

Two-Level Cache Structure based on User Logs

Zhang Weizhe, He Hui, Qiu Zhijie, Zhang Deqiang

School of Computer Science and Technology, Harbin Institute of Technology, Harbin,
China. 150001
wzzhang@hit.edu.cn

Abstract. A two-stage structure of cache is proposed based on the queries of the users' logs. The highest queries are extracted by users based on static cache, in which the queries are the most popular. We adopt dynamic cache as an auxiliary to optimize the distribution of the cache data. A distribution strategy of the cache data is proposed.

Keywords: distributed retrieve; cache system; cache structure.

1 Introduction

Cache technology is a key technology to improve response speed, processing efficiency and system performance. It has an import position in the distributed text retrieval process. Sergey Melnik and Sriram Raghavan^[1] designed a pipe water to create the inverted index, it improves the efficiency of index creation. However, such measures just optimized dictionaries and indexes and other related factors.

Cache technology is widely used in various fields of computers; it can effectively solve the system bottlenecks and enhance the handling capacity. Currently, cache technology has widely used in the search engine performance improvement. It has become a key technology to enhance handling capacity and shorten the response time for search engine. In recent years, through statistical analysis of search engine user behavior, the study found that user queries follow a high degree of repeatability and locality principle^[2]. Lots of Internet users query a large number of repetitive queries, and often concentrated in certain local content. These hot contents can maintain a high number during a certain period. Thus, cache technology can be used, it can ease the pressure of search engines, and it is also one of the best ways to improve retrieval performance^[3].

2 Architecture

The structure of the cache design ideas are as follows: Within each cluster, each cache server, the cache structure using two-level cache structure, cache are consisted of the static cache and dynamic cache, static cache and dynamic cache works together to

cache data, to achieve the goal to enhance the processing speed. In entire cache structure, static cache will work at first, then dynamic cache, with the arrival of each query, our system will look up it in the static cache, if hits the cached data is returned; If it does not hit, and then access dynamic cache to see if it hits, according to different situation, cache system use different process. The static cache stores in the highest frequency query and retrieve the results. By analyzing the query log in the cluster, extraction of the hottest query, and the corresponding search results in pairs stored in the static cache initialization data as a static cache. Extract the hottest query and corresponding search results in pairs to store them in the static cache, initialize the data as a static cache. Data in static cache is relatively fixed, the content will change only when it need to rebuild the cache and replace some of the cache. The dynamic cache dynamically changed according to user queries, it stores high frequency user queries and retrieve results. The store is dynamic, and after a period of time, with the role of replacement algorithm, the content in dynamic storage is also relatively hot.

Two-level cache structure mainly consists of four parts:

- 1) **Memory buffer, to create a static cache and dynamic cache**
- 2) **With cached data distribution strategy, static cache and dynamic cache are initialized**
- 3) **Design coordination mechanism between static cache and dynamic cache**
- 4) **Static cache and dynamic cache survival, destruction mechanism**

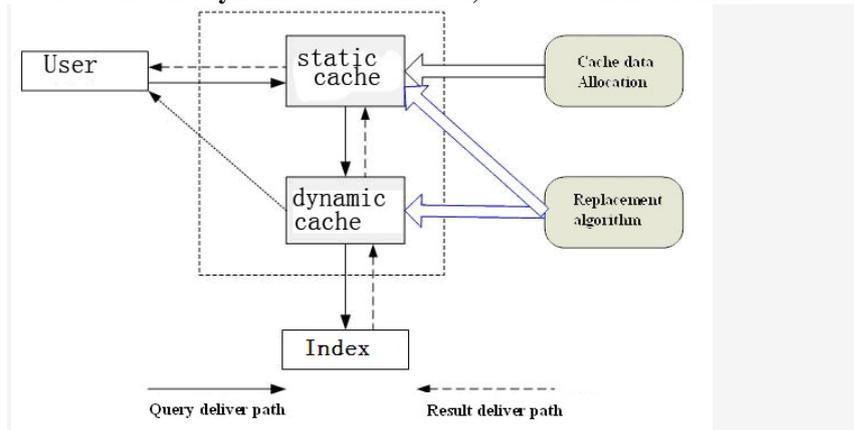


Fig.1 Coordination mechanism of static cache and dynamic cache

3 Experiments

The experiment using two different cache structures. Case 1 is using two level cache structure.

In the experiment, the cache in two-level cache structure is divided into 2 halves, 500 records in static cache and 500 records in dynamic cache. But cache with just dynamic cache holds all the 1000 records. Static cache in two-level structure will not change and the dynamic cache changes with LRU algorithm. The dynamic cache structure also changes with LRU algorithm.

Test 1: Test two-level cache structure and pure dynamic cache 2 cases. For different total queries, the result of single query processing time is show in table 1.

Tab.1 The processing time of single query

Total queries (times)	Processing time of single query by using two-level cache structure (ms)	Processing time of single query by just using dynamic cache (ms)
1000	32.3	59.6
2000	31.6	51.2
5000	39.5	54.5
10000	46.3	48.6
20000	53.1	50.3
30000	55.2	52.6
50000	50.8	50.4
80000	49.1	48.7

Acknowledgments. This work was supported by the National Natural Science Foundation of China (NSFC) under grant No. 61173145, the National High Technology Research and Development Program of China under Grant No. 2010AA012504 and No. 2011AA010705, and the National Basic Research Program of China under Grant No.G2011CB302605.

References

1. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148--1158. Springer, Heidelberg (2006)
2. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)
3. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181--184. IEEE Press, New York (2001)