

국내 슈퍼컴퓨팅센터 소개

# 선진 일기예보의 초석, 기상청 슈퍼컴퓨터 3호기

오하영 (기상청 슈퍼컴퓨터운영과)

## 1. 기상청 슈퍼컴퓨터의 역사

### 기상청 슈퍼컴퓨터와 일기예보의 만남

역사의 비극, 3년간의 동족상잔 6.25전쟁이 막 끝나고 온통 잿더미로 변한 나라에서 하루하루를 어렵게 살아야만 했던 1954년 대한민국, 바로 그때 지구 반대편에서는 당시 대한민국 국민들이 상상도 할 수 없는 놀라운 일, 지구를 감싸고 있는 대기(공기)의 흐름을 방정식화하여 컴퓨터로 계산해 내는 일이 현실로 나타나게

되었다. 일기예보(수치예보)를 생산하기 위하여 슈퍼컴퓨터를 세계 최초로 이용한 나라는 놀랍게도 미국도, 영국도 아닌, 북유럽의 작은 국가인 스웨덴이었다. 뒤를 이어 1957년 미국에서도 슈퍼컴퓨터를 이용한 일기예보(수치예보) 실시간 현업 운영이 시작되었다. 그리고 약 30년이 더 지난 1988년, 당시 올림픽이라는 국제적인 행사도 성공적으로 유치했던 대한민국이었지만 기상분야에서는 컴퓨터를 이용한 일기예보(수치예보)라는 매우 생소한 분야에 첫 발을 내딛게 되는 시기가기도 했다. 이 때 기상청이 사용한 최초의 슈퍼컴퓨터는 한국과학기술정보연구원(KIST)의 슈퍼컴퓨터 1호기(Cray 2S)였다. 1994년 처음으로 기상청이 자체적으로 도입한 기상 분석용 컴퓨터(Fujitsu, VPX220/10, 1.25Gflops)를 활용하여 전지구 규모의 수치모델 시험운영을 시작하게 되었다. 당시 사용된 기상 분석용 컴퓨

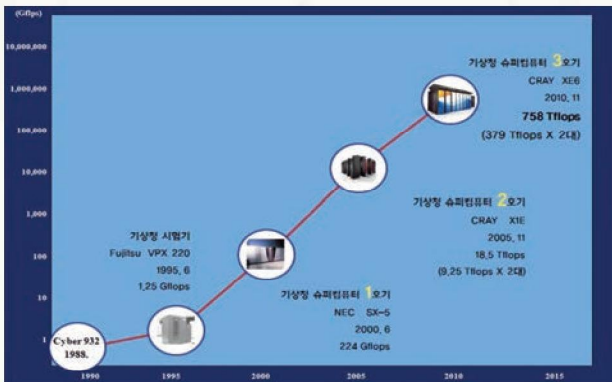


그림 1. 기상청 슈퍼컴퓨터 시스템 도입 역사



그림 2. 기상청 슈퍼컴퓨터 1호기(상) / 기상청 슈퍼컴퓨터 2호기(하)

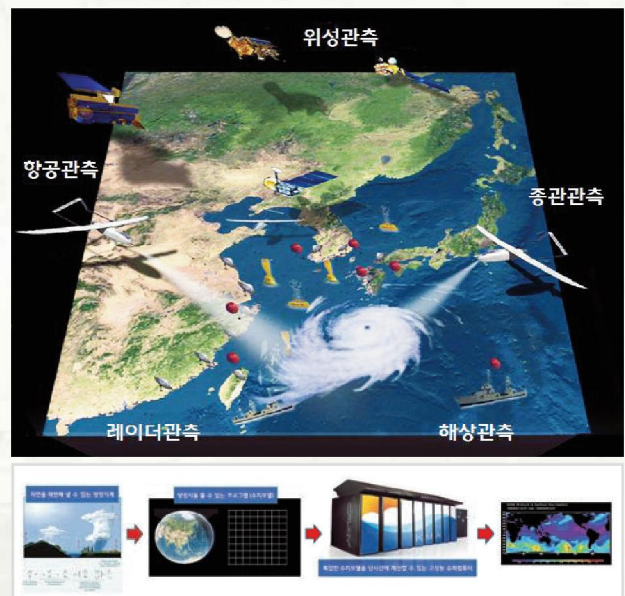


그림 3. 일기예보를 위한 다양한 관측 장비와 일기예보(수치예보) 생산과정

터의 성능 1.25Gflops는 요즘 일반 국민들이 일상생활에 흔히 사용하는 노트북 1대 보다도 못한 성능이라고 생각하면 IT기술이 정말 빠르게 세상을 변화시키고 있다는 것에 다시 한번 놀라움을 감출 수 없다.

### 기상청 슈퍼컴퓨터 도입 10년, 10년이면 강산도 변한다.

1999년 드디어 기상청도 기상청 슈퍼컴퓨터 1호기(NEC, SX-5/28)의 도입을 시작으로 본격적인 수치예보 실시간 현업운동을 시작하게 되었다. 우리나라는 주요 기상 선진국에 비해 약 40여년 늦게 수치예보 현업 운영이 시작되었다고 할 수 있다. 2004년~2005년 기상청 슈퍼컴퓨터 2호기(CRAY, X1E, 18.5TF)를 도입하여 이전보다 정교해진 전지구 예보모델 뿐만 아니라 앙상블 예보모델도 실시간으로 현업 운영을 하게 되었다. 하지만 슈퍼컴퓨터 2호기 도입 당시만 해도 슈퍼컴퓨터를 설치할 자체 전산실 공간이 부족하여 서초동에 위치한 KDC라는 민간 전산실을 임대해서 운영하는 수준이었다.



그림 4. 국가기상슈퍼컴퓨터센터 진경

### 기상청 슈퍼컴퓨터 3호기와 국가기상슈퍼컴퓨터센터

기상청은 2010년 슈퍼컴퓨터 3호기 설치 시기에 맞춰서 2008년부터 2009년까지 충청북도 청원군 오창과 학산업단지에 국가기상슈퍼컴퓨터센터라는 슈퍼컴 전용 건물을 건축했다. 국가기상슈퍼컴퓨터센터는 기상

용 슈퍼컴퓨터 운영에 최적화된 환경 구현을 위한 대규모 항온항습시설, 이중화된 전력공급장치 뿐만 아니라 비상시 외부전력이 완전히 차단된 상태에서도 48시간동안 슈퍼컴퓨터를 안정적으로 운영할 수 있는 비상용 발전기 등 안정적인 시스템 운영에 필수적인 대규모 기반설비를 완벽하게 갖춘 명실상부한 국가기상슈퍼컴퓨터센터의 모습을 갖추고 2010년 3월 센터 준공식을 개최했다.

## 2. 기상청 슈퍼컴퓨터 3호기 (Cray XE6, 758TF) 도입

2010년 12월 하반기에 최종적으로 도입 완료된 기상청 슈퍼컴퓨터 3호기는 미국 CRAY사의 XE6이다. 계산능 758TF, 공유자료 저장장치 약 2.5PB, 백업 저장장치 4.5PB로 구성된 2012년 현재 국내 최대 규모이며, 2012년 6월 기준 전 세계 슈퍼컴퓨터 성능 순위 55위, 56위에 랭크되어있다(참고 <http://www.top500.org>).



그림 5. 슈퍼컴퓨터 3호기 CRAY XE6 “해운”



그림 6. 슈퍼컴퓨터 3호기 CRAY XE6 “해담”



### 기상청 슈퍼컴퓨터 3호기 제원

슈퍼컴퓨터 3호기 최종본	
시스템명	Cray XE6
설치시기	2010년 9월
프로세서타입	AMD 2.1GHz (12 core)
캐비닛 수	40 (20 + 20)
Core 수	90,240
이론성능	758TF
메모리용량	119.8TB
공유저장장치	2.5PB (Lustre file system)
백업용량	4.5PB

그림 7. 슈퍼컴퓨터 3호기 시스템 제원

기상청 슈퍼컴퓨터 3호기는 AMD 사의 최신 Multi-core (12-core) 프로세서와 Gemini 인터커넥트 기술로 구성된 슈퍼컴퓨터 시스템이다. 슈퍼컴퓨터에서 계산을 담당하는 컴퓨팅 블레이드(그림 9(상))에는 4개의 컴퓨팅 노드가 탑재되어 있으며, 각각의 노드에는 AMD 사의 2.1GHz 12-core 소켓 2개와 DDR3 32GB의 메모리가 장착되어 있다. 부트 영역, I/O 영역, 로그인 영역 등 그림 Cray XE6 Gemini 구조 다양한 서비스를 제공하는 서비스 블레이드(그림 9(하))에는 4개의 각각 서비스 노드가 장착되어 있으며, 각 노드에는 AMD사의 2.2GHz

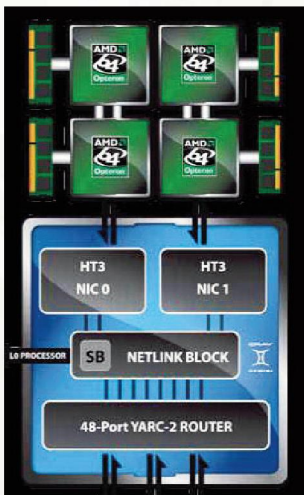


그림 8. Cray XE6 Gemini 구조

6-core 소켓 1개와 DDR2 16GB의 메모리가 장착되어 있다. 하나의 서비스 블레이드는 컴퓨팅 블레이드와 공유 저장장치(Lustre file system)간의 I/O 대역은 DDR 4X(16Gbps) 기반 Infiniband 스위치로 연결되어 있으며, 이론적으로 최대 32GB/s까지 제

공할 수 있다.

모든 컴퓨팅 노드를 연결하는 인터커넥트는 AMD 사의 Hyper Transport 3.0 기반의 Gemini(그림 8)로 구성되어 있으며, 각각의 컴퓨팅 노드마다 최대 20GB/s의 대역폭을 제공한다. 인터커넥트의 네트워크 구조는 일반적인 Fat-tree 방식과는 매우 다른 크레이사 고유의 3-D torus 방식이 적용되어 있으며, 인터커넥트를 통해 연결된 특정 노드에 장애 발생시 자동적으로 라우팅 경로를 갱신하여 재설정하는 동적 라우팅 기능도 포함되어 있다. 참고로 기상청 전체 계산시스템과 공유저장장치간의 I/O 실제 성능 벤치마크 실험은 HPC 벤치마크 툴인 IOZONE을 활용하였으며, 실험 결과는 단일 스트림 실험 시 약 1GB/s 이상, 멀티 스트림 실험 시 25GB/s 이상의 실제 대역폭을 제공하는 것으로 측정된 바 있다. 기상청 슈퍼컴퓨터 3호기 전체 컴퓨팅 블레이드의 core 개수는 90,240개, 전체 이론성능은 758TF이며, 전체 메모리는 119.8TB로 구성되어 있다.

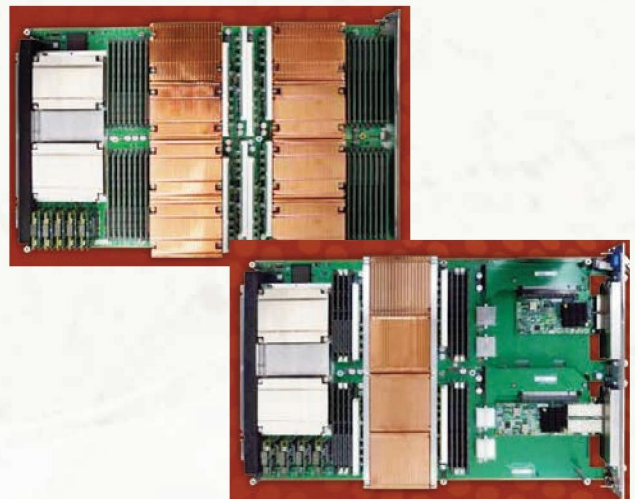


그림 9. Cray XE6 컴퓨팅 블레이드(상) / 서비스 블레이드(하)

기상청 슈퍼컴퓨터 3호기의 공유저장 장치는 리눅스를 기반으로 하는 슈퍼컴퓨터 시스템의 고성능 I/O에 적합한 러스터 파일시스템(Lustre file system)으로 구성되어 있다. 러스터 파일 시스템은 크게 3가지 구성요소인 클라이언트 서버, MDS(Meta Data Server), OSS(Object

ject Storage Server)로 구분할 수 있다. 기상청의 경우 슈퍼컴퓨터, 전-후처리시스템 및 로그인 시스템이 러스터 파일시스템의 클라이언트 서버에 해당되며, 메타 데이터 정보를 관리하는 MDS와 실제 슈퍼컴퓨터 사용자의 자료가 저장되는 스토리지 관리서버인 OSS는 리눅스 기반의 Dell사의 블레이드 서버로 구성되어 있다. 공유 저장 장치의 디스크 RAID는 RAID 6(8+2) 형태로 구성하여 저장 장치의 성능과 안정성을 극대화 시켰다. 저장 장치의 하드디스크는 400GB SAS(15,000rpm) 디스크와 1TB 용량의 SATA(7,200rpm)으로 구성되어 있다.

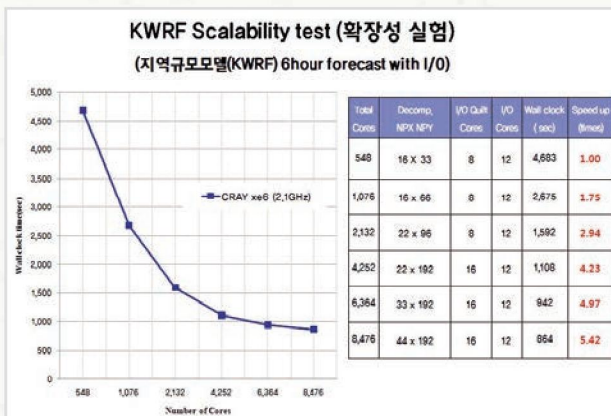


그림 10. 기상청 수치모델 중 지역규모예측용 KWRF 확장성 실험

기상청에서 운영하는 중요 모델의 실제 계산 성능을 검증하기 위하여 슈퍼컴퓨터 3호기의 컴퓨팅 블레이드 core 갯수를 늘려가면서 시스템과 사용자 어플리케이션의 확장성을 테스트한 벤치마크 실험(그림 10)을 실시했고, MPI 병렬 프로그램을 이용한 경우 단일 기상청 사용자 어플리케이션에 대하여 약 10,000개까지의 core를 동시에 안정적으로 수행할 수 있는 것을 실제로 확인하였다.