Caching and Materialization for Web Databases

Caching and Materialization for Web Databases

Alexandros Labrinidis

University of Pittsburgh
USA
labrinid@cs.pitt.edu

Qiong Luo

Hong Kong University of Science and Technology
Hong Kong
luo@cse.ust.hk

Jie Xu

University of Pittsburgh
USA
xujie@cs.pitt.edu

Wenwei Xue

Hong Kong University of Science and Technology
Hong Kong
wwxue@cse.ust.hk

now
the essence of knowledge

Boston – Delft

Full text available at: http://dx.doi.org/10.1561/1900000005

Foundations and Trends® in Databases

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is A. Labrinidis, Q. Luo, J. Xu and W. Xue, Caching and Materialization for Web Databases, Foundation and Trends® in Databases, vol 2, no 3, pp 169–266, 2009

ISBN: 978-1-60198-324-4
© 2010 A. Labrinidis, Q. Luo, J. Xu and W. Xue

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Foundations and Trends® in Databases
Volume 2 Issue 3, 2009

Editorial Board

Editorial Scope

Foundations and Trends® in Databases covers a breadth of topics relating to the management of large volumes of data. The journal targets the full scope of issues in data management, from theoretical foundations, to languages and modeling, to algorithms, system architecture, and applications. The list of topics below illustrates some of the intended coverage, though it is by no means exhaustive:

- Data Models and Query Languages
- Query Processing and Optimization
- Storage, Access Methods, and Indexing
- Transaction Management, Concurrency Control and Recovery
- Deductive Databases
- Parallel and Distributed Database Systems
- Database Design and Tuning
- Metadata Management
- Object Management
- Trigger Processing and Active Databases
- Data Mining and OLAP
- Approximate and Interactive Query Processing
- Data Warehousing
- Adaptive Query Processing
- Data Stream Management
- Search and Query Integration
- XML and Semi-Structured Data
- Web Services and Middleware
- Data Integration and Exchange
- Private and Secure Data Management
- Peer-to-Peer, Sensornet and Mobile Data Management
- Scientific and Spatial Data Management
- Data Brokering and Publish/Subscribe
- Data Cleaning and Information Extraction
- Probabilistic Data Management

Information for Librarians

Foundations and Trends® in Databases, 2009, Volume 2, 4 issues. ISSN paper version 1931-7883. ISSN online version 1931-7891. Also available as a combined paper and online subscription.

Foundations and Trends® in
Databases
Vol. 2, No. 3 (2009) 169–266

DOI: 10.1561/1900000005

Caching and Materialization for Web Databases

Alexandros Labrinidis[1], **Qiong Luo**[2],
Jie Xu[3] and **Wenwei Xue**[4]

[1] *University of Pittsburgh, USA, labrinid@cs.pitt.edu*
[2] *Hong Kong University of Science and Technology, Hong Kong, luo@cse.ust.hk*
[3] *University of Pittsburgh, USA, xujie@cs.pitt.edu*
[4] *Hong Kong University of Science and Technology, Hong Kong, wwxue@cse.ust.hk*

Abstract

Database systems have been driving dynamic websites since the early 1990s; nowadays, even seemingly static websites employ a database back-end for personalization and advertising purposes. In order to keep up with the high demand fuelled by the rapid growth of the Internet, a number of caching and materialization techniques have been proposed for web databases over the years. The main goal of these techniques is to improve performance, scalability, and manageability of database-driven dynamic websites, in a way that the quality of data is not compromised. Although caching and materialization are well-understood concepts in the traditional database and networking/operating systems literature, the Web and web databases bring forth unique characteristics that warrant new techniques and approaches.

In this monograph, we adopt a data management point of view to describe the system architectures of web databases, and analyze the research issues related to caching and materialization in such architectures. We also present the state-of-the-art in caching and materialization for web databases and organize current approaches according to the fundamental questions, namely how to store, how to use, and how to maintain cached/materialized web data. Finally, we associate work in caching and materialization for web databases to similar techniques in other related areas, such as data warehousing, distributed systems, and distributed databases.

Contents

1

Introduction

Database systems have been driving dynamic websites since the early 1990s, and caching and materialization have been the major techniques to improve the performance, scalability, and manageability of such web databases. Different from a traditional database environment, the software components of a web database, including web servers, database servers, application servers, and possibly additional middleware, are largely independent from one another, even though they work together as a holistic system (Figure 1.1). Caching and materialization techniques for such web databases consider a number of issues at different parts of the system and they bring interesting challenges and opportunities.

In addition to the inherent architectural uniqueness, web databases also come with stringent demands for near-real-time performance (at the speed of thought) and ability to withstand high request volumes (e.g., due to *flash crowds*, giving caching and materialization techniques a pivotal role in such environments.

In this monograph, we look at the entire process of caching and materialization for web databases as a sequence of actions (or verbs). For each action, we identify the possible options and present

Full text available at: http://dx.doi.org/10.1561/1900000005

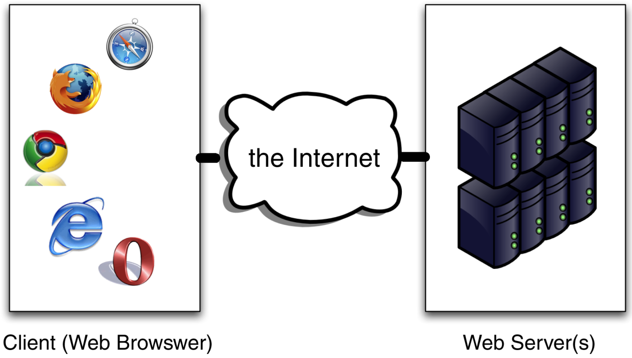

Fig. 1.1 10,000-feet view of a typical web-database architecture.

representative techniques and papers for each one. Specifically, we look into the following actions, as they relate to caching and materialization for web databases:

- **Store:** How is data cached/materialized, including where and at what granularity to cache/materialize data, and how to select the data items to cache/materialize.
- **Use:** How do we use cached/materialized data items to answer new requests?
- **Maintain:** How are caches maintained up-to-date, including when and how to handle updates, how to organize the cache to answer new requests efficiently, and how to perform cache replacement as necessary.

When presenting existing techniques on these three actions, we focus on the performance and quality aspect (in Section 6). In brief, we examine the following three types of evaluation metrics: (1) Quality of Service (QoS) metrics, including response time, throughput, and availability; (2) Quality of Data (QoD) metrics, mainly about data freshness and accuracy, and (3) user-centric metrics, usually in the form of quality contracts.

Among the above evaluation metrics, throughput is a major one to consider for web databases, and one way to achieve a high throughput is through high-performance hardware configurations. However, enhancing system throughput using high-end configurations is not always desirable; a major factor is the *Total Cost of Ownership (TCO)* [44]. TCO is the direct and hidden lifetime IT cost of purchasing, operating, managing, and maintaining a computing system [31]. Example cost items include hardware, software, network communication, administration, personnel training, and technical support. For a database-backed website, an increase in the throughput through high-end configurations may result in the increase of all those TCO cost items and in turn the TCO itself. Therefore, system throughput should be considered together with TCO. In particular, transparent and adaptive caches are often an ideal candidate to increase system throughput automatically with little increase on TCO.

Note that an important factor to consider in any caching and materialization technique for web databases is whether the scheme supports database transaction semantics, even though many web applications today work well with a weaker consistency guarantee. Eric Brewer proposed the CAP Theorem [23], which states that, for any distributed computing system, only two out of the following three key requirements can be satisfied: Consistency, Availability, and Partition tolerance. The theorem was later formally proved by Gilbert and Lynch [61] for both asynchronous and partially synchronous network models. As such, in real-world distributed web databases, where availability and scalability are crucial, it is unavoidable to make compromises on consistency. Therefore, caching and materialization are a good match to such systems in that they can help on both availability and partition tolerance and can benefit from the relaxed consistency requirement. In practice, many commercial database-backed websites today have focused their technical development, including deploying caches throughout all levels of the system, on ensuring service availability and tolerance upon data partitioning/distribution, while sacrificing consistency to various degrees. Well-known examples of such websites include Amazon, EBay, and Twitter.

Roadmap: The rest of this monograph is structured into four separate components as follows:

- **Architecture** — this part describes a typical web-database architecture and discusses some of the complexities intrinsic in such setups (Section 2).

- **Taxonomy** — this part presents the three “verbs” in detail, along with the possible alternatives. We also give a detailed description of a typical web-database architecture (Sections 3–5, followed by a discussion on metrics, in Section 6).
- **Projects** — this part describes some representative web-database projects, while explaining the choices each project adopted under the presented taxonomy (Section 7).
- **Related work** — this section presents related work from other areas (Section 8).

We conclude this monograph with a short discussion on some open problems and future directions.

References

[1] R. K. Abbott and H. Garcia-Molina, “Scheduling real-time transactions: A performance evaluation,” *ACM TODS*, vol. 17, no. 3, pp. 513–560, 1992.

[2] S. Abiteboul, J. McHugh, M. Rys, V. Vassalos, and J. L. Wiener, “Incremental maintenance for materialized views over semistructured data,” in *VLDB*, pp. 38–49, 1998.

[3] S. Adali, K. S. Candan, Y. Papakonstantinou, and V. S. Subrahmanian, “Query caching and optimization in distributed mediator systems,” in *SIGMOD Conference*, pp. 137–148, 1996.

[4] B. Adelberg, H. Garcia-Molina, and B. Kao, “Applying update streams in a soft real-time database system,” in *SIGMOD Conference*, pp. 245–256, 1995.

[5] B. Adelberg, B. Kao, and H. Garcia-Molina, “Database support for efficiently maintaining derived data,” in *EDBT*, pp. 223–240, 1996.

[6] Ajax. http://en.wikipedia.org/wiki/AJAX.

[7] Akamai. http://www.akamai.com.

[8] M. Altinel, C. Bornhövd, S. Krishnamurthy, C. Mohan, H. Pirahesh, and B. Reinwald, “Cache tables: Paving the way for an adaptive database cache,” in *VLDB*, pp. 718–729, 2003.

[9] K. Amiri, S. Park, R. Tewari, and S. Padmanabhan, “DBProxy: A dynamic data cache for web applications,” in *ICDE*, pp. 821–831, 2003.

[10] K. Amiri, S. Park, R. Tewari, and S. Padmanabhan, “Scalable template-based query containment checking for web semantic caches,” in *ICDE*, pp. 493–504, 2003.

[11] D. Andresen and T. McCune, “Towards a hierarchical scheduling system for distributed WWW server clusters,” in *HPDC*, p. 301, 1998.

[12] J. Anton, L. Jacobs, X. Liu, J. Parker, Z. Zeng, and T. Zhong, "Web caching for database applications with oracle web cache," in *SIGMOD Conference*, pp. 594–599, 2002.

[13] R. Bakalova, A. Chow, C. Fricano, P. Jain, N. Kodali, D. Poirier, S. Sankaran, and D. Shupp, "Websphere dynamic cache: Improving J2EE application performance," *IBM Systems Journal*, vol. 43, no. 2, pp. 351–370, 2004.

[14] A. Balmin, F. Ozcan, K. S. Beyer, R. Cochrane, and H. Pirahesh, "A framework for using materialized xpath views in XML query processing," in *VLDB*, pp. 60–71, 2004.

[15] G. Banga, F. Douglis, and M. Rabinovich, "Optimistic deltas for WWW latency reduction," in *USENIX Annual Technical Conference*, 1997.

[16] D. Barbara and T. Imieliski, "Sleepers and workaholics: Caching strategies in mobile environments," *ACM SIGMOD Record*, vol. 23, no. 2, pp. 1–12, 1994.

[17] P. A. Bernstein, A. Fekete, H. Guo, R. Ramakrishnan, and P. Tamma, "Relaxed-currency serializability for middle-tier caching and replication," in *SIGMOD Conference*, pp. 599–610, 2006.

[18] A. Bestavros, "Demand-based document dissemination to reduce traffic and balance load in distributed information systems," in *SPDP*, p. 338, 1995.

[19] M. Bhide, P. Deolasee, A. Katkar, A. Panchbudhe, K. Ramamritham, and P. Shenoy, "Adaptive push–pull: Disseminating dynamic web data," *IEEE Transactions on Computers*, vol. 51, no. 6, pp. 652–668, 2002.

[20] J. A. Blakeley, N. Coburn, and P.-A. Larson, "Updating derived relations: Detecting irrelevant and autonomously computable updates," *ACM Transactions on Database Systems*, vol. 14, no. 3, pp. 369–400, 1989.

[21] J. A. Blakeley, P.-Å. Larson, and F. W. Tompa, "Efficiently updating materialized views," in *SIGMOD Conference*, pp. 61–71, 1986.

[22] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *INFOCOM*, pp. 126–134, 1999.

[23] E. A. Brewer, "Towards robust distributed systems (abstract)," in *PODC '00: Proceedings of the Nineteenth Annual ACM Symposium on Principles of Distributed Computing*, p. 7, New York, NY, USA: ACM, 2000.

[24] L. Bright and L. Raschid, "Using latency–recency profiles for data delivery on the web," in *VLDB*, pp. 550–561, 2002.

[25] M. Burrows, "The Chubby lock service for loosely-coupled distributed systems," in *OSDI*, pp. 335–350, 2006.

[26] K. S. Candan, W.-S. Li, Q. Luo, W.-P. Hsiung, and D. Agrawal, "Enabling dynamic content caching for database-driven web sites," in *SIGMOD Conference*, pp. 532–543, 2001.

[27] P. Cao and S. Irani, "Cost-aware WWW proxy caching algorithms," in *USENIX Symposium on Internet Technologies and Systems*, 1997.

[28] P. Cao and C. Liu, "Maintaining strong cache consistency in the world wide web," *IEEE Transactions on Computers*, vol. 47, no. 4, pp. 445–457, 1998.

[29] P. Cao, J. Zhang, and K. Beach, "Active cache: Caching dynamic contents on the web," *Distributed Systems Engineering*, vol. 6, no. 1, pp. 43–50, 1999.

[30] C. Cappiello, M. Comuzzi, and P. Plebani, "On automated generation of web service level agreements," in *Proceedings of IEEE International conference on Advanced Information Systems Engineering (CAiSE)*, pp. 264–278, 2007.

[31] D. Cappucio, B. Keyworth, and W. Kirwin, Total cost of ownership: The impact of system management tools, 1996.

[32] V. Cate, "Alex — A global filesystem," in *USENIX File System Workshop*, pp. 1–12, 1992.

[33] S. Ceri and J. Widom, "Deriving production rules for incremental view maintenance," in *VLDB*, pp. 577–589, 1991.

[34] U. Cetintemel and P. Keleher, "Efficient distributed precision control in symmetric replication environments," in *21st IEEE Symposium on Reliable Distributed Systems (SRDS'02)*, 2002.

[35] J. Challenger, A. Iyengar, and P. Dantzig, "A scalable system for consistently caching dynamic web data," in *INFOCOM*, pp. 294–303, 1999.

[36] J.-H. Chen, R. Anane, K.-M. Chao, and N. Godwin, "Architecture of an agent-based negotiation mechanism," in *ICDCSW '02: Proceedings of the 22nd International Conference on Distributed Computing Systems*, pp. 379–384, Washington, DC, USA: IEEE Computer Society, 2002.

[37] B. Chidlovskii, C. Roncancio, and M.-L. Schneider, "Semantic cache mechanism for heterogeneous web querying," *Computer Networks*, vol. 31, no. 11–16, pp. 1347–1360, 1999.

[38] J. Cho and H. Garcia-Molina, "Synchronizing a database to improve freshness," in *SIGMOD Conference*, pp. 117–128, 2000.

[39] E. Cohen and H. Kaplan, "Aging through cascaded caches: Performance issues in the distribution of web content," in *SIGCOMM*, pp. 41–53, 2001.

[40] E. Cohen, B. Krishnamurthy, and J. Rexford, "Improving end-to-end performance of the web using server volumes and proxy filters," in *SIGCOMM*, pp. 241–253, 1998.

[41] M. Conti, M. Kumar, S. K. Das, and B. A. Shirazi, "Quality of service issues in internet web services," *IEEE Transactions on Computers*, vol. 51, no. 6, pp. 593–594, 2002.

[42] S. Dar, M. J. Franklin, B. T. Jónsson, D. Srivastava, and M. Tan, "Semantic data caching and replacement," in *VLDB*, pp. 330–341, 1996.

[43] A. Datta, K. Dutta, H. M. Thomas, D. E. VanderMeer, and K. Ramamritham, "Proxy-based acceleration of dynamically generated content on the world wide web: An approach and implementation," *ACM Transactions on Database Systems*, vol. 29, no. 2, pp. 403–443, 2004.

[44] J. S. David, D. Schuff, and R. St. Louis, "Managing your total it cost of ownership," *Communications of the ACM*, vol. 45, no. 1, pp. 101–106, 2002.

[45] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[46] J. Dean and S. Ghemawat, "Mapreduce: A flexible data processing tool," *Communications of the ACM*, vol. 53, no. 1, pp. 72–77, 2010.

[47] P. Deshpande, K. Ramasamy, A. Shukla, and J. F. Naughton, "Caching multidimensional queries using chunks," in *SIGMOD Conference*, pp. 259–270, 1998.

[48] J. Dilley, M. Arlitt, S. Perret, and T. Jin, "The distributed object consistency protocol," in *HP Labs Technical Report*, 1999.

[49] J. Dilley, B. Maggs, J. Parikh, H. Prokop, R. Sitaraman, and B. Weihl, "Globally distributed content delivery," *IEEE Internet Computing*, vol. 6, no. 5, pp. 50–58, 2002.

[50] A. Dingle and T. Partl, "Web cache coherence," in *World Wide Web Conference*, Paris, France, 1996.

[51] DNS: Domain Name System. http://en.wikipedia.org/wiki/Domain_name_system.

[52] F. Douglis, A. Feldmann, B. Krishnamurthy, and J. Mogul, "Rate of change and other metrics: A live study of the world wide web," in *USENIX Symposium on Internet Technologies and Systems*, pp. 147–158, 1997.

[53] V. Duvvuri, P. Shenoy, and R. Tewari, "Adaptive leases: A strong consistency mechanism for the World Wide Web," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 5, pp. 1266–1276, 2003.

[54] Edge Side Includes. http://www.esi.org.

[55] K. Elhardt and R. Bayer, "A database cache for high performance and fast restart in database systems," *ACM Transactions on Database Systems*, vol. 9, no. 4, pp. 503–525, 1984.

[56] C. Elkan, "Independence of logic database queries and update," in *PODS*, pp. 154–160, 1990.

[57] D. F. Ferguson, C. Nikolaou, J. Sairamesh, and Y. Yemini, *Economic Models for Allocating Resources in Computer Systems.* River Edge, NJ, USA: World Scientific Publishing Co., Inc., 1996.

[58] M. J. Franklin, M. J. Carey, and M. Livny, "Transactional client–server cache consistency: Alternatives and performance," *ACM Transactions on Database Systems*, vol. 22, no. 3, pp. 315–363, 1997.

[59] L. Gao, A. Nayate, and J. Zheng, "Improving availability and performance with application-specific data replication," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 1, pp. 106–120, 2005.

[60] J. Garreau, "Our Cells, Ourselves: Planet's fastest revolution speaks to the human heart," *The Washington Post*, February 2008.

[61] S. Gilbert and N. Lynch, "Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services," *SIGACT News*, vol. 33, no. 2, pp. 51–59, 2002.

[62] J. Goldstein and P.-Å. Larson, "Optimizing queries using materialized views: a practical, scalable solution," in *SIGMOD Conference*, pp. 331–342, 2001.

[63] C. Gray and D. Cheriton, "Leases: An efficient fault-tolerant mechanism for distributed file cache consistency," in *SOSP*, pp. 202–210, 1989.

[64] T. Griffin and L. Libkin, "Incremental maintenance of views with duplicates," in *SIGMOD Conference*, pp. 328–339, 1995.

[65] T. Griffin, L. Libkin, and H. Trickey, "An improved algorithm for the incremental recomputation of active relational expressions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 9, no. 3, pp. 508–511, 1997.

[66] J. Grundy, S. Newby, T. Whitmore, and P. Grundeman, "Extending a persistent object framework to enhance enterprise application server performance," in *ADC*, pp. 57–64, 2002.

[67] S. Guirguis, M. A. Sharaf, P. K. Chrysanthis, A. Labrinidis, and K. Pruhs, "Adaptive scheduling of web transactions," in *Proceedings of the 25th International Conference on Data Engineering*, April 2009.

[68] H. Guo, P.-A. Larson, R. Ramakrishnan, and J. Goldstein, "Relaxed currency and consistency: How to say "good enough" in SQL," in *SIGMOD Conference*, pp. 815–826, 2004.

[69] A. Gupta, D. Katiyar, and I. S. Mumick, "Counting solutions to the view maintenance problem," in *Workshop on Deductive Databases, JICSLP*, 1992.

[70] A. Gupta and I. S. Mumick, "Maintenance of materialized views: Problems, techniques, and applications," *IEEE Bulletin on Data Engineering*, pp. 145–157, 1995.

[71] A. Gupta and I. S. Mumick, eds., *Materialized Views: Techniques, Implementations, and Applications.* Cambridge, Mass: MIT Press, 1999.

[72] A. Gupta, I. S. Mumick, and V. S. Subrahmanian, "Maintaining views incrementally," in *SIGMOD Conference*, pp. 157–166, 1993.

[73] J. Gwertzman and M. Seltzer, "World wide web cache consistency," in *USENIX Annual Technical Conference*, 1996.

[74] H. Qu and A. Labrinidis, Preference-aware query and update scheduling in web-databases, April 2007.

[75] A. Y. Halevy, "Answering queries using views: A survey," *The VLDB Journal*, vol. 10, no. 4, pp. 270–294, 2001.

[76] E. N. Hanson, "A performance analysis of view materialization strategies," in *SIGMOD Conference*, pp. 440–453, 1987.

[77] J. R. Haritsa, M. Livny, and M. J. Carey, "Earliest deadline scheduling for real-time database systems," in *Proceedings of RTSS '91*, pp. 232–243, IEEE Computer Society, 1991.

[78] J. M. Hellerstein and J. F. Naughton, "Query execution techniques for caching expensive methods," in *SIGMOD Conference*, pp. 423–434, 1996.

[79] J. M. Hellerstein and M. Stonebraker, *Readings in Database Systems.* Cambridge, Mass: MIT Press, 2005.

[80] B. C. Housel and D. B. Lindquist, "WebExpress: A system for optimizing web browsing in a wireless environment," in *International Conference on Mobile Computing and Networking*, pp. 108–116, 1996.

[81] J. H. Howard, M. L. Kazar, S. G. Menees, D. A. Nichols, M. Satyanarayanan, R. N. Sidebotham, and M. J. West, "Scale and performance in a distributed file system," *ACM Transactions on Computer Systems*, vol. 6, no. 1, pp. 51–81, 1988.

[82] HTTP: Hypertext Transfer Protocol. http://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol.

[83] H. Hu, J. Xu, W. S. Wong, B. Zheng, D. L. Lee, and W.-C. Lee, "Proactive caching for spatial queries in mobile environments," in *ICDE*, pp. 403–414, 2005.

[84] P. C. K. Hung, H. Li, and J.-J. Jeng, "WS-negotiation: An overview of research issues," in *HICSS '04: Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04)*, p. 10033.2, Washington, DC, USA: IEEE Computer Society, 2004.

[85] A. Iyengar and J. Challenger, "Data update propagation: A method for determining how changes to underlying data affect cached objects on the web," in *IBM Research Technical Report*, 1998.

[86] S. Iyer, A. Rowstron, and P. Druschel, "Squirrel: A decentralized, peer-to-peer web cache," in *PODC*, 2002.

[87] M. Ji, "Affinity-based management of main memory database clusters," *ACM Transactions on Internet Technology*, vol. 2, no. 4, pp. 307–339, 2002.

[88] L.-J. Jin, V. Machiraju, and A. Sahai, "Analysis on service level agreement of web services," Technical report, HP Laboratories, 2002.

[89] J. Jing, A. Elmagarmid, A. S. Helal, and R. Alonso, "Bit-sequences: An adaptive cache invalidation method in mobile client/server environments," *Mobile Networks and Applications*, vol. 2, no. 2, pp. 115–127, 1997.

[90] P. Kalnis, W. S. Ng, B. C. Ooi, D. Papadias, and K.-L. Tan, "An adaptive peer-to-peer network for distributed caching of OLAP results," in *SIGMOD Conference*, pp. 25–36, 2002.

[91] P. Kalnis and D. Papadias, "Proxy-server architectures for OLAP," in *SIGMOD Conference*, pp. 367–378, 2001.

[92] K.-D. Kang, S. H. Son, and J. A. Stankovic, "Managing deadline miss ratio and sensor data freshness in real-time databases," *IEEE TKDE*, vol. 16, no. 10, pp. 1200–1216, 2004.

[93] A. Keller and H. Ludwig, "The WSLA framework: Specifying and monitoring service level agreements for web services," *Journal of Network and Systems Management*, vol. 11, no. 1, pp. 57–81, 2003.

[94] A. M. Keller and J. Basu, "A predicate-based caching scheme for client–server database architectures," *The VLDB Journal*, vol. 5, no. 1, pp. 35–47, 1996.

[95] D. Kossmann, M. J. Franklin, and G. Drasch, "Cache investment: Integrating query optimization and distributed data placement," *ACM Transactions on Database Systems*, vol. 25, no. 4, pp. 517–558, 2000.

[96] S. Kounev, B. Weis, and A. Buchmann, "Performance tuning and optimization of J2EE applications on the JBoss platform," *Journal of Computer Resource Management*, vol. 113, 2004.

[97] H. Kreger, Web services conceptual architecture (WSCA 1.0), 2001.

[98] B. Krishnamurthy and J. Wang, "On network-aware clustering of web clients," in *SIGCOMM*, pp. 97–110, 2000.

[99] B. Krishnamurthy and C. Wills, "Study of piggyback cache validation for proxy caches in the WWW," in *USENIX Symposium on Internet Technology and Systems*, pp. 1–12, 1997.

[100] B. Krishnamurthy and C. Wills, "Piggyback server invalidation for proxy cache coherency," in *World Wide Web Conference*, pp. 185–194, 1998.

[101] B. Krishnamurthy and C. E. Wills, "Proxy cache coherency and replacement — towards a more complete picture," in *ICDCS*, pp. 332–339, 1999.

[102] A. Labrinidis, “Web views,” in *Encyclopedia of Database Systems*, (L. Liu and M. T. Özsu, eds.), pp. 3524–3525, Springer US, 2009.

[103] A. Labrinidis, H. Qu, and J. Xu, “Quality contracts for real-time enterprises,” in *First International Workshop on Business Intelligence for the Real Time Enterprise*, 2006.

[104] A. Labrinidis and N. Roussopoulos, “Webview materialization,” in *SIGMOD Conference*, pp. 367–378, 2000.

[105] A. Labrinidis and N. Roussopoulos, “Adaptive webview materialization,” in *WebDB*, pp. 85–90, 2001.

[106] A. Labrinidis and N. Roussopoulos, “Update propagation strategies for improving the quality of data on the web,” in *VLDB*, pp. 391–400, 2001.

[107] A. Labrinidis and N. Roussopoulos, “Exploring the tradeoff between performance and data freshness in database-driven web servers,” *The VLDB Journal*, vol. 13, no. 3, pp. 240–255, 2004.

[108] P.-Å. Larson, J. Goldstein, and J. Zhou, “MTCache: Transparent mid-tier database caching in SQL server,” in *ICDE*, pp. 177–189, 2004.

[109] P.-Å. Larson and H. Z. Yang, “Computing queries from derived relations,” in *VLDB*, pp. 259–269, 1985.

[110] D. Lee and W. W. Chu, “Semantic caching via query matching for web sources,” in *CIKM*, pp. 77–85, 1999.

[111] A. Y. Levy and Y. Sagiv, “Queries independent of updates,” in *VLDB*, pp. 171–181, 1993.

[112] D. Li and D. R. Cheriton, “Scalable web caching of frequently updated objects using reliable multicast,” in *USENIX Symposium on Internet Technologies and Systems*, pp. 1–12, 1999.

[113] H. Li, S. Y. W. Su, and H. Lam, “On automated e-business negotiations: Goal, policy, strategy, and plans of decision and action,” *Journal of Organizational Computing and Electronic Commerce*, vol. 13, pp. 1–29, 2006.

[114] W.-S. Li, O. Po, W.-P. Hsiung, K. S. Candan, and D. Agrawal, “Freshness-driven adaptive caching for dynamic content web sites,” *Data & Knowledge Engineering*, vol. 47, no. 2, pp. 269–296, 2003.

[115] C. Liu and P. Cao, “Maintaining strong cache consistency in the world-wide web,” in *ICDCS*, p. 12, 1997.

[116] J. Liu, “Performance and scalability measurement of COTS EJB technology,” in *SBAC-PAD*, p. 212, 2002.

[117] T. Loukopoulos, P. Kalnis, I. Ahmad, and D. Papadias, “Active caching of on-line-analytical-processing queries in WWW proxies,” in *ICPP*, pp. 419–426, 2001.

[118] Q. Luo, S. Krishnamurthy, C. Mohan, H. Pirahesh, H. Woo, B. G. Lindsay, and J. F. Naughton, “Middle-tier database caching for e-business,” in *SIGMOD Conference*, pp. 600–611, 2002.

[119] Q. Luo and J. F. Naughton, “Form-based proxy caching for database-backed web sites,” in *VLDB*, pp. 191–200, 2001.

[120] Q. Luo, J. F. Naughton, R. Krishnamurthy, P. Cao, and Y. Li, “Active query caching for database web servers,” in *WebDB (Selected Papers)*, pp. 92–104, 2000.

[121] Q. Luo, J. F. Naughton, and W. Xue, “Form-based proxy caching for database-backed web sites: Keywords and functions,” *The VLDB Journal*, 2007.

[122] Q. Luo and W. Xue, “Template-based proxy caching for table-valued functions,” in *DASFAA*, pp. 339–351, 2004.

[123] T. Malik, R. C. Burns, and A. Chaudhary, “Bypass caching: Making scientific databases good network citizens,” in *ICDE*, pp. 94–105, 2005.

[124] T. Malik, R. C. Burns, N. V. Chawla, and A. Szalay, “Estimating query result sizes for proxy caching in scientific database federations,” in *Supercomputing*, p. 36, 2006.

[125] J. C. Mogul, “Recovery in spritely NFS,” *Computing Systems*, vol. 7, no. 2, pp. 201–262, 1994.

[126] J. C. Mogul, F. Douglis, A. Feldmann, and B. Krishnamurthy, “Potential benefits of delta encoding and data compression for http,” in *SIGCOMM*, pp. 181–194, 1997.

[127] M. N. Nelson, B. B. Welch, and J. K. Ousterhout, “Caching in the sprite network file system,” *ACM Transactions on Computer Systems*, vol. 6, no. 1, pp. 134–154, 1988.

[128] M. Nottingham, “Optimizing object freshness controls in web caches,” in *Web Caching and Content Delivery Workshop*, 1999.

[129] C. Olston and J. Widom, “Best-effort cache synchronization with source cooperation,” in *SIGMOD Conference*, pp. 73–84, 2002.

[130] Ózgúr Ulusoy and G. G. Belford, “Real-time transaction scheduling in database systems,” *Information Systems*, vol. 18, no. 9, pp. 559–580, 1993.

[131] S. Papastavrou, G. Samaras, P. Evripidou, and P. K. Chrysanthis, “A decade of dynamic web content: A structured survey on past and present practices and future trends,” *IEEE Communications Surveys and Tutorials*, vol. 8, no. 1–4, pp. 52–60, 2006.

[132] S. Patro and Y. C. Hu, “Transparent query caching in peer-to-peer overlay networks,” in *IPDPS*, p. 32, 2003.

[133] A. Pavlo, E. Paulson, A. Rasin, D. J. Abadi, D. J. DeWitt, S. Madden, and M. Stonebraker, “A comparison of approaches to large-scale data analysis,” in *SIGMOD ’09: Proceedings of the 35th SIGMOD International Conference on Management of Data*, pp. 165–178, New York, NY, USA: ACM, 2009.

[134] X. Qian and G. Wiederhold, “Incremental recomputation of active relational expressions,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 3, no. 3, pp. 337–341, 1991.

[135] M. Rabinovich and O. Spatschek, *Web Caching and Replication.* Addison-Wesley Longman Publishing Co., Inc., 2002.

[136] M. Rabinovich, Z. Xiao, F. Douglis, and C. R. Kalmanek, “Moving edge-side includes to the real edge — The clients,” in *USENIX Symposium on Internet Technologies and Systems*, 2003.

[137] A. Rajaraman, Y. Sagiv, and J. D. Ullman, “Answering queries using templates with binding patterns,” in *PODS*, pp. 105–112, 1995.

[138] L. Ramaswamy, A. Iyengar, L. Liu, and F. Douglis, “Automatic fragment detection in dynamic web pages and its impact on caching,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 859–874, 2005.

[139] L. Ramaswamy and L. Liu, "An expiration age-based document placement scheme for cooperative web caching," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 5, pp. 585–600, 2004.

[140] L. Ramaswamy, L. Liu, and A. Iyengar, "Cache clouds: Cooperative caching of dynamic documents in edge networks," in *ICDCS*, pp. 229–238, 2005.

[141] L. Ramaswamy, L. Liu, and J. Zhang, "Efficient formation of edge cache groups for dynamic content delivery," in *ICDCS*, p. 43, 2006.

[142] G. D. Rodosek and L. Lewis, "Dynamic service provisioning: A user-centric approach," in *Proceedings of the 12th International Workshop on Distributed Systems: Operations and Management (DSOM'01)*, 2001.

[143] P. Rodriguez, C. Spanner, and E. W. Biersack, "Analysis of web caching architectures: Hierarchical and distributed caching," *IEEE/ACM Transactions on Networking*, vol. 9, no. 4, pp. 404–418, 2001.

[144] N. Roussopoulos and H. Kang, "Principles and techniques in the design of ADMS±," *IEEE Computer*, vol. 19, no. 12, pp. 19–25, 1986.

[145] M. Satyanarayanan, "The evolution of Coda," *ACM Transactions on Computer Systems*, vol. 20, no. 2, pp. 85–124, 2002.

[146] A. Sawires, J. Tatemura, O. Po, D. Agrawal, and K. S. Candan, "Incremental maintenance of path-expression views," in *SIGMOD Conference*, pp. 443–454, 2005.

[147] B. Schroeder and M. Harchol-Balter, "Web servers under overload: How scheduling can help," *ACM Transactions on Internet Technology*, vol. 6, no. 1, pp. 20–52, 2006.

[148] S. Shah, K. Ramamritham, and P. J. Shenoy, "Maintaining coherency of dynamic data in cooperating repositories," *VLDB*, pp. 526–537, 2002.

[149] M. A. Sharaf, P. K. Chrysanthis, A. Labrinidis, and K. Pruhs, "Algorithms and metrics for processing multiple heterogeneous continuous queries," *ACM Transactions on Database Systems (TODS)*, pp. 5.1–5.44, March 2008.

[150] A. Sherman, P. A. Lisiecki, A. Berkheimer, and J. Wein, "ACMS: The Akamai configuration management system," in *NSDI*, pp. 245–258, 2005.

[151] B. Smith, A. Acharya, T. Yang, and H. Zhu, "Exploiting result equivalence in caching dynamic web content," in *USENIX Symposium on Internet Technologies and Systems*, 1999.

[152] G. F. Stefano Di Paola, "Subverting Ajax," in *CCC*, 2006.

[153] M. Stonebraker, D. Abadi, D. J. DeWitt, S. Madden, E. Paulson, A. Pavlo, and A. Rasin "Mapreduce and parallel DBMSs: Friends or foes?" *Communications of the ACM*, vol. 53, no. 1, pp. 64–71, 2010.

[154] M. Stonebraker, P. M. Aoki, W. Litwin, A. Pfeffer, A. Sah, J. Sidell, C. Staelin, and A. Yu, "Mariposa: A wide-area distributed database system," *The VLDB Journal*, vol. 5, no. 1, 1996.

[155] S. Y. Su, C. Huang, J. Hammer, Y. Huang, H. Li, L. Wang, Y. Liu, C. Pluempitiwiriyawej, M. Lee, and H. Lam, "An Internet-based negotiation server for e-commerce," *The VLDB Journal*, vol. 10, no. 1, pp. 72–90, 2001.

[156] A. S. Szalay, J. Gray, A. Thakar, P. Z. Kunszt, T. Malik, J. Raddick, C. Stoughton, and J. vandenBerg, "The SDSS skyserver: Public access to the sloan digital sky server data," in *SIGMOD Conference*, pp. 570–581, 2002.

[157] X. Tang and S. T. Chanson, "Coordinated management of cascaded caches for efficient content distribution," in *ICDE*, pp. 37–48, 2003.
[158] X. Tang, S. T. Chanson, H. Chi, and C. Lin, "Session-affinity aware request allocation for web clusters," in *ICDCS*, pp. 142–149, 2004.
[159] Tangosol. http://www.tangosol.com.
[160] The TimesTen Team, "In-memory data management in the application tier," in *ICDE*, p. 637, 2000.
[161] The TimesTen Team, "Mid-tier caching: The timesten approach," in *SIGMOD Conference*, pp. 588–593, 2002.
[162] P. Triantafillou and C. Neilson, "Achieving strong consistency in a distributed file system," *IEEE Transactions on Software Engineering*, vol. 23, no. 1, pp. 35–55, 1997.
[163] D. Verma, "Service level agreements on IP networks," *Proceedings of the IEEE*, vol. 92, pp. 1382–1388, 2004.
[164] D. Vista, "Optimizing incremental view maintenance expressions in relational databases," PhD thesis, University of Toronto, 1997.
[165] J. Wang, "A survey of web caching schemes for the Internet," *ACM SIGCOMM Computer Communication Review*, vol. 29, no. 5, pp. 36–46, 1999.
[166] Y. Wang and L. A. Rowe, "Cache consistency and concurrency control in a client/server DBMS architecture," in *SIGMOD Conference*, pp. 367–376, 1991.
[167] W. K. Wilkinson and M.-A. Neimat, "Maintaining consistency of client-cached data," in *VLDB*, pp. 122–133, 1990.
[168] A. Wolman, "Sharing and caching characteristics of internet content," PhD thesis, University of Washington, 2002.
[169] A. Wolman, G. Voelker, N. Sharma, N. Cardwell, M. Brown, T. Landray, D. Pinnel, A. Karlin, and H. Levy, "Organization-based analysis of web-object sharing and caching," *USENIX Symposium on Internet Technologies and Systems*, 1999.
[170] K. Worrell, "Invalidation in large scale network object caches," Master thesis, University of Colorado, Boulder, 1994.
[171] S. Wu and C.-C. Liao, "Virtual proxy servers for WWW and intelligent agents on the internet," in *HICSS*, p. 200, 1997.
[172] L. Xiao, X. Zhang, A. Andrzejak, and S. Chen, "Building a large and efficient hybrid peer-to-peer internet caching system," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 6, pp. 754–769, 2004.
[173] K. Yagoub, D. Florescu, V. Issarny, and P. Valduriez, "Caching strategies for data-intensive web sites," in *VLDB*, pp. 188–199, 2000.
[174] J. Yin, L. Alvisi, M. Dahlin, and C. Lin, "Using leases to support server-driven consistency in large-scale systems," in *ICDCS*, pp. 285–294, 1998.
[175] J. Yin, L. Alvisi, M. Dahlin, and C. Lin, "Volume leases for consistency in large-scale systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 4, pp. 563–576, 1999.
[176] H. Yu and A. Vahdat, "Efficient numerical error bounding for replicated network services," in *VLDB*, pp. 123–133, 2000.

[177] Q. Zhang, W. Sun, E. Smirni, and G. Ciardo, "Workload-aware load balancing for clustered web servers," *IEEE Transactions on Parallel and Distributed Systems*, vol. 16, no. 3, pp. 219–233, 2005.

[178] Y. Zhuge and H. Garcia-Molina, "Graph structured views and their incremental maintenance," in *ICDE*, pp. 116–125, 1998.

[179] F. Zulkernine, P. Martin, C. Craddock, and K. Wilson, "A policy-based middleware for web services SLA negotiation," in *Proceedings of the IEEE International Conference on Web Services (ICWS08)*, 2008.