

Distributed Metadata Management of Mass Storage System in High Energy of Physics



Qiulan Huang¹, Ran Du¹, YaoDong Cheng¹, Wenxiao Kan^{1,2}, Jingyan Shi¹, Gang Chen¹

¹ IHEP, 19B Yuquan Road, Beijing, 100049 China

² University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049

1. Introduction

Metadata management is quite important to overall system performance in large-scale distributed storage systems, especially in the big data era. Metadata performance would produce a big effect on the scalability, availability and high performance of the massive storage system. In order to manage metadata effectively, so that data can be allocated and accessed efficiently, we design and implement a dynamic and scalable distributed metadata management system(StarFS) to HEP mass storage system based on Gluster.

2. Architecture

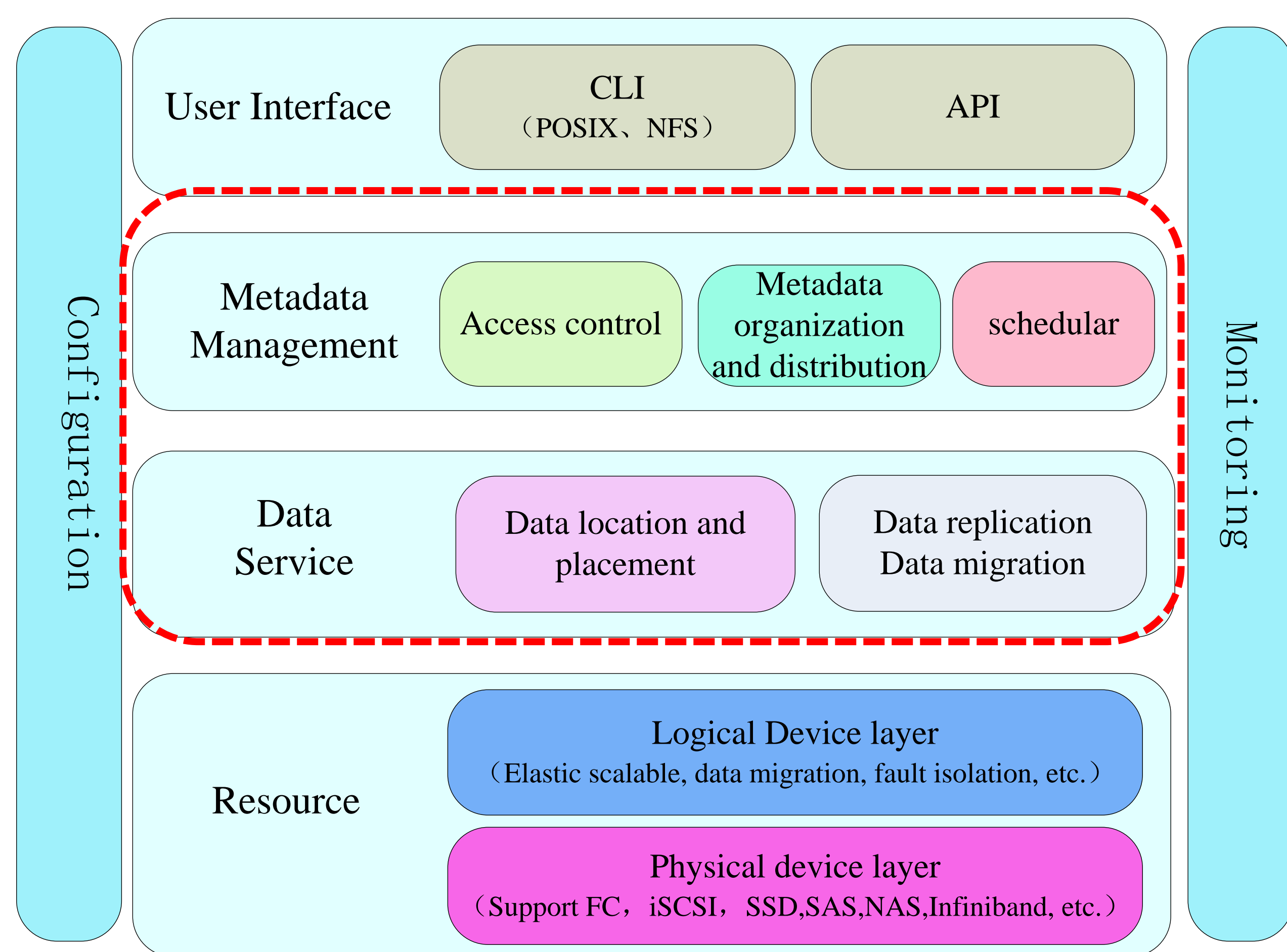


Figure 1: StarFS Architecture

How to distribute metadata/data?

Adaptive Directory Sub-tree Partition algorithm(ADSP)

ADSP divides the filesystem namespace into sub-trees with directory granularity. Sub-trees will be stored on storage devices in flat structure, whose locality information and file attributes are recorded as extended attributes. ADSP is an improved sub-tree partition algorithm with low computational complexity, also easy to be implemented.

Distributed Unified Layout(DULA)

DULA could provide uniform data distribution and efficient data positioning. DULA is an improved consistent hashing algorithm which is able to locate data in O(1) without the help of routing information.

3. Namespace

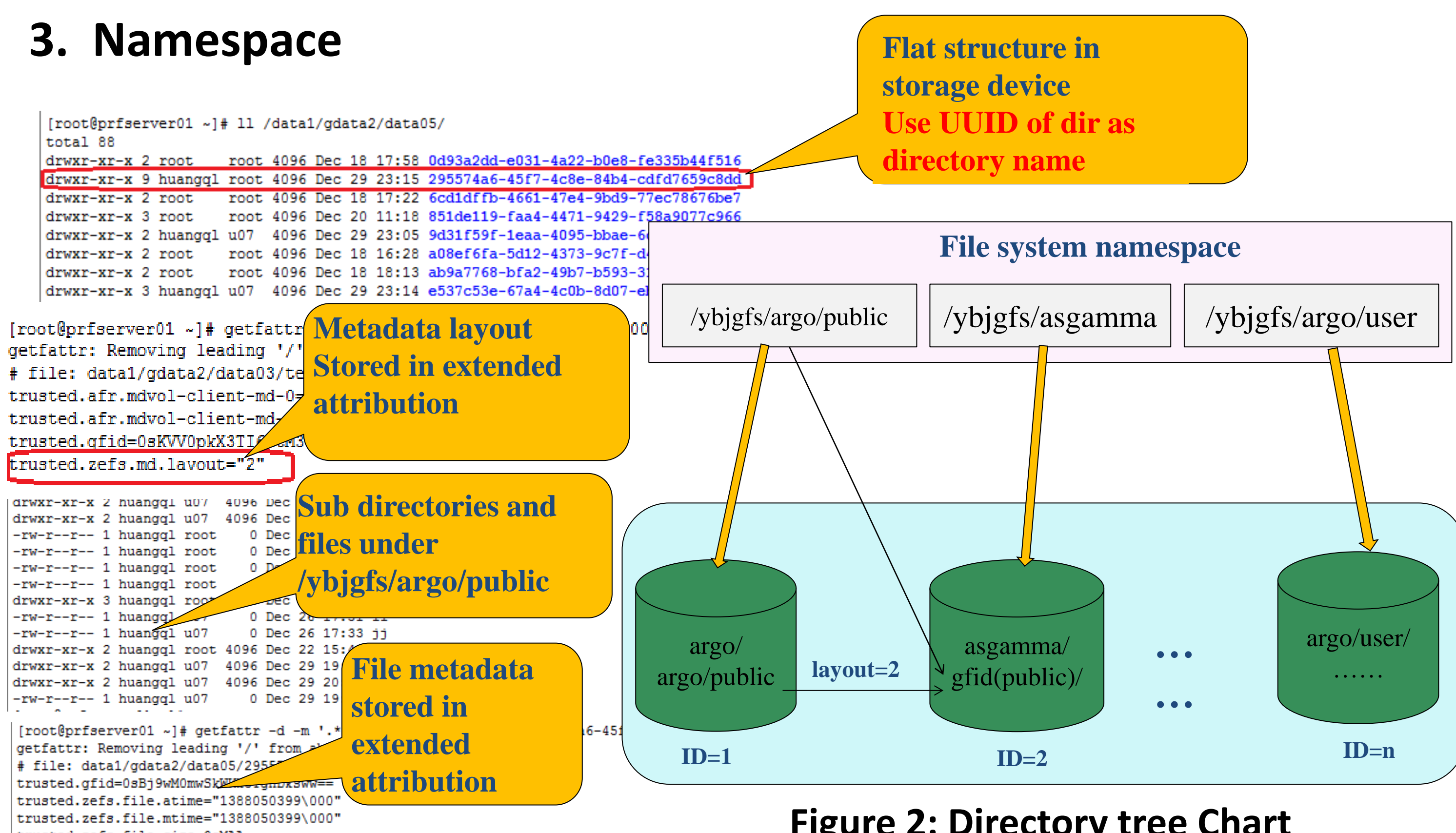


Figure 2: Directory tree Chart

4. Result 1: Metadata Performance

The result shows the Metadata performance is greatly improved by ADSP. From the Figure 3, we know performance in directory operation of ADSP is about 2~3 times than GLUSTER and the performance in file operation of ADSP is about 2 times than GLUSTER and ADSP is better than LUSTRE.

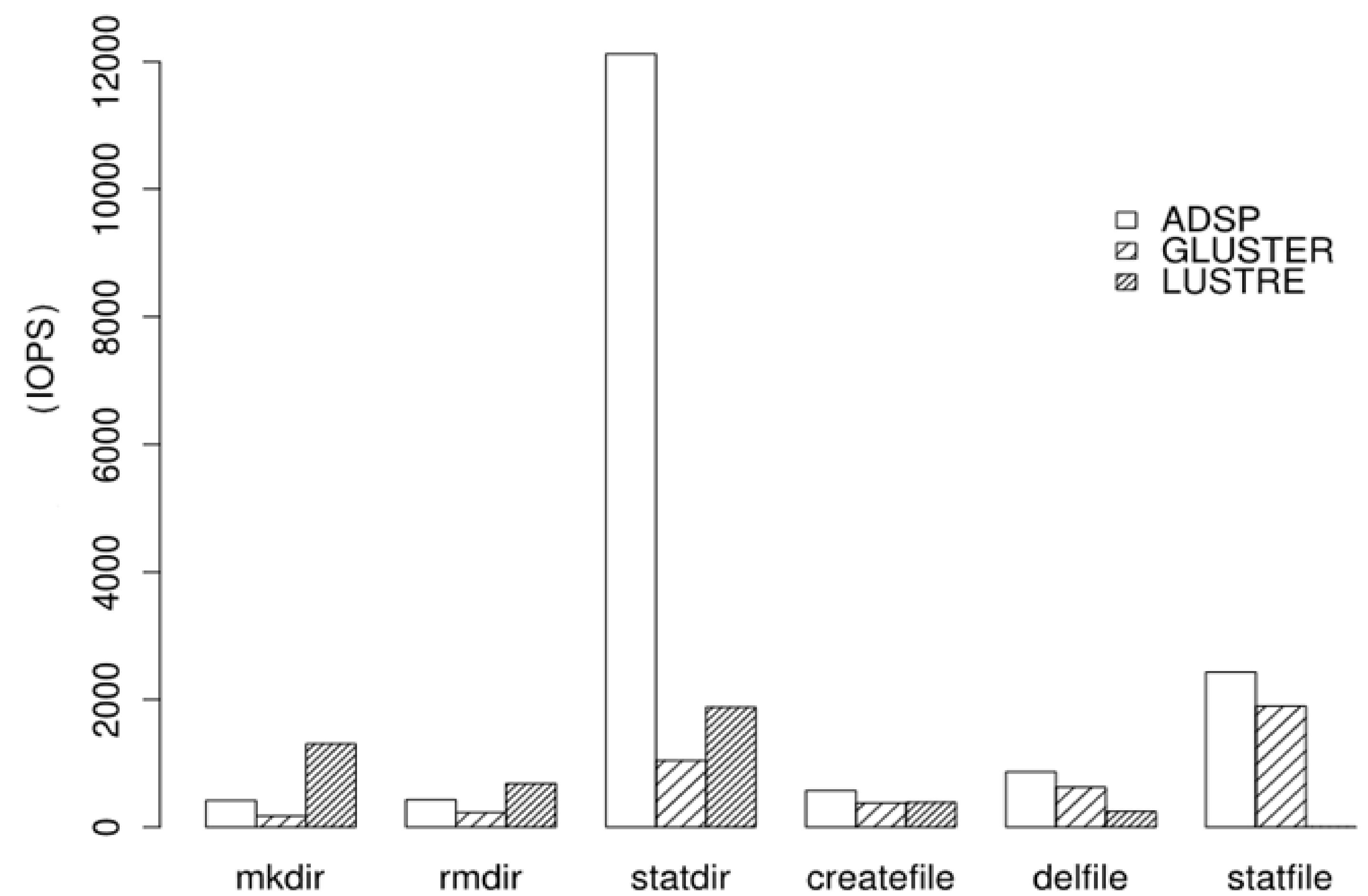


Figure 3: Metadata performance

5. Result 2: Scalability of StarFS

The result shows the scalability of StarFS. Figure 4 illustrate StarFS with ADSP algorithm has the better scalability than Gluster.

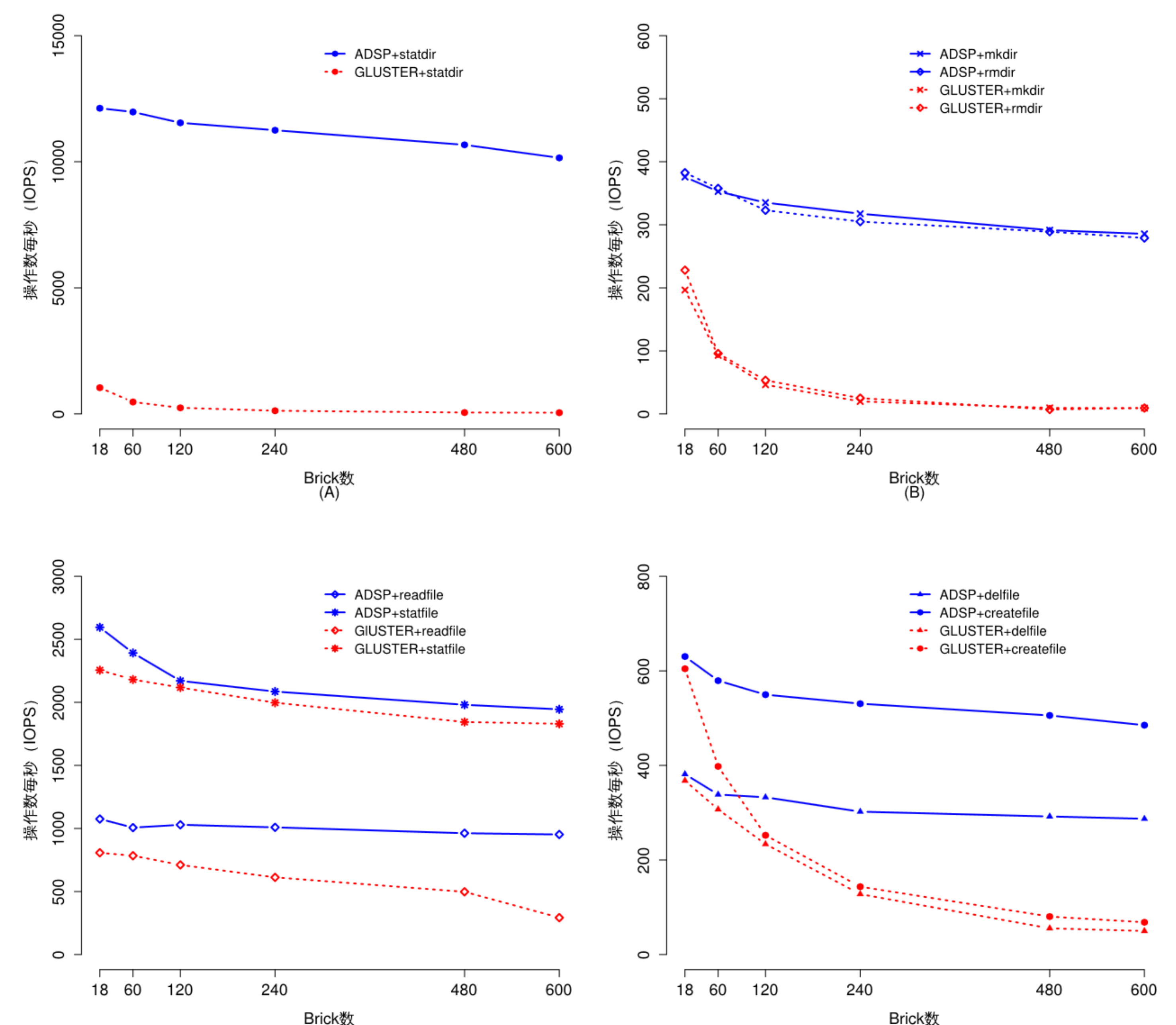


Figure 4: Scalability Evaluation between StarFS and Gluster

6. Conclusion

- Expansive Metadata module in Gluster framework
- Metadata performance greatly improved with ADSP. The performance in directory operation of the system is about 2~3 times than GLUSTER and about 2 times than GLUSTER in file operation.
- DULA algorithm achieve good data placement in cluster
- Better scalability than Gluster