Thoughts after reading Google MapReduce

It is the second paper I ‘ve read, and I feel that I can gradually understand some idea. Though I still have to often search for some difficult words, and have to search the Internet for some explanation, but I really think I’ve learned something I won’t learn from the class, and something that more classic and basic.

And I really think I should use the holiday to read it again, because it is quite useful for the future for example the company interview. And I really thank our teacher because he gives us a preliminary understanding of the software industry, and we really learned something different from him, for example some tips for code typing, including reading this three academic paper. And since I think I understand some of the paper, I will talk about it bellow.

I think the main idea of MapReduce can be summed up in four words, "breaking up the whole into parts", which is realized through map and reduce.

First of all, MapReduce seems to solve the problem that Google programmers deal with a large amount of data every day, such as website query protocol, Require computers to process different types of data, collect most requests every day, and so on.

The difficulty in solving these problems lies in the large amount of data. According to the traditional method, this will be a huge project, which requires a lot of computer processing and a lot of time and time. Therefore, Google has developed MapReduce function. Compared with the traditional communication database, MapReduce has a larger amount of data It adopts batch processing, saves the time of writing and multiple reading, and has a more flexible dynamic mode structure and wider linear space expansion. Map reduce mainly adopts the concept of division of labor and management, divides a task into several small tasks, and finally summarizes the results. Mapping functions perform and reduce by creating mappings. The mapping function divides the job into multiple tasks, scales them and summarizes the results. For example, in a bookstore, the bookshelf conforms to the drawing function, People in the bookstore should calculate and summarize the different types and quantities of books on the shelf, and tell the boss what function abbreviations are. Each grid on each shelf has a corresponding book, and each layer corresponds to a book. MapReduce inherits this idea and decomposes it layer by layer Each level corresponds to a keyword, followed by multiple values. Therefore, a large amount of data can be distributed to reduce the workload of programmers, while map reduce has less hardware requirements. Every computer can work and save a lot of money.

This is my superficial view of MapReduce. Since I really not that familiar with these basic theory, that’s what I think, and might be wrong in some way. And in the future I will not only complete the class homework, I will spend more time to learn some techniques that are used now by the company, to prepare myself better for the future work.