

P.O.Box 8718, Beijing 100080, China	Journal of Software, Nov. 2006,17(11):2302-2313
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2006 by <i>Journal of Software</i>

## 支持多领域动态数据集成的数据库网格系统

申德荣, 于 戈, 聂铁铮, 寇 月

[Full-Text PDF](#) [Submission](#) [Back](#)

申德荣, 于 戈, 聂铁铮, 寇 月

(东北大学 信息科学与工程学院, 辽宁 沈阳 110004)

作者简介: 申德荣(1964—), 女, 辽宁铁岭人, 博士, 教授, CCF高级会员, 主要研究领域为数据网格, Web服务. 于戈(1962—), 男, 教授, 博士生导师, CCF高级会员, 主要研究领域为数据流, 数据挖掘, 分布式数据库. 聂铁铮(1980—), 男, 博士生, 主要研究领域为分布式数据处理. 寇月(1980—), 女, 博士生, 主要研究领域为数据网格.

联系人: 申德荣 Phn: +86-24-83687776, Fax: +86-24-23895654, E-mail: shenderong@ise.neu.edu.cn, http://www.neu.edu.cn

Received 2006-06-10; Accepted 2006-08-25

### Abstract

With the richness of common database resources, distributed users in wide areas hope to transparently access and use these data resources on demand. DS\_Grid (database grid) is an SOA (service-oriented architecture) based database grid system for data sharing in multiple application domains. DS\_Grid adopts a P2P (peer-to-peer) Multi-Chord (MultiChord) architecture to realize the distributed storage, query processing and dynamic data integration of data resources. According to the text similarity, the data resources are registered to the corresponding domains to realize rapidly discovering data resources. The domain ontology knowledge and reasoning rules are used to support the semantics based intelligent query. A multi-root and multi-peer maintenance based data resources replica management mechanism is applied to improve the reliability of the system. A keyword filter based distributed data integration strategy is adopted to reduce the communication cost. A distributed clustering technique is used to summarize the huge data information. The experiments demonstrate the feasibility and effectiveness of the key techniques of DS\_Grid.

Shen DR, Yu G, Nie TZ, Kou Y. A database grid system for multi-domain dynamic data integration. *Journal of Software*, 2006,17(11):2302-2313.

DOI: 10.1360/jos172302

<http://www.jos.org.cn/1000-9825/17/2302.htm>

### 摘要

随着公有数据库资源的丰富,广泛分布的用户希望能够按需地、透明地访问和使用这些丰富的数据资源.DS\_Grid(database grid)是一个采用SOA(service-oriented architecture)思想、支持多应用领域数据共享的数据库网格系统.系统采用一种P2P(peer-to-peer)多Chord(MultiChord)网格体系结构,实现数据资源的分布存储、查询处理和动态数据集成;基于文本相似性,可分领域地注册数据资源,实现资源的快速发现;根据领域本体知识和推理规则,实现基于语义的智能查询;采用多根节点多点维护的数据资源副本管理机制,提高系统可靠性;基于关键字过滤的数据集成策略,减少通信代价;采用分布式聚类技术,实现大数据量信息的概要显示.通过实验验证了DS\_Grid中所采用的关键技术可行性和有效性.

基金项目: Supported by the National Natural Science Foundation of China under Grant Nos.60473073, 60673139 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2003AA414210 (国家高技术研究发展计划(863))

### References:

- [1] Wang S, Zhang KL. Database system on the grid. *Computer Applications*, 2004,24(10):1-3 (in Chinese with English abstract).
- [2] Yang DH, Li JZ, Zhang WP. Join algorithm based on data grid. *Journal of Computer Research and Development*, 2004,41(10): 1848-1855 (in Chinese with English abstract).

[3] Ren H, Li ZG, Xiao N. Database grid: The multi-database built on the grid. *Computer Engineering and Applications*, 2006,42(2): 171-175 (in Chinese with English abstract).

[4] Meng XF, Zhou LX, Wang S. State of the art and trends in database research. *Journal of Software*, 2004,15(12):1822-1836 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/15/1822.htm>

[5] Watson P. Databases and the grid. 2002. [http://www.nesc.ac.uk/technical\\_papers/](http://www.nesc.ac.uk/technical_papers/)

[6] Global grid forum: Grid data service specification. 2004. <https://forge.gridforum.org/projects/dais-wg>

[7] The OGSA-DAI project. 2005. <http://www.ogsa-dai.org.uk/>

[8] Kojima I, Pahlevi SM. Design and implementation of OGSA-WebDB. 2004. <http://www.nesc.ac.uk/events/GGF10-DA/>

[9] OGSA-DQP. 2004. <http://www.ogsa-dai.org.uk/dqp/>

[10] MyGrid. 2001. <http://www.mygrid.org.uk/>

[11] Smith J, Gounaris A, Watson P. Distributed query processing on the grid. *The Int'l Journal of High Performance Computing Applications*, 2003,17(4):353-367.

[12] Comito C, Talia D. XML data integration in OGSA grids. In: Pierson JM, ed. *Data Management in Grids: 1st VLDB Workshop, DMG 2005*. Heidelberg: Springer-Verlag, 2005. 16-29.

[13] Fomkin R, Risch T. Framework for querying distributed objects managed by a grid infrastructure. In: Pierson JM, ed. *Data Management in Grids: 1st VLDB Workshop, DMG 2005*. Heidelberg: Springer-Verlag, 2005. 58-72.

[14] Porto F, Silva VFV, Dutra ML, Schulze B. An adaptive distributed query processing grid service. In: Pierson JM, ed. *Data Management in Grids: 1st VLDB Workshop, DMG 2005*. Heidelberg: Springer-Verlag, 2005. 45-57.

[15] G?res J. Towards dynamic information integration. In: Pierson JM, ed. *Data Management in Grids: 1st VLDB Workshop, DMG 2005*. Heidelberg: Springer-Verlag, 2005. 16-29.

[16] Dartgrid. 2003. <http://ccnt.zju.edu.cn/projects/dartgrid/>

[17] Scientific data grid (SDG). 2005. <http://www.sdg.ac.cn/document/sdg/>

[18] Sandholm T, Gawor J. Globus toolkit 3 core—A grid service container framework. 2003. [http://www-unix.globus.org/toolkit/3.0/ogsa/docs/qt3\\_core.pdf](http://www-unix.globus.org/toolkit/3.0/ogsa/docs/qt3_core.pdf)

[19] Stoica I, Morris R, Karger D, Kaashoek MF, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for internet applications. 2001. [http://pdos.csail.mit.edu/papers/chord:sigcomm01/chord\\_sigcomm.pdf](http://pdos.csail.mit.edu/papers/chord:sigcomm01/chord_sigcomm.pdf)

[20] Falbo RA, Guizzardi G, Duarte KC. An ontological approach to domain engineering. In: *Proc. of the 14th Int'l Conf. of Software Engineering and Knowledge Engineering*. New York: ACM Press, 2002. 351-358.

[21] Bilke A, Naumann F. Schema matching using duplicates. In: Kawada S, ed. *Proc. of the 21th Int'l Conf. on Data Engineering*. Los Alamitos: IEEE Computer Society Press, 2005. 69?-80.

[22] Madhavan J, Bernstein PA, Rahm E. Generic schema matching with cupid. 2001. <http://research.microsoft.com/~philbe/CupidVLDB01.pdf>

[23] Ng WS, Ooi BC, Tan KL, Zhou AY. PeerDB: A P2P-based system for distributed data sharing. In: Dayal U, ed. *Proc. of the 19th Int'l Conf. on Data Engineering (ICDE)*. Bangalore: IEEE Computer Society Press, 2003. 633-644.

[24] Zhao BY, Kubiawicz J, Joseph AD. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical Report, UCB/CSD-01-1141, Berkeley: University of California, 2001.

[25] Keim DA, Panse C, Schneidewind J, Sips M, Hao MC, Dayal U. Pushing the limit in visual data exploration: Techniques and applications. In: Günter A, et al., eds. KI 2003: Advances in Artificial Intelligence. Heidelberg: Springer-Verlag, 2003. 37-51.

[26] Guo QJ, Yu HB, Wu K. Research & application of distributed condition-based maintenance open system. Computer Integrated Manufacturing Systems, 2005,11(3):416-421 (in Chinese with English abstract).

附中文参考文献:

[1] 王珊,张坤龙.网络环境下的数据库系统.计算机应用,2004,24(10):1-3.

[2] 杨东华,李建中,张文平.基于数据网格环境的连接操作算法.计算机研究与发展,2004,41(10):1848-1855.

[3] 任浩,李志刚,肖依.数据库网格:基于网格的多数据库系统.计算机工程与应用,2006,42(2):171-175.

[4] 孟小峰,周龙骧,王珊.数据库技术发展趋势.软件学报,2004,15(12):1822-1836. <http://www.jos.org.cn/1000-9825/15/1822.htm>

[26] 郭前进,于海斌,徐皓.基于状态维修的开放系统研究与实现.计算机集成制造系统,2005,11(3):416-421.