

Reliable and High Performance Flash Storage Support for Big Data

Abstract:

Data storage and I/O system performance and reliability have been a critical concern for large scale data centers and big data applications. Compared to traditional hard drives, NAND flash-based solid state drives (SSDs) exhibit better performance, particularly on random workloads. Over the past decade, SSDs have been gaining popularity in massive storage. This trend is attributed to increasing storage density resulting from scaling manufacturing process technology, as well as the employment of multi-level cell (MLC and TLC) techniques. However, the deployment of SSDs in enterprise storage systems or data centers is still limited due to SSD endurance and reliability concerns. In this talk, I will discuss the challenges of big data applications from storage aspects and how SSD based flash storage can alleviate the challenges, and present several effective solutions for big data management to achieve high performance and reliability.