Lecture 13 Summary

This lecture talked about burst buffers in leadership-class storage systems. Burst buffers, which is a new storage tier in HPC systems, serve as an intermediate, high-bandwidth store between computational resources and the primary parallel file system and are aimed to absorb periodic high-intensity I/O phases form application, which alleviate the need of parallel file systems to scale to short-lived peak I/O rates. In Burst Buffer I/O paper, a model was developed by using CODES interconnect and storage simulation suite, which provides the ability to efficiently execute detailed and large-scale discrete-event simulations. To study the I/O and network interference, they dragonfly network model and the burst buffer storage model are used to simulate the interference analysis. In the experiments, authors execute jobs that generate I/O traffic in the form of checkpoint writes to the burst buffer nodes as well as jobs that generate uniform random network traffic. Each compute node entity generates either checkpointing traffic or communication traffic. To warm up the simulation, background traffic starts a few simulated seconds before the checkpoint workload and continues until that the checkpoint workload completed. In the performance analysis of communication traffic, there are three results: 1. The combination of nearest burst buffer allocation and minimal routing causes the least perturbation to network communication, regardless of job allocation policy. 2. Adaptive routing produces longer worst-case latency than minimal routing in all cases. 3. Contiguous job placement brings the least perturbation to background traffic in all cases, while random node job placement brings maximum jitter. In the performance analysis of I/O traffic, there are 3 observations: 1. Smaller jobs checkpoint faster with a random burst buffer allocation policy, while larger jobs checkpoint faster with a nearest burst buffer allocation policy. 2. Adaptive routing offers a slight improvement to checkpoint traffic time for randomly allocated burst buffers; for nearest burst buffer allocation policy, adaptive routing offers improvement with random node and random router job placements but brings a slow down with contiguous job placement. 3. Contiguous job placement over-utilizes selected burst buffer nodes while the rest stay un-utilized. Therefore, there is a tradeoff between minimum perturbation to communication traffic and efficiently utilizes the burst buffer nodes.