

Building adaptable cache services

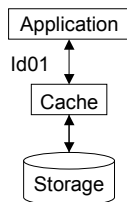
Laurent d'Orazio
Fabrice Jouanot
Cyril Labbé
Claudia Roncancio
(firstname.lastname@imag.fr)

1

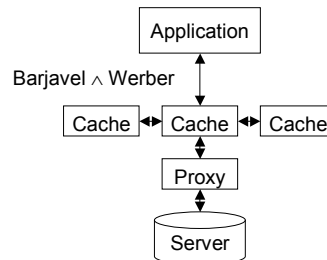
Motivations (1)

- Observations
 - Caching is used to improve performances in many applications
 - Performances result from specific behaviors of a cache
 - Grids may use caches at different levels

Persistent object manager



Web searching system



2

Motivations (2)

- Consequence
Caches are usually built from scratch for a given application
⇒ Time (and money) consuming
- Solution
Static adaptation of cache services
- Proposal
A framework for building adaptable cache services

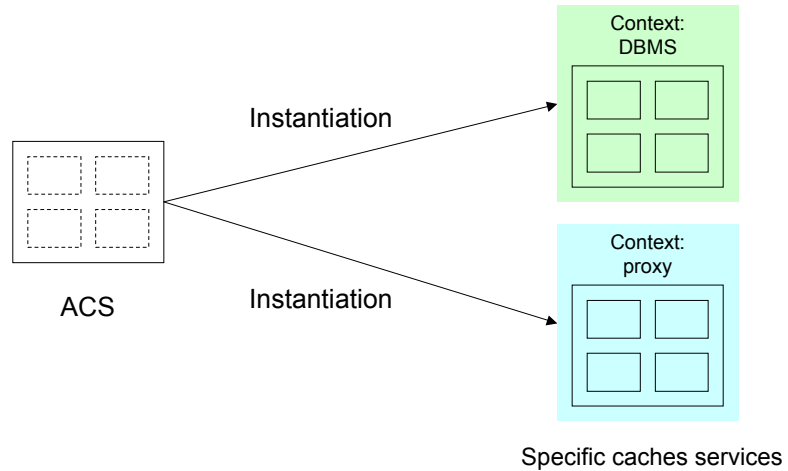
3

Adaptable cache service motivations

- Observation
Applications and system environment using a cache may change
- Consequence
Behaviors of a cache may no longer be well suited
- Solution
Dynamic adaptation of cache services
- Proposal
Using a component model to build cache services

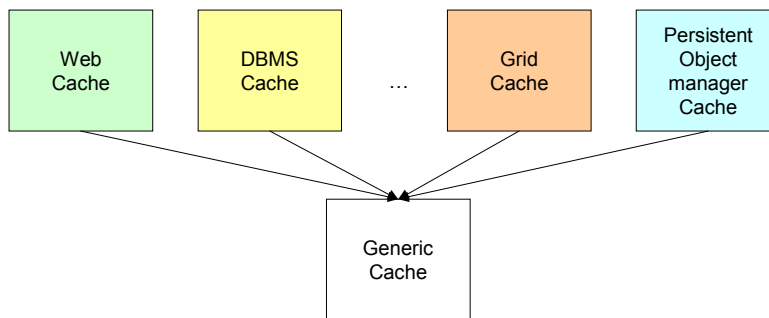
4

ACS a framework for building Adaptable Cache Services (1)



5

ACS a framework for building Adaptable Cache Services (2)



Basic concerns of a cache service

6

Plan

- Motivations
- Basic concerns of a cache service
 - Addressing
 - Searching
 - Replacement
 - Resolution
- ACS a framework for building adaptable cache services
- Using ACS in a data grid middleware
- Related work
- Conclusions and perspectives

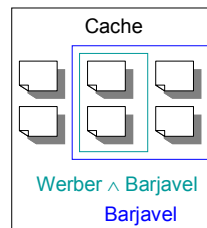
7

Addressing in the cache

- Role
Identifying an entry.
- State of the art

Identity-based	Semantic
----------------	----------

- Example
Semantic caching



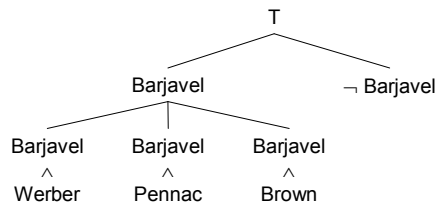
8

Searching in the cache

- Role
Finding an entry.
- State of the art

Direct Search	Indirect search					
	Translation table	Compact table			Index	
		Unsorted table	Sorted table	Chained table	Tree based	Hash table

- Example
R-tree



9

Replacement of cache entries

- Role
Choosing entries to remove when the cache is full.
- State of the art

Random	Entry moment	Use moment	Use frequency	Semantic locality	Cost / Benefit	Hybrid
--------	--------------	------------	---------------	-------------------	----------------	--------

- Example
Cost / benefit
 - Retrieval time
 - Size
 - Frequency use

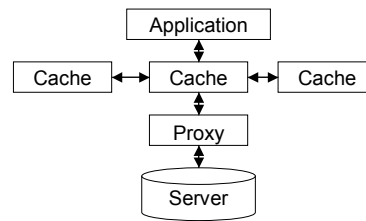
10

Cache miss resolution

- Role
Finding where to look when a cache miss occurs.
- State of the art

Direct resolution	Hierarchical resolution			
	ICP	CARP	CD	Other

- Example
Internet Cache Protocol (ICP)



11

Synthesis

- Observation
Several possibilities for each concern
- Interpretation
Each concern can be considered as an adaptation point for a cache service

12

Example 1: Caching in a persistence object manager

Addressing

Identity-based	Semantic
----------------	----------

Searching

Direct Search	Indirect search					
	Translation table	Compact table			Index	
		Unsorted table	Sorted table	Chained table	Tree based	Hash table

Replacement

Random	Entry moment	Use moment	Use frequency	Semantic locality	Cost / Benefit	Hybrid
--------	--------------	------------	---------------	-------------------	----------------	--------

Resolution

Direct resolution	Hierarchical resolution			
	ICP	CARP	CD	Other

13

Example 2: Caching in a web searching system

Addressing

Identity-based	Semantic
----------------	----------

Searching

Direct Search	Indirect search					
	Translation table	Compact table			Index	
		Unsorted table	Sorted table	Chained table	Tree based	Hash table

Replacement

Random	Entry moment	Use moment	Use frequency	Semantic locality	Cost / Benefit	Hybrid
--------	--------------	------------	---------------	-------------------	----------------	--------

Resolution

Direct resolution	Hierarchical resolution			
	ICP	CARP	CD	Other

14

Plan

- Motivations
- Basic concerns of a cache service
- [ACS a Framework for building adaptable cache services](#)
- Using ACS in a data grid middleware
- Related work
- Conclusions and perspectives

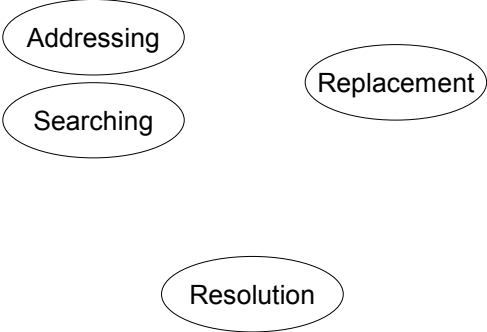
15

Architecture motivation

- Needs
Dynamic and static adaptation
- A solution: component model
Based on the separation of concerns principle, enables minimizing parts of the service to modify during an adaptation.

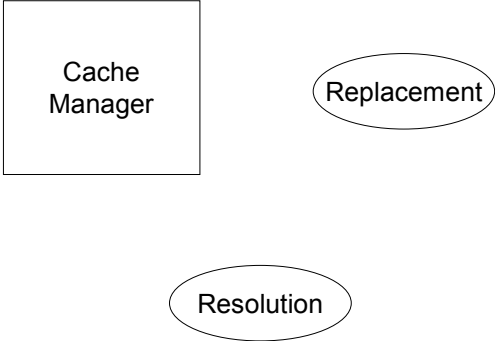
16

Basic concerns of a cache service



17

Cache manager



- Role
To Bind and find entries in the cache

18

Replacement manager



- Role
To capture the replacement strategy used to determine items to remove when the cache is full



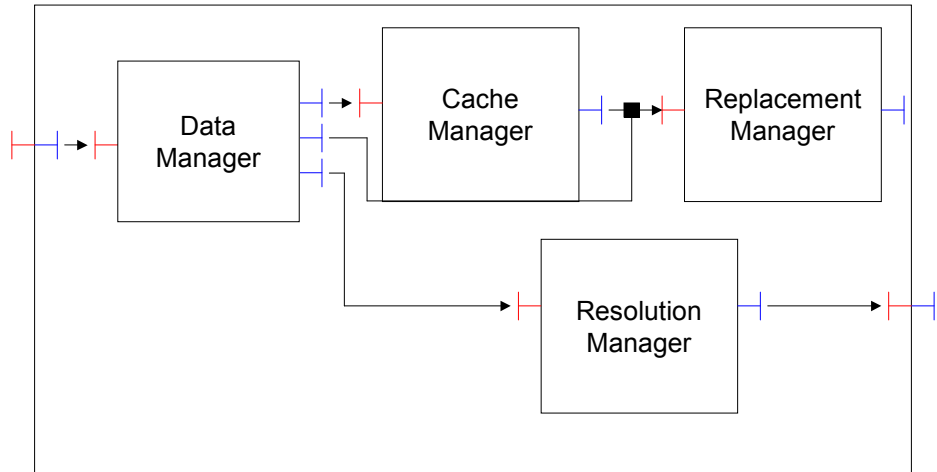
Resolution manager



- Role
To capture the resolution protocol used when a cache miss occurs

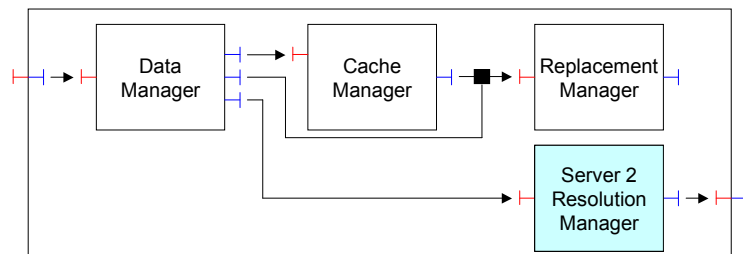
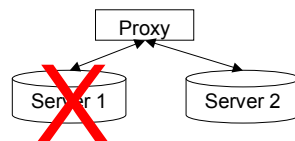


ACS Architecture



21

Adaptation of a cache service



22

Prototype

- Behaviors
 - Java
 - Fractal
- Motivations
 - Re use of Perseus cache framework
 - Fractal enables static and dynamic adaptation

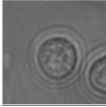
23

Plan

- Motivations
- Basic concerns of a cache service
- ACS a Framework for building adaptable cache services
- [Using ACS in a data grid middleware](#)
- Related work
- Conclusions and perspectives

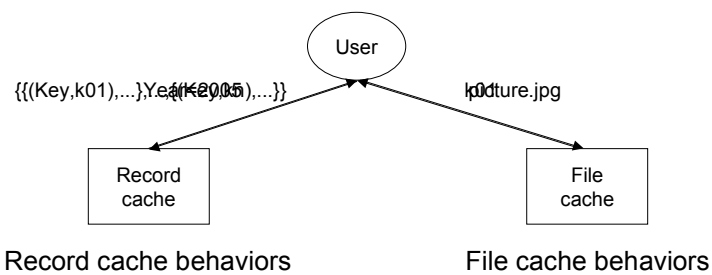
24

Using ACS in a data grid middleware

- Context
Gedeon project: middleware for data management in grids
- Entities manipulated
 - Files
Example: picture.jpg 
 - Records: a set of meta data
Example: {(Key,k01),(Author,Dupond),(Year,2005)}

25

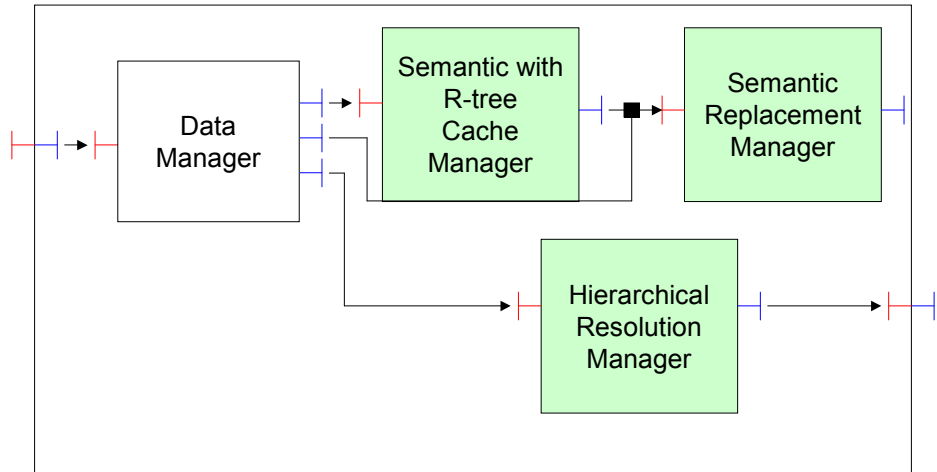
Caching in Gedeon



- Semantic addressing
- R-tree searching
- Semantic replacement
- Hierarchical resolution
- Identity based addressing
- Hash table searching
- LRU replacement
- Hierarchical resolution

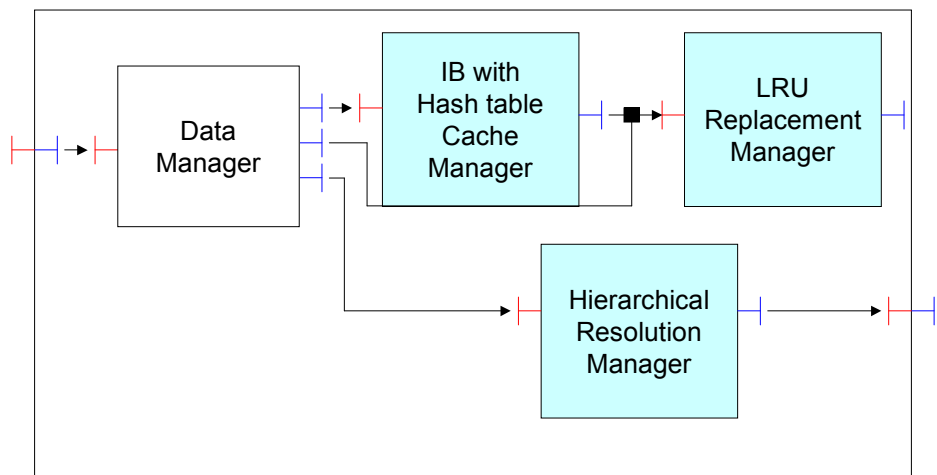
26

Record cache



27

File cache



28

Plan

- Motivations
- Basic concerns of a cache service
- ACS a framework for building adaptable cache services
- Using ACS in data grid middleware
- [Related works](#)
- Conclusions and perspectives

29

Caching

- Very rich literature on the topic
Example: some references about semantic caching
[KB96] [MS05] [RD98] [RD00] [R91] [AK94] [CB98] [CRS99]
[DRSN98] [F82] [GG97] [XZLL05]
- Context dependant caches
 - DBMS [CFLS91]
 - Web [C]
 - Mediator [AKS98]
- Caching mechanisms
 - Tuple, page and semantic caching [DFJST96]
 - Hierarchical and distributed caching [CDNSW96]
- Framework for building cache services
 - Perseus [GDC03]
 - CaLi [Z04]

30

Perseus [GDC03]

- Definition
A persistence object manager framework.
- Advantages
Separation of concerns, capturing
 - Addressing
 - Searching
 - Replacement
- Difference with ACS
Resolution not captured in the framework

31

CaLi [Z04]

- Definition
A cache framework
- Advantages
Separation of concerns, capturing
 - Addressing
 - Searching
 - Replacement
 - Distributed caching
- Difference with ACS
 - No dynamic adaptation
 - Distributed entries and communications VS resolution manager

32

Plan

- Motivations
- Basic concerns of a cache service
- ACS a framework for building adaptable cache services
- Using ACS in data grid middleware
- Related work
- [Conclusions and perspectives](#)

33

Conclusions

- Basic concerns of a cache service
 - Addressing
 - Searching
 - Replacement
 - Resolution
- Architecture of a framework
 - Static adaptation
 - Dynamic adaptation
- Prototype

34

Perspectives

- Experimentation
 - Gedeon, middleware for data management in grids
 - Record caches
 - Data caches
- Interactions with other concerns
 - Prefetching
 - Concurrency
 - Consistency

35

Thank you

Questions ?

36

References (1)

- [C] Cern server on-line. URL : <http://www.w3.org/Daemon/>.
- [Z04] J. Zola. Cali, efficient library for cache implementation. In Proc. of PCS 2004, pages 415–420, 2004.
- [CDNSW96] A. Chankhunthod, P. B. Danzig, C. Neerdaels, M. F. Schwartz, and K. J. Worrell. A hierarchical internet object cache. In USENIX Annual Technical Conf., pages 153–164, 1996.
- [GDC03] L. Garcia-Banuelos, P.-Q. Duong, and C. Collet. A Component based Infrastructure for Customized Persistent Object Management. In Proc. of the 5th Int'l WS on Parallel and Distributed Databases: innovative applications and new architectures, Czech Republic, 2003.
- [DFJST96] S. Dar, M. J. Franklin, B. Jonsson, D. Srivastava, and M. Tan. Semantic data caching and replacement. In VLDB'96: Proceedings of the 22nd Inter'l Conference on Very Large Data Bases, pages 330–341, San Francisco, CA, USA, 1996. Morgan Kaufmann Publishers Inc.
- [AKS98] N. Ashish, C. A. Knoblock, and C. Shahabi. Intelligent caching for information mediators: A kr based approach. In Workshop on Knowledge Representation meets Databases, Seattle, WA, 1998.
- [CFLS91] M. J. Carey, M. J. Franklin, M. Livny, and E. J. Shekita. Data caching tradeoffs in client-server dbms architectures. In SIGMOD '91: Proceedings of the 1991 ACM SIGMOD int'l conference on Management of data, pages 357–366, New York, NY, USA, 1991.

37

References (2)

- [AK94] Y. Arens and C. A. Knoblock. Intelligent caching: selecting, representing, and reusing data in an information server. In CIKM '94: Proceedings of the third international conference on Information and knowledge management, pages 433–438, 1994.
- [CB98] B. Chidlovskii and U. M. Borghoff. Signature file methods for semantic query caching. In European Conference on Digital Libraries, pages 479–498, 1998.
- [CRS99] B. Chidlovskii, C. Roncancio, and M.-L. Schneider. Semantic cache mechanism for heterogeneous web querying. Computer Networks, 31(11-16):1347–1360, 1999.
- [DRSN98] P. M. Deshpande, K. Ramasamy, A. Shukla, and J. F. Naughton. Caching multidimensional queries using chunks. In SIGMOD '98: Proceedings of the 1998 ACM SIGMOD international conference on Management of data, pages 259–270, 1998.
- [F82] S. Finkelstein. Common expression analysis in database applications. In SIGMOD '82: Proceedings of the 1982 ACM SIGMOD international conference on Management of data, pages 235–245, New York, NY, USA, 1982.
- [GG97] P. Godfrey and J. Gryz. Semantic query caching for heterogeneous databases. In Knowledge Representation Meets Databases, pages 6.1–6.6, 1997.
- [XZLL05] 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 '05: Proceedings of the 21st International Conference on Data Engineering (ICDE'05), pages 403–414, Washington, DC, USA, 2005.

38

References (3)

- [KB96] A. M. Keller and J. Basu. A predicate-based caching scheme for client-server database architectures. The VLDB Journal, 5(1):035–047, 1996.
- [MS05] B. Mandhani and D. Suciu. Query caching and view selection for xml databases. In VLDB '05: Proceedings of the 31st International Conference on Very Large Data Bases, pages 469–480, 2005.
- [RD98] Q. Ren and M. Dunham. Semantic caching and query processing, 1998.
- [RD00] Q. Ren and M. H. Dunham. Using semantic caching to manage location dependent data in mobile computing. In Mobi- Com '00: Proceedings of the 6th annual international conference on Mobile computing and networking, pages 210– 221, 2000.
- [R91] N. Roussopoulos. An incremental access method for viewcache: concept, algorithms, and cost analysis. ACM Trans. Database Syst., 16(3):535–563, 1991.