Abstract

Search engines require computers with high computation resources for processing to crawl web pages and huge data storage to store billions of pages collected from the World Wide Web after parsing and indexing these pages. The indexer is one of the main components of the search engine that come intermediate between the crawler and the searcher. Indexing is the process of organizing the collected data to facility information retrieval and minimizes the time of query. Indexing requires huge processing and storage resources, and the indexing has a high effect on the performance of the search engine, this effect differs based on the structure and the process index construction. Distribution of the indexing process over a cluster of computers in grid computing will improve the performance through distributing the parsing load over a number of computers in a grid environment, and distributing the indexed data over distributed memory.
Dynamic and Distributed Indexing Architecture in Search Engine using Grid Computing

according to terms over a number of computers remotely. Due to the search engine data collections with frequent changes, the indexer require dynamic indexing. So the merge of the distributed and dynamic indexing in architecture over grid computing will give a better performance utilizing the available resources without need to computers with high cost such as supercomputers.

References

- E. Adar, J. Teevan, and S. T. Dumais, "Resonance on the Web: Web Dynamics
Index Terms

Computer Science  Information Sciences

Keywords

Indexer  World Wide Web  Search engine  Grid Computing  Web pages

Secondary index

Main index

Alchemi
Manager

Executor