

국외 출장 보고서

A Survey on GPU-Accelerated Data Management Technologies

Sung-Soo Kim

sungsoo@etri.re.kr

요약

GPU 상의 범용계산 접근법 (GPGPU)은 일반적으로 컴퓨터 그래픽스를 위한 계산만 맡았던 그래픽 처리 장치(GPU)를, 전통적으로 CPU가 맡았던 응용 프로그램들의 계산에 사용하는 기술이다. 기존 CPU 클러스터 기반의 하둡을 GPGPU를 적용하여 더 빠른 처리능을 얻을 수 있을까? 이론적으로 처리해야 할 프로세스가 병렬 컴퓨팅에 최적화되어 있는 경우 GPU는 CPU보다 50배에서 100배까지 빠르게 연산을 수행한다. 하지만, GPGPU에서 주요 병목현상은 GPU 메모리와 CPU 메모리 사이의 PCIe 버스를 통한 느린 데이터 전송에서 발생한다. 그래서 하둡과 같은 대규모 데이터 처리에 단순히 GPGPU의 처리방식만 적용하면 성능개선을 기대하기 어렵다. 현재까지 하둡과는 별개로 GPGPU기반 데이터관리 사례로는 하버드와 MIT에서 개발한 대규모 스트림데이터 처리/분석/가시화를 수행할 수 있는 시스템인 MapD [1]가 있지만, 대부분 GPGPU 기반 하둡 연구들은 실험결과 제시수준이다. GPGPU를 통해 하둡기반 빅 데이터 시스템의 성능을 높이기 위해서는 GPU상에서 다루어야 할 데이터 저장/압축모델, 처리 방식 뿐만 아니라, 질의 처리 방식도 GPU에 최적화되어야 한다.

1 본론

참고 문헌

- [1] T. Mostak. Introduction to MapD. *MapD White Papers*, 2015.