C-Grid: Enabling iRODS-based Grid Technology for Community Health Research

Nitin Sukhija[1], and Arun K. Datta[2*]

[1]Mississippi State University, CAVS, Starkville, MS 39759, USA
[2]National University (NUCRI), La Jolla, CA 92037, USA
nitin@cavs.msstate.edu, adatta@nu.edu

Abstract. A Community Grid web portal, C-Grid, has been developed in this study for storing, managing and sharing large amounts of distributed community health related data in a data grid, thus facilitating further analysis of these datasets by health researchers in a collaborative environment. Remote management of this data grid is performed using the middleware iRODS, the Integrated Rule-Oriented Data System. A PHP-based wrapper, ez-iRODS, has been created as a component of C-Grid to interact with this middleware through PRODS, a client application programming interface (API). C-Grid serves as a gateway for the XSEDE resources, and also helps the users via ez-iRODS to create and manage 'virtual data collection' that can be stored in heterogeneous data resources across the distributed network. This web-based system has been developed with an objective of long-term data preservation, unified data access and sharing domain specific data amongst the scientific research collaborators of myCHOIS project.

Keywords: iRODS, cyberinfrastructure, data grid, XSEDE, community health, grid computing, data integration, portal, virtual, obesity, mobile technology.

1 Introduction

In collaboration with various Community Health Organizations, we are working on health-IT as a solution to the burgeoning problems in healthcare industry, particularly to address Childhood Obesity [1-4], which has become a national concern in US. Recent survey suggests almost 17% (or 12.5 million) of children and adolescents aged 2—19 years are obese [5]. The economic impact of this condition is staggering: in 2008 alone, it has been estimated that more than 147 billion dollars were spent just in the United States for medical costs related to obesity [6]. President Obama has recently established a task force to address this issue (White house, Office of the Press Secretary, February 09, 2010). Nevertheless, the escalating cost of providing healthcare, particularly to the patients with chronic health problems, is draining the resources of this nation. Studies suggest that the health-IT can improve the quality of

---

* For correspondence. Preliminary work was presented at the 8th IEEE International Conference on eScience 2012, October 8 – 12, Chicago (IL), USA.

4 Conclusion

In this paper, we report on a web-based data grid system, termed Community Grid (C-Grid) for community health research, which is developed as a portlet of myCHOIS [1] utilizing grid technology to manage and store health related data. Among the various data management systems, we used iRODS because of multiple advantages as elaborated by Chiang and colleagues [35]. It has been implemented to serve as a distributed computing environment and data management system for sharing resources, data and computing power with the collaborators. It provides the collaborative data sharing and maintenance of distributed storage resource collections. The ez-iRODS of C-Grid conveniently control and interact with the iRODS of the data grid located at the San Diego Supercomputer Center (SDSC). This system utilizing the Data Grid Technologies provides a long-term data preservation and allows user community to access valuable data objects conveniently through the user-friendly intuitive user interface from anywhere. The ez-iRODS simplifies the complicated operating steps and approaches of the iRODS services for the users. This is now deployed for the DHS in Illinois and Operation Samahan in California.

References

1. Datta, A.K., Jackson, V., Nandkumar, R., Sproat, J., Zhu, W., Krahling, H.: CHOIS: Enabling grid technologies for obesity surveillance and control. In 'Healthgrid Applications and Core Technologies (Eds., T. Solomonides, I. Blanquer, V. Breton, T. Glatard, and Y. Legre), vol 159, p191-202, IOS Press, Washington D.C. (2010).
2. Datta, A.K., Jackson, V., Nandkumar, R., Zhu, W.: Cyberinfrastructure for CHOIS- a global health initiative for obesity surveillance and control. Proceedings in the PRAGMA 18, 3-4 (2010).
3. Datta, A.K., Sumargo, A., Jackson, V., Dey, P.P.: mCHOIS: An application of mobile technology for childhood obesity surveillance. Procedia Computer Science 5, 653-660 (2011).
4. Datta, A.K., J.V., Rimmer, J.: Informatics challenges and opportunities for childhood obesity research. Proceedings of Public Health Informatics Conference (PHIN 2011), August 20-24, 2011, http://cdc.confex.com/cdc/phi2011/webprogram/Session12554.html; retrieved on June 4, 2013.

5. Center for Disease Control, http://www.cdc.gov/obesity/data/childhood.html; retrieved on June 4, 2013.

6. Wang, Y., Beydoun, M.A., Liang, L., Caballero, B., Kumanyika, S.K.: Will all americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic. Obesity 16(10), 2323-2330 (2008).

7. Berner, E.S.: Clinical decision support systems: State of the art. ahrq. In: Centers for Disease Control. Publication No.09-0069-EF. Rockville, Maryland: Agency for Healthcare Research and Quality (June, 2009).

8. The Health Care Law. http://www.healthcare.gov/law/; retrieved on June 4, 2013.

9. HIMSS, http://www.himss.org; retrieved on June 4, 2013.

10. Blumenthal, D., Glaser, J.P.: Information technology comes to medicine. New England Journal of Medicine 356(24), 2527-2534 (2007).

11. Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., Taylor, R.: Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. Health A airs 24(5), 1103-1117 (2005).

12. Homann, L.: Implementing electronic medical records. Communications of the ACM, 52(11), 18-20 (2009).

13. The case for Personalized Medicine, http://www.ageofpersonalizedmedicine.org/objects/pdfs/Case_for_PM_3rd_edition.pdf; retrieved on June 4, 2013.

14. Weston, A.D., Hood, L.: Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. Journal of proteome research, 3(2), 179-196 (2004).

15. Hey, A.J., Trefethen, A.E.: The data deluge: An e-science perspective (2003).

16. Chen, K.Y., Janz, K.F., Zhu, W. and Brychta, R.J.: Redefining the roles of sensors in objective physical activity monitoring. Med Sci Sports Exerc. 2012 January; 44(1 Suppl 1): S13–S23.

17. Foster, I., Kesselman, C.: The grid 2: Blueprint for a new computing infrastructure. Morgan Kaufmann (2003).

18. a. General Parallel File System, http://www-03.ibm.com/systems/software/gpfs/; retrieved on June 4, 2013; b. Lustre, http://wiki.lustre.org/index.php/Main_Page; retrieved on June 4, 2013.

19. Chervenak, A., Deelman, E., Foster, I., Guy, L., Hoschek, W., Iamnitchi, A., Kessel- man, C., Kunszt, P., Ripeanu, M., Schwartzkopf, B., Stockinger, H., Stockinger, K., Tierney, B.: Giggle: A framework for constructing scalable replica location services. In: Supercomputing, ACM/IEEE 2002 Conference. p. 58 (nov 2002).

20. Schmuck, F., Haskin, R.: Gpfs: A shared-disk file system for large computing clusters. In: Proceedings of the First USENIX Conference on File and Storage Technologies. pp. 231-244 (2002).

21. Cu , J.A., C.G.C.T., Rae, M.: In: The Ensembl computing architecture. Genome Res. vol. 14, pp. 971-975 (2004).

22. Baru, C., Moore, R., Rajasekar, A., Wan, M.: The sdsc storage resource broker. In: Proceedings of the 1998 conference of the Centre for Advanced Studies on Collaborative research. pp. 5. CASCON '98, IBM Press (1998); Also at, http://dl.acm.org/citation.cfm?id=783160.783165; retrieved on June 4, 2013.

23. Rajasekar, A., Moore, R.,Wan, M., Schroeder, W., Hasan, A.: Applying rules as policies for large-scale data sharing. In: Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference on. pp. 322-327. IEEE (2010).

24. Foster, I., Kesselman, C.: Globus: A metacomputing infrastructure toolkit. International Journal of High Performance Computing Applications 11(2), 115-128 (1997).

25. Grimshaw, A.S., Wulf, W.A., et al.: The legion vision of a worldwide virtual computer. Communications of the ACM 40(1), 39-45 (1997).

26. Hoschek, W., Jaen-Martinez, J., Samar, A., Stockinger, H., Stockinger, K.: Data management in an international data grid project. Grid ComputingGRID 2000, pp. 333-361 (2000).

27. Huber, V.: Unicore: A grid computing environment for distributed and parallel computing. Parallel Computing Technologies pp. 258-265 (2001).

28. Rajasekar, A., Moore, R., Hou, C.y., Lee, C.A., Marciano, R., de Torcy, A., Wan, M., Schroeder, W., Chen, S.Y., Gilbert, L., et al.: irods primer: integrated rule-oriented data system. Synthesis Lectures on Information Concepts, Retrieval, and Services 2(1), 1-143 (2010).

29. Southern California Earthquake Center, http://www.scec.org/; Retrieved on June 4, 2013.

30. Biomedical Informatics Research Network, http://www.birncommunity.org/; retrieved on June 4, 2013.

31. GEON, http://www.geongrid.org/; retrieved on June 4, 2013.

32. Moore, R.W., Rajasekar, A., Wan, M.: Data grids, digital libraries, and persistent archives: An integrated approach to sharing, publishing, and archiving data. Proceedings of the IEEE 93(3), 578-588 (2005).

33. Franklin, K., Datta, A.K. Introduction to HASTAC. Cyberinfrastructure for Humanities, Arts Social Sciences Summer Institute. San Diego Supercomputing Center, University of California San Diego. La Jolla, California (July 24-28 2006).

34. Hedges, M., Blanke, T., Hasan, A.: Rule-based curation and preservation of data: A data grid approach using irods. Future Generation Computer Systems 25(4), 446-452 (2009).

35. Chiang, G.T., Clapham, P., Qi, G., Sale, K., Coates, G.: Implementing a genomic data management system using irods in the wellcome trust sanger institute. BMC Bioinformatics 12(1), 361 (2011); also at, http://www.biomedcentral.com/1471-2105/12/361; retrieved on June 4, 2013.

36. Hunich, D., Muller-Pfeerkorn, R.: Managing large datasets with iRODS - a performance analysis. In: Computer Science and Information Technology (IMCSIT), Proceedings of the 2010 International Multiconference on. pp. 647-654. IEEE (2010).

37. Globus, http://www.globus.org/toolkit/security/index.html; retrieved on June 4, 2013.

38. Kerberos, http://web.mit.edu/kerberos/; retrieved on June 4, 2013.

39. Jargon, https://www.iRODS.org/index.php/Jargon; retrieved on June 4, 2013.

40. Prods, https://www.irods.org/prods_doc/; retrieved on June 4, 2013

41. Pyrods, http://code.google.com/p/irodspython/wiki/PyRods; retrieved on June 4, 2013.

42. iRODS. https://www.irods.org/; retrieved on June 4, 2013.

43. Health Information Privacy, http://www.hhs.gov/ocr/privacy/; retrieved on June 4, 2013.

44. Family Educational Rights and Privacy Act (FERPA), http://www.ed.gov/policy/gen/guid/fpco/ferpa/index.html; retrieved on June 4, 2013.

45. Foster, I., Kesselman, C., Nick, J.M., Tuecke, S.: The physiology of the grid. Grid computing: making the global infrastructure a reality pp. 217-249 (2003).

46. NHGRI, http://www.genome.gov/10001504; retrieved on June 4, 2013.

47. Hammond, R.A.: Peer reviewed: complex systems modeling for obesity research. Preventing chronic disease 6(3) (2009).

48. Fielding, R.T., Taylor, R.N.: Principled design of the modern web architecture. ACM Transactions on Internet Technology (TOIT) 2(2), 115-150 (2002).

49. Salje, E., Artacho, E., Austen, K., Bruin, R., Calleja, M., Chappell, H., Chiang, G.T., Dove, M., Frame, I., Goodwin, A., et al.: Escience for molecular-scale simulations and the eminerals project. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 367(1890), 967{985 (2009).

50. Kawai, Y., Iwai, G., Sasaki, T., Watase, Y.: Saga-based application to use resources on dierent grids. In: Proc. International Symposium on Grids and Clouds (ISGC), in series Proceedings of Science, Taipei, Taiwan (2011).