論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A Study on Data Placement and Data Management for Power-proportional Distributed File Systems
著者(和文)	LE HIEU HANH
Author(English)	LE HIEU HANH
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9920号, 授与年月日:2015年4月30日, 学位の種別:課程博士, 審査員:横田　治夫,佐伯　元司,権藤　克彦,吉瀬　謙二,金子　晴彦
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9920号, Conferred date:2015/4/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

(論文博士)

論　文　要　旨（英文）　(800語程度)

(Summary)

報告番号	乙 第　　号	氏　名	Hieu Hanh Le

（要　旨）

In the era of Big Data, distributed file systems, such as the Google File System and the Hadoop Distributed File System (HDFS), have been widely used for efficient processing over a huge amount of unstructured data. In such systems, not only high performance but also power consumption has gained much attention from both academia and industry. Among these systems, power-aware distributed file systems are progressively moving towards power-proportional design, in which the data placement methods are designed to control the total number of active nodes to store and retrieve the data. The common idea in these methods is to make the system operate in multiple gears where each gear contains a different number of groups of nodes. Here, a higher gear has a larger number of groups of active nodes, hence is able to deliver better throughput performance.

However, the current power-proportional works do not well consider the effects of gear-shifting on the performance of the system. During gear-shifting, in order to achieve desired performance at a high gear, the system may have to perform updated data reflection that re-transfers the updated data modified in a low gear when a subset of the nodes was powered off. Specifically, it needs to internally re-transfer those updated data to the reactivated nodes to share the load equally among all of the active nodes in the systems. Inefficient updated data reflection with large amounts of re-transferred data degrades the performance of power-proportional distributed file systems greatly. Furthermore, the updated data reflection is still constrained by a centralized architecture as all the metadata management is centrally maintained by a single node, e.g. in the default HDFS. In addition to normal operations, the updated data reflection increases the costs of metadata management for data transference inside the system. Carrying out this process effectively is vital in realizing power proportionality for distributed file systems.

In this Ph.D. dissertation, I introduce solutions for the problems of insufficient gear-shifting at multiple-gear power proportional distributed file systems. In the solutions, the amount of re-transferred data and the centralization management of metadata are focused.

At first, I propose a novel data placement method, named ERIGS_HDFS, which uses data replication to arrange the data layout to provide efficient gear-shifting. The ERIGS_HDFS reduces the amount of re-transferred data, which shortens the period required for reflecting updated data in gear-shifting. Empirical experiments on actual machines indicate that the ERIGS_HDFS gains more than 40% better performance than an existing comparative method.

However, as the ERIGS_HDFS lacks the flexibility to be applied in large-scale systems, I further propose a novel data placement method named Accordion. Aimed for large-scale systems, the Accordion also uses data replication to arrange the data layout for efficient gear-shifting but with more flexibility. Through carefully designing the locations of primary data and the backup data, compared with existing current methods, the Accordion reduces the amount of data transferred, shortens the period required to re-transfer the updated data during gear-shifting, then is able to improve the throughput performance of the systems. Extensive empirical experiments using actual machines with an Accordion prototype based on the HDFS demonstrate that our proposed method significantly reduces the period required for updated data reflection, i.e., by 66% compared with existing methods. Especially, the Accordion is also able to improve the throughput performance of the system during gear-shifting by more than 30%.

Thirdly, I propose a new coupling architecture of metadata management and data management, named NDCouplingHDFS, which effectively reflects the updated data when the system goes into the high gear. In the normal HDFS design, when the system changes power mode, the updated data reflection process is ineffectively restrained by a single node because of the access congestion of the metadata information of data. On the other hand, the NDCouplingHDFS at first eliminates such bottleneck by utilizing a distributed metadata management method. Then, through coupling both metadata management and data management at each node, the NDCouplingHDFS efficiently localizes the range of the updated data maintained by the metadata management. Experiments using actual machines show that the NDCouplingHDFS is able to significantly reduce the execution time required to move updated data by 46% relative to the normal HDFS. Moreover, the NDCouplingHDFS is capable of increasing the throughput of the system supporting MapReduce by applying an index in metadata management.

Finally, I present a multiple-gear distributed file system that efficiently integrates both Accordion and NDCouplingHDFS to provide high throughput performance during gear-shifting. Although the amount of re-transferred data is reduced in the Accordion, efficient metadata management is required to increase the efficiency of updated data reflection further for better power-proportional throughput performance. In the proposed system, this is achieved with support from the NDCouplingHDFS. The effect of the proposed system is optimized against changes in the amount of updated data at low gear and the load of metadata. It is observed that the proposed system gains better performance for large amounts of updated data under a heavy metadata load; and for small amounts of updated data under a light metadata load.

The works in this dissertation contribute directly for designing the power-proportional distributed file systems and lead to the elastic computing which has recently become a new trend for Cloud applications.