The three papers on big data released by Google, Google MapReduce, Google BigTable and Google file system, guide the road of big data and the necessity of developing big data.

Google MapReduce, a programming model, has long been used by a large number of programmers, and even derived a huge number and methods to deal with massive data. This model is inspired by LISP and many other functional languages. Most operations can be implemented under this programming model. The intermediate data can be merged only through the reduce operation with the same key value, so as to achieve the purpose of data processing.

As we all know, it is the application of this model that Google has the largest search engine in the world. The massive documents captured by web crawlers are indexed by MapReduce model.

MapReduce's greatest effect on programmers may be to reduce the number of programs, which is very friendly to programmers' hair. Moreover, MapReduce can reduce machine failure, slow machine processing speed and even network congestion, and improve the overall performance.

Google file system, Google File management system and windows tree file management system share the same design ideas, performance, scalability, reliability and availability, but GFS is obviously different from the early file management system. Through different choices, completely different design ideas have been derived, and GFS came into being.

GFS provides users with a form of file management. The stored files are divided into chunks of fixed size, and then the master assigns an exclusive code. At the same time, Linux files are saved on their own hard disk. Moreover, users can also connect with other chunk systems through chunks to achieve the connection between chunks.

Google BigTable, a distributed structured data storage system, has the following characteristics: wide applicability, scalability, high performance and high availability. In fact, BigTable is similar to the database. The content of the database can be realized through BigTable, and BigTable shows the scalability and high performance of the database at the same time. However, contrary to the database, BigTable is the correlation of data defined or speculated by the user. Take Google Earth for example. Google provides users with satellite images through Google Earth and Google Maps on the web page. The map requires high precision, which directly leads to the huge amount of data contained. BigTable perfectly solves this problem. BigTable stores the image through image compression, and can give users high-speed response.

Among the three carriages of Google, BigTable is the storage mode, MapReduce is the programming model, and file system is a file management mode. Through the simple reading of these three papers, I have generally understood the great development prospect of big data, and clearly realized how Google created the first year of big data. Through three papers, Google provides a bright path for future big data.