

A New Data Access Mechanism for HDFS

Li Qiang^{1,2}, Sun Zhen-yu^{1,2}, Wei Zhan-chen^{1,2}, Sun Gong-xing¹

¹Institute of High Energy Physics, Beijing, China

²University of Chinese Academy of Sciences, Beijing, China

Email: liqiang88@ihep.ac.cn

Introduction

With the era of big data emerging, Hadoop has become de facto standard of big data processing. However, it is still difficult to get High Energy Physics (HEP) applications run efficiently on HDFS platform. There are two reasons to explain. Firstly, Random access to events data is not supported by HDFS platform. Secondly, it is difficult to make HEP applications adequate to Hadoop data processing mode. In order to address this problem, a new read and write mechanism of HDFS is proposed. With this mechanism, data access is done on local filesystem instead of through HDFS streaming interface. With this mechanism, data access is done on local filesystem instead of through HDFS streaming interface.

Methods

In High Energy Physics (HEP) requires to avoid chunking of binary data files. So we set the HDFS Blocks size equal or greater than the file size, a Block is a file.

We design HDFS Service module to communicate with the NameNode or DataNode. The data reading process is shown in Figure 1.

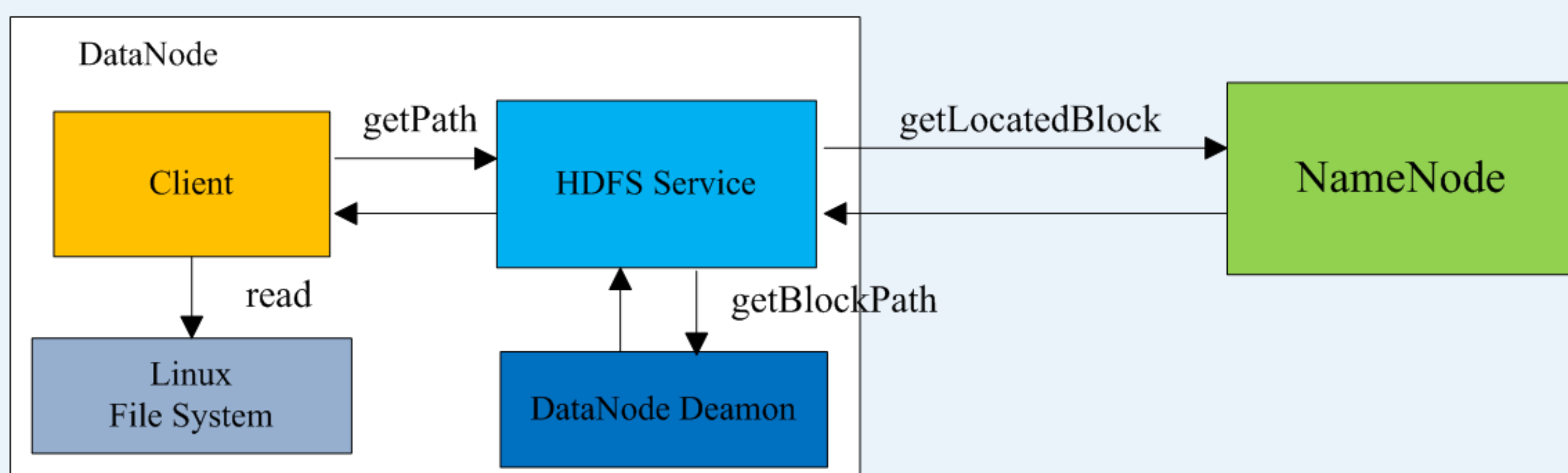


Figure 1. Data Reading Process

For data writing (Figure 2), the first file replica is written to the local DataNode, the rest replicas produced by copy of the first replica stored on other DataNodes. The first replica is written under the Blocks storage directory and calculates data checksum after write completion.

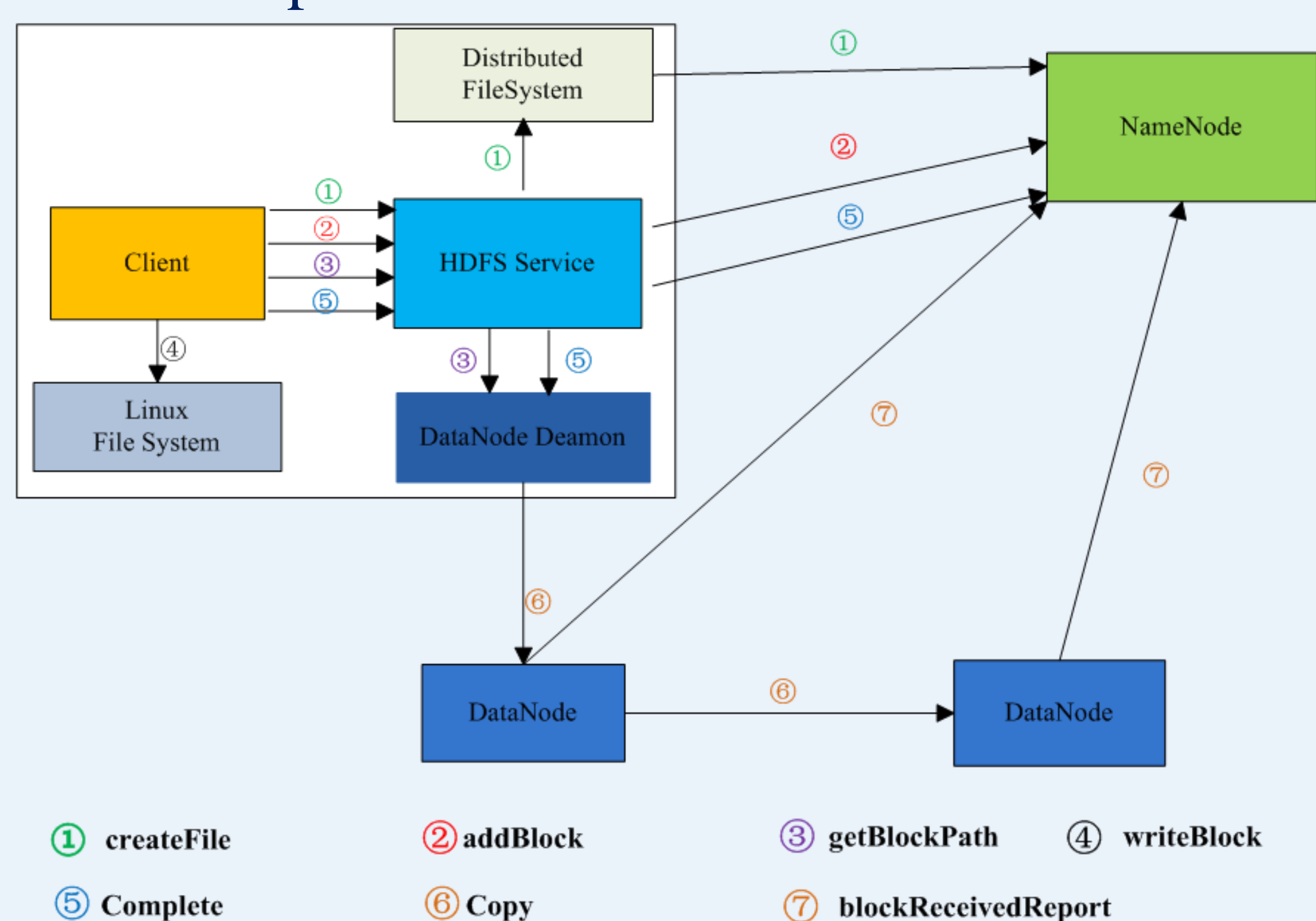


Figure 2. Data Writing Process

Performance

We use the High Energy Physics data process tool ROOT for testing.

In the premise of the data file is local, the test result (Figure 3) show that read performance through the HDFS Service is improved **about 10%** than FUSE.

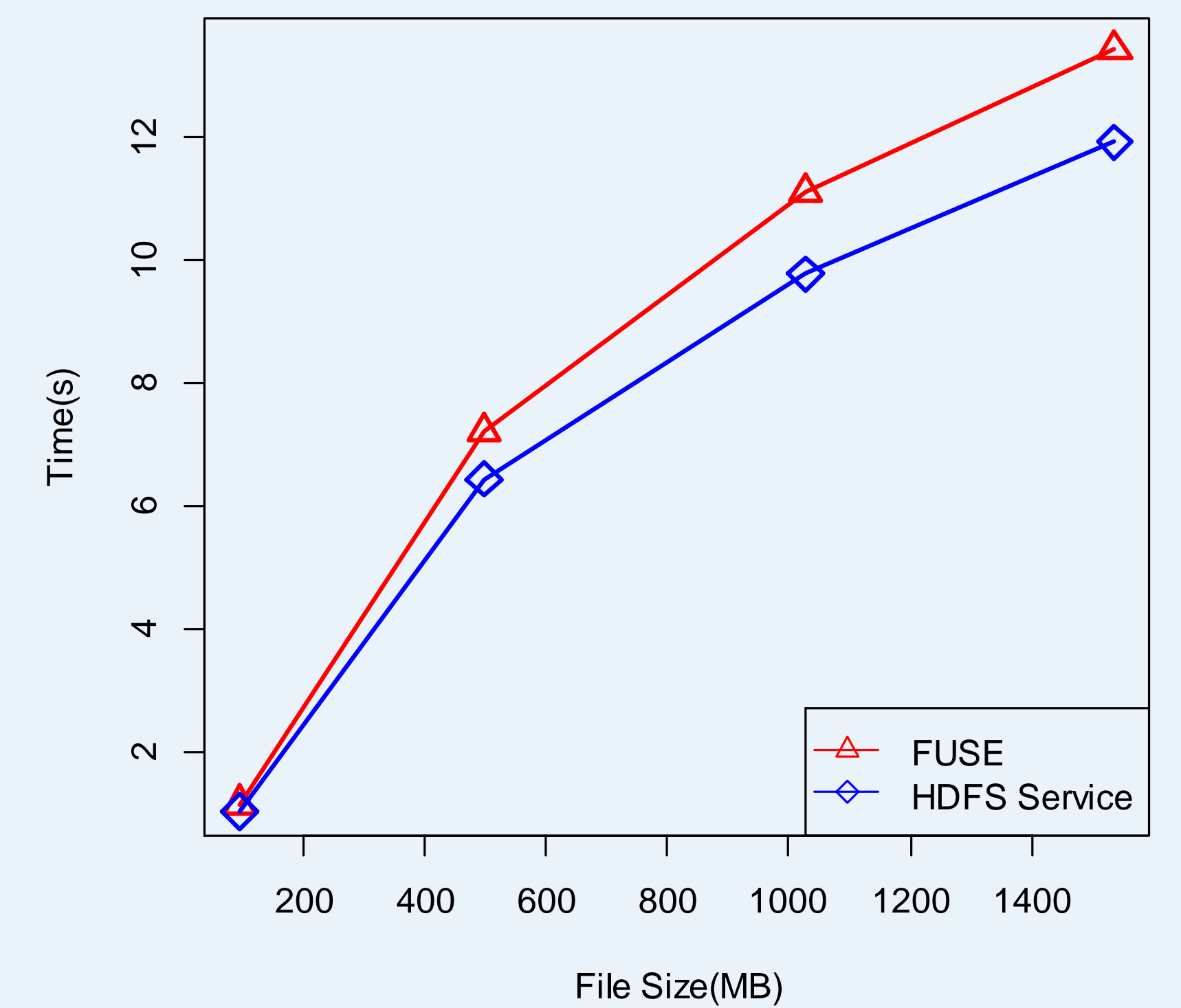


Figure 3. Read performance

In HEP, the traditional way to write the file to HDFS is via temporary file, first write the file to local File System, then copy the file to HDFS.

We use Temporary file method and HDFS Service method to write HEP event to HDFS separately to test write performance. The relationship between the number of events and the file size is show in Table 1.

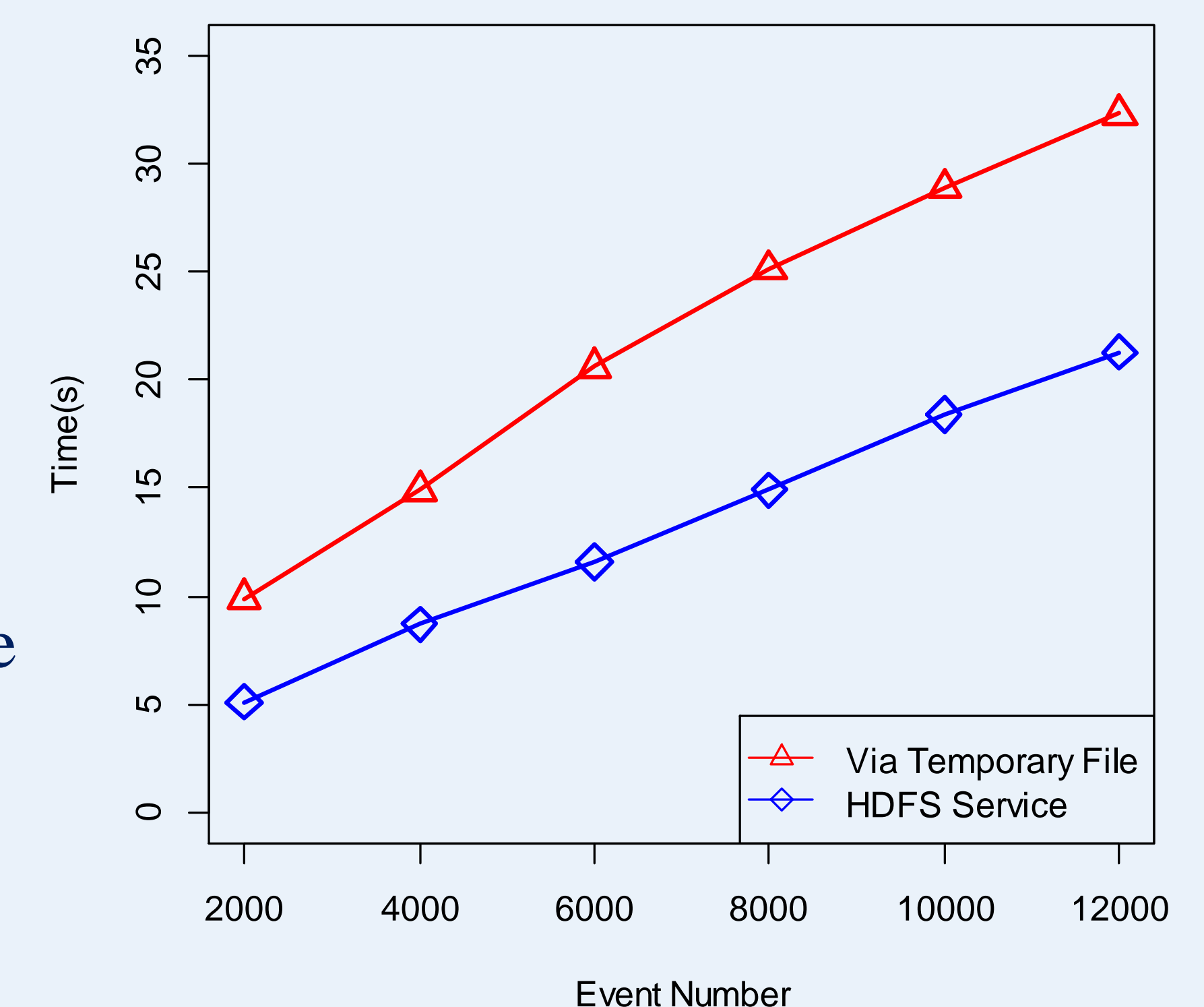


Figure 4. Write performance

Table 1. Event Number vs File Size

Event Number	Size (MB)
2000	147
4000	293
6000	440
8000	586
10000	733
12000	879

The test result (Figure 4) show that write performance is improved **more than 30%** when use HDFS Service.

Conclusions

The new data access mechanism for HDFS not only meet the needs of the high energy physics experiments random write data to HDFS, also improved the HDFS read and write performance. It have This have great significance for the application and extension of Hadoop in High Energy Physics.