

사이언스 클라우드 컴퓨팅 환경에서 대규모 데이터 관리를 위한 스토리지 운영 기술 연구

A study on the storage operation technique managing large-scale data in the Science Cloud Computing environment

정용환, 윤준원, 곽재혁, 함재균*
한국과학기술정보연구원*

Jung yong-hwan, Yoon jun-weon, Kwak jae-hyuck,
Hahm jae-gyoon*
Korea Institute of Science and Technology
Information*

요약

최근 국내외에서 클라우드 컴퓨팅 기술을 활용한 과학기술 응용 연구가 점차 이루어지고 있다. 본 연구에서는 대규모 데이터 처리를 필요로 하는 과학기술 응용을 지원하기 위한 스토리지 구축 및 운영 기술을 다루고자 한다. 본 연구를 위하여 많은 과학기술 분야 중 대규모 데이터의 처리를 필요로 하는 천문우주 분야 응용을 대상으로 선정하였으며, 물리적인 컴퓨팅 자원의 스토리지 구축 및 운영 방안뿐만 아니라, 응용 연구자가 직접 접근 및 사용하는 가상 클러스터 환경의 스토리지를 대규모 데이터 처리에 최적화시킬 수 있는 방안을 제시하고자 한다.

I. 서론

클라우드 컴퓨팅 기술은 다수의 상이한 컴퓨팅 자원들을 가상화함으로써 사용자에게 온-디맨드 방식으로 서비스를 제공하는 기술로 자원 효율성의 극대화 및 관리비용 최소화라는 장점으로 최신 IT 기술 트렌드로 부상하고 있다. 이미 비즈니스 영역에서는 많은 회사들(Amazon, MS, Google 등)이 인프라(IaaS), 플랫폼(PaaS), 서비스(SaaS) 형태로 클라우드 컴퓨팅 서비스 중에 있다.

High Performance Computing(HPC) 및 High Throughput Computing(HTC) 위주의 과학기술 응용 분야에서도 처리하고자 하는 데이터의 크기 및 규모가 커짐에 따라 데이터 분석을 위하여 스토리지와 대규모 컴퓨팅 자원이 결합된 형태의 컴퓨팅 자원 수요가 요구되고 있다. 이러한 수요를 충족시키기 위하여 클라우드 컴퓨팅과 같이 확장성 있고 유연한 컴퓨팅 기술을 적용하고자 하는 연구가 점차 진행되고 있다.

본 연구에서는 대규모 데이터의 처리를 필요로 하는 많은 과학기술 응용 연구 분야 중에서 천문우주 분야를 대상으로 선정하였으며 해당 분야 연구자에게 클라우드 컴퓨팅 연구 환경을 제공함에 있어 최적화된 스토리지 구축 및 운영 방안을 제시하고자 한다. 물리적인 자원뿐만 아니라 천문응용 연구자가 직접 접근하여 사용하는 가상 클러스터 환경의 스토리지를 대규모 데이터 처리에 최적화시킬 수 있는 방안을 제시하고자 한다.

II. 본론

1. 우주천문분야 응용 특성 분석

천체의 모든 별들은 기본적으로 시간에 따라 그 밝기가 변화하는데 이것을 변광 현상이라 하며, 이 변광 현상은 별들의 물리적인 특성을 이해하는 중요한 자료로 활용된다. 변광 특성 분석 응용은 밝은 별 약 1,600만 개를 대상으로 하며, 각 별에 대한 변광 현상 데이터의 크기는 1MB 미만이며, 전체 데이터의 크기는 대략 12TB의 규모이다.

해당 응용을 수행하기 위해서는 12TB 규모의 입력 데이터 스토리지 뿐만 아니라 비슷한 규모로 생성되는 결과 데이터를 저장하기 위한 스토리지가 필요하다.



▶▶ 그림 1. 천문우주 응용수행 절차

변광 특성 분석 응용의 단일 작업 프로세스는 그림 1과 같이 단순한 형태를 취하지만, 전체 작업 횟수는 입력 데이터의 수와 동일하게 1,600만 번 수행되기 때문에 전체 작업 수행 중 소요되는 Disk I/O 시간은 전체 매우

