware (DIANA) grid scheduling, Journal of Grid Computing 5 (2007) 43–64.
- GILDA, Visited 2011. <https://gilda.ct.infn.it/>.
- CERN, Visited 2011, Compact Muon Solenoid (CMS). <http://public.web.cern.ch/public/en/lhc/CMS-en.html>.
- J. Taheri, Y. C. Lee, A. Y. Zomaya, Simultaneous job and data allocation in grid environments, The University of Sydney, Sydney, Australia, TR 6712011.
- E. Deelman, H. Lamahemedi, B. Szymanski, S. Zujun, Data replication strategies in grid environments, in: Proceedings of 5th International Conference on Algorithms and Architecture for Parallel Processing, ICA3PP'2002, IEEE Computer Science Press, Beijing, China, 2002, pp. 378–383.
- N. N. Dang, S. B. Lim, Combination of replication and scheduling in data grids, International Journal of Computer Science and Network Security (IJCSNS) 7 (2007) 304–308.
- L. Tseng, Y. Chin, and S. Wang, "The anatomy study of high performance task scheduling algorithm for grid computing system," Computer Standards and Interfaces, vol. 31, no. 4, pp. 713 – 722, 2009.
- O. Cordon, F. Herrera, and P. Villar, "Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base," Fuzzy Systems, IEEE Transactions on, vol. 9, no. 4, pp. 667–674, Aug 2001.
- S. Abdi, S. Mohamadi, Two level job scheduling and data replication in data grid, International Journal of Grid Computing & Applications (IJGCA) 1 (2010) 23–37.
- Reddy, Hemant Kumar, Manas Patra, and Diptendu Sinha Roy. "Adaptive execution and performance tuning of parallel jobs in computational desktop grid using GridGain. " Parallel Distributed and Grid Computing (PDGC), 2012 2nd IEEE International Conference on. IEEE, 2012.
- Reddy, K. Hemant K. , et al. "An Adaptive Scheduling Mechanism for Computational Desktop Grid Using GridGain. " Procedia Technology 4 (2012): 573-578.

Index Terms

Computer Science

Distributed System

Keywords

Heuristics Data Replication Adaptive Desktop Grid Distributed Systems Reliability.