Week 12

Paper Critique

Ceph: A Scalable, High-Performance Distributed File System (OSDI '06)

This paper presents the distributed file system called Ceph. It efficiently separates the data and metadata management by replacing allocation tables with a random data distribution function called CRUSH. Ceph utilizes a distributed metadata cluster architecture that improves the scalability of the access to the metadata, as a whole the scalability of entire system.

The advantage of suggested paper is that it widely used in a data service platform for the experiment. It considers QoS requirement, but it does not have features to improve the recovery which is essential to satisfy QoS.

Since the current trend in the computer system area, is to bypass the software stack that lies between the application and the storage device. Ceph is implemented in the user space so it can have advantage of bypassing the file system and access the raw block device directly.

Barrier-Enabled IO Stack for Flash Storage (FAST '18)

This paper suggests that the overhead lies in the I/O stack to guarantee the storage order especially in flash storage. This reordering occurs by the I/O scheduler, the command queue manager and the writeback cache manager in the I/O stack. It suggests novel filesystem to preserve order. It efficiently distinguish order-preserving write to orderless write. Therefore, while preserving order, the I/O scheduler issues I/O in the order inserted into the queue.

The contribution of this paper is that it suggests that order is not preserved in I/O stack and tried to keep the order of write considering the journaling and the caching effect. However, it has a limit of keeping order for only journaling based filesystem.

Lots of enterprises still use flash storage as their second storage devices. However, the development of NVMe to be commercialized is not a long day later story. Therefore, since it only considers the journaling based filesystem on the flash storage devices, it can consider how JFS works on NVMe, and if the order is still crashed.