분산 컴퓨팅 기술을 이용한 고해상도 강수량 예측

윤준원*, 송의성**

요약

분산컴퓨팅은 네트워크로 연결된 여러 컴퓨터들의 연산 능력을 이용하여 거대 계산 문제를 해결하려는 분산처리 모델이다. 인터넷에 연결된 수많은 컴퓨팅 자원들의 참여를 통해 대규모의 계산이 필요한 기상, 바이오, 천문학, 암호학 등과 같은 다양한 분야의 어플리케이션 들을 병렬로 처리할 수 있다.

본 논문에서는 기상 분야의 고해상도 강수량 예측 어플리케이션 수행을 위해 인터넷 분산컴퓨팅 모델을 구성하여 성능을 분석하였다. 한반도의 강수량 예측을 위해서 중규모 예측 모형인 QPM(Quantitative Precipitation Model)을 적용하였으며 이 모형은 한반도의 지형을 격자 간격 27km로 나누고 각 격자의 특성인 고도, 기온, 강수, 강도 습도 등을 기반으로 강수량을 예측하게 된다. 그러나 QPM의 격자 간격에 따른 분석은 모형 구축에 많은 시간이 소요 되고 한번에 수행되어야 할 계산량이 많아 효율성이 저하된다. 따라서 이 모형을 기반으로 하여 3km 간격의 상세 지형을 반영하는 모형으로 소규모 지형 효과를 표현함으로써 상세 지역에서의 강수량 산출과 지형에 따른 강수량의 분포 파악이 용이해지며 계산 효율성을 개선시킬 수 있다. 이렇게 상세지역으로 세분화 된 모형은 병렬처리가 필요하며 계산노드의 수가 많아질수록 그 효율은 선형적으로 증가된다. 이 모형은 20X20의 아격자 도메인의 분산된 단위작업들로 나뉘어 분산되고 네트워크로 연결된 컴퓨팅 자원에서 수행되게 된다.

키워드 : 인터넷 분산컴퓨팅, 병렬처리, 분산 어플리케이션, 기후예측, 계산과학

High Resolution Rainfall Prediction Using Distributed Computing Technology

JunWeon Yoon*, Ui-Sung Song**

Abstract

Distributed Computing attempts to harness a massive computing power using a great numbers of idle PCs resource distributed linked to the internet and processes a variety of applications parallel way such as bio, climate, cryptology, and astronomy.

In this paper, we develop internet-distributed computing environment, so that we can analyze High Resolution Rainfall Prediction application in meteorological field. For analyze the rainfall forecast in Korea peninsula, we used QPM(Quantitative Precipitation Model) that is a mesoscale forecasting model. It needs to a lot of time to construct model which consisted of 27KM grid spacing, also the efficiency is degraded. On the other hand, based on this model it is easy to understand the distribution of rainfall calculated in accordance with the detailed topography of the area represented by a small terrain model reflecting the effects 3km radius of detail and terrain can improve the computational efficiency. The model is broken down into detailed area greater the required parallelism and increases the number of compute nodes that efficiency is increased linearly.. This model is distributed divided in two sub-grid distributed units of work to be done in the domain of 20X20 is networked computing resources.

Keywords : Internet-Distributed computing, Parallel Processing, Distributed Application, Climate Prediction, Computational Science

※ Corresponding Author: Ui-Sung Song
Received : February 04, 2015

Revised : February 25, 2016
Accepted : February 29, 2016
* Dept. of Supercomputing Center, KISTI

1. 서론

분산컴퓨팅은 네트워크상에 연결된 수많은 계산 자원을 활용하여 계산을 수행함으로써 빠른 시간 안에 문제를 해결할 수 있을 뿐만 아니라 문제의 규모를 확장하여 처리할 수 있다[1]. 분산컴퓨팅은 여러 형태의 모델로 발전하고 있으며 대표적인 기술로는 그리드 컴퓨팅, 클라우드 컴퓨팅, 슈퍼컴퓨팅 등이 있다. 그리드 컴퓨팅은 특정 조직내 대규모의 분산된 자원을 가상 조직(Virtual Organization)으로 구성함으로써 고성능의 컴퓨팅 파워를 제공하는 컴퓨팅 패러다임이다[2]. 특히, 인터넷 분산컴퓨팅은 인터넷에 연결된 일반 사용자의 컴퓨팅 자원을 이용하여 고성능의 컴퓨팅 자원을 창출하여 고처리의 분산 컴퓨팅을 수행하는 컴퓨팅 패러다임이다. 이런 인터넷 분산컴퓨팅은 자발적 컴퓨팅(Volunteer Computing)으로도 불리며, 대표적인 사례로는 외계의 신호를 찾는 SETI@home[3]이나 암호 메시지 해독을 위한 distributed.net[4]과 같은 프로젝트들이 성공적 수행됨으로써 많은 관심과 연구가 이루어지고 있다.

2. 관련연구

인터넷 분산컴퓨팅은 1996년 2월 인터넷을 통해 무료로 다운로드 할 수 있는 Prime95와 MPrime과 같은 소프트웨어를 사용해 메르센 소수를 찾는 GIMPS(Great Internet Mersenne Prime Search) 공동 프로젝트를 진행하면서 시작되었다[5]. 이어 1997년 암호화 키를 찾는 distributed.net 프로젝트가 진행되어 1998년 1월 RSA DES Challenge III를 22 시간 15분 만에 해독하면서 유사한 프로젝트들이 개발되기 시작하였다. 특히, SETI@home 프로젝트가 성공적으로 수행되면서 인터넷에 물려있는 수많은 일반 PC 사용자들이 공공의 문제들을 해결하기 위해 자발적으로 참여하기 시작하였다.

SETI@home은 우주로부터의 복사 전파에 숨어있는 이성적인 무선전파를 분석하여 외계 지

성체의 존재를 찾아내려는 프로젝트로 수백만의 자원자들이 SETI@home의 화면보호기를 각자의 컴퓨터에 내려받고 서버로부터 일을 받아 처리한 후 그 결과를 버클리 대학교에 있는 데이터 서버로 보내는 방식으로 전체 연산을 분산화하여 수행하는 프로젝트이다[3].

Folding@home은 단백질의 접힘 기능을 규명하여 질병의 원인을 찾아내는 프로젝트로 PC의 CPU자원 뿐만 아니라 그래픽카드의 GPU 자원 또한 이용하여 연산을 수행하게 된다[6].

3. 작업 수행

(그림 1)과 같이 인터넷 분산컴퓨팅 컴퓨팅 구성은 크게 어플리케이션 수행자, 자원제공자, 중앙관리서버로 구성된다. 어플리케이션 수행자는 자원제공자에게 작업을 요청하는 주체이며, 자원제공자는 자신의 유휴 컴퓨팅 자원을 제공해주는 자발적 자원제공자이다. 일반적으로 자원제공자는 컴퓨팅 자원 제공과 연산을 수행을 위해 경량의 에이전트 프로그램을 설치하게 된다. 중앙관리서버는 인터넷에 연결된 자원제공자들의 작업을 통제하는 관리자이다. 일반적으로 어플리케이션 수행자는 서버에게 병렬처리가 필요한 작업을 위탁한다. 중앙관리서버는 어플리케이션 작업을 각각의 세부작업으로 구성한 뒤 스케줄링 알고리즘에 따라 자원제공자들에게 분배한다. 자원제공자는 할당받은 작업에 대해 연산을 수행하고 그 결과를 중앙관리서버에게 반환한다. 마지막으로 중앙관리서버는 자원 제공자로부터 받은 결과들을 정리하여 해당 어플리케이션 수행자에게 최종 작업 결과를 반환한다. 본 논문에서는 고해상도 강수량 예측을 위해 인터넷 분산컴퓨팅 환경을 구축하였다.

Tel: +82-42-869-0581, Fax: +82-42-869-0569
email: jwyoona@kisti.re.kr

** Dept. of Computer Education, Busan National University of Education / ussong@bnu.ac.kr

(그림 1) 인터넷 분산컴퓨팅 개요도

(Figure 1) Internet-Distributed Computing Overview

4. 고해상도 강수량 예측

4.1 연구개발의 배경 및 목적

우리나라에서 해마다 발생하고 있는 풍수해에는 연평균 강우량의 2/3가 하절기에 집중된다는 기상학적인 원인과 전국토의 70% 이상이 산지로 구성되어 있다는 지형학적인 원인이 크게 작용한다. 또한 우리나라와 같이 국토는 좁은 반면, 지형지세가 복잡한 구조로 되어 있는 지역에서의 정확한 기후예측 자료를 생산하기 위해서는 지형의 특성을 반영한 충분한 해상도의 기상자료가 필요하다. 특히 강수량의 경우는 언제, 어디서 발생할지를 결정하는 많은 물리적 과정들이 일반적인 일기예측 모형에서 표현하는 것보다 훨씬 작은 공간적 규모의 대기 현상들에 의해 결정되고 공간적 편차가 매우 크기 때문에 다른 기상요소의 예측에 비해 매우 까다롭다. 최근 자연 재해 발생 및 피해는 나날이 증가하고 있고 이와 더불어 강수예측의 중요성과 관심이 증가하고 있어 이와 관련된 획기적인 연구가 절실히 요구되고 있다.

지구시스템의 변화를 예측하는 과학적 방법들 중 가장 정밀하게 구성된 것이 컴퓨터 프로그램을 통한 기상모델의 모의 계산이다. 컴퓨터를 이용한 수치모델은 지구의 대기에서 일어나는 공기입자의 역학적 운동과 대류, 강수, 난류혼합,

열 복사과정 등을 설명하는 물리 과정을 수학적으로 기술한 것으로 이런 모델을 통해서 기상변화의 구체적인 원인과 결과를 정량적으로 검증하는 것이 가능하다. 포트란 언어로 작성되어있는 수치모델은 배 정밀도로 정의된 실수(64-bit real) 변수를 사용하고 많은 실수연산과 긴 실행시간의 특성을 갖는다. 일반적으로, 앞서 언급한 물리 과정을 어떻게 수학적으로 유사하게 표현하여 나타낼 것인가를 결정하는 모수화(parameterization)가 갖는 특정한 계통적 편향성 때문에 기후모의 계산의 결과는 달라진다. 각 모수화로 얻어지는 수치결과에 어떤 편향성이 있는 것인지, 그리고 이런 편향성을 어떻게 제거할 것인지 이 해결책을 모색하는 노력이 보다 정확한 기후예측을 위해서 필요하다. 그리고 수치모델을 이용하여 기상/기후를 예측할 때 그 정확성을 향상시키기 위해서는 무엇보다도 모델 가능한 모든 경우가 검토되어야 한다. 분산컴퓨팅 시스템은 그 방대한 계산력을 이용하여 모델 모수화에 대한 계산 검색을 가능하게 하므로 이러한 요구를 만족시켜줄 수 있는 거의 유일한 수단일 것이다. 또한 분산컴퓨팅 환경을 통해서 각 모수화에 대한 장단점 및 민감도를 거의 완벽하게 이해하는 것도 가능할 것이다. 따라서 지금까지와 다른 새로운 차원의 장기 기상분석에 대한 연구가 이루어 질 수 있다. 국내의 기상예측 연구 분야에서는 컴퓨팅 자원이 절대적으로 공급부족인 상태이다. 정확한 기상현상의 재현이나 예측을 위한 계산은 많은 모델 파라미터 종류들에 대한 계산이 요구되지만 그렇게 하지 못하는 것이 현실이다. 이러한 상황에서 방대한 계산력을 갖춘 인터넷 분산컴퓨팅 환경은 미래 기상분야에서의 전산모사 영역에 매우 귀한 자원이 될 것이다. (그림 2)는 계산자원을 통한 기상학 연구의 활용성을 보여주고 있다.

(그림 2) 계산 자원과 기상학 연구의 상관관계

(Figure 2) Correlation between computational resources and research in meteorology

4.2 연구개발 수행 내용 및 결과

4.2.1 인터넷 분산컴퓨팅을 활용한 고해상도 강수량예측 시스템의 개선

① 정량적 강수 진단모델의 튜닝

본 연구에서는 한반도의 상세 지역 강수량 예측을 위한 모형으로 중규모 수치 모형의 결과를 이용하여 상세 지역의 지형 효과에 의한 강수를 예측하는 QPM(Quantitative Precipitation Model)을 사용하였다. 강수량 산출 모형인 QPM은 Misumi et al.(2001), Bell(1978), Collier(1975) 등이 제안한 바 있는 Collier-type의 모형으로서 이들 모형은 소규모 지형 효과를 고려한 강수량을 산출하는 진단 모형이다[7]. (그림 3)과 같이 QPM은 중규모 예측 모형으로부터 계산된 수평 바람장, 고도, 기온, 강수 강도, 상대습도 등의 예측 자료를 이용하고 중규모 예측 모형에서는 잘 표현되지 않는 소규모 지형 효과를 고려함으로써 중규모 예측 모형에서 생산된 상대적으로 성긴 격자의 강수량 예측 값을 상세 지역의 지형을 고려한 강수량 예측 값으로 재구성한다.

(그림 3) 기상예측 시스템의 개략도

(Figure 3) Weather forecasting System Overview

중규모 예측 모형은 최소 격자 간격이 27km 정도로 모든 물리적, 역학적 과정을 포함한 모형이고 모형 구축에 많은 시간이 소요될 뿐 아니라, 계산 효율성이 저하될 수 있는 단점이 있다. 이를 개선하기 위해 QPM 중규모 모형으로부터 나온 자료를 초기 자료로 이용하고 3km 간격의 상세 지형을 반영하는 모형으로 소규모 지형 효과를 표현함으로써 상세 지역에서의 강수량 산출과 지형에 따른 강수량의 분포 파악이 용이해질 뿐 아니라, 계산 효율성을 개선시킬 수 있다[8]. 한편, 강수가 적은 시기인 겨울에서 초봄(11월~4월)에 motel의 강수가 계통적 오류에 의한 과다 예측되는 사례가 있었다. 이를 추가적으로 보완하기 위해 최대 가능 강수가 $\pm 2\alpha$ 로 경계를 두어 중규모 motel에 있을 수 있는 계통적 오류를 줄일 수 있도록 motel 내 비현실적인 강수 제어하는 프로그램을 추가하였다.

② 후처리 과정을 포함한 전 시스템의 자동화
기 고해상도 강수량 예측 시스템의 결과는 문
숫자로 마무리되었다. 하지만 이는 기상전문가들
도 이해하기에 불가능하며 격자 데이터로 된 문
숫자들은 최종적으로 조합하여 그래픽 분석을
통한 공간분포 정보 해석과정이 필요하다. 그래
서 한 격자가 3km인 한반도상의 보정된 강수에
대한 문숫자 결과를 위경도로 합성하였다. 20 x
20의 아격자 도메인으로 분산된 단위작업들을
통합하는 프로그램을 거쳐 하나의 결과를 재구
성하였고 GrADS(Grid Analysis and Display
System)를 활용하여 일련의 과정을 실행프로그
램으로 자동화하였다. 이렇게 자동으로 생성되는
이미지는 강수의 국지현상을 분석하고 원인을
파악하여 앞으로의 대책을 마련하는데 필요한

정보를 쉽게 제공할 수 있다. 고해상도 강수량예측 시스템은 기상청에서 제공해주는 원시자료 전송시간에 따라 크론탭의 자동제어로 매일 00시와 12시 2회 자료를 받아와서 인터넷 분산컴퓨팅 시스템으로 자료 전송 후, 수집된 분산시스템의 모의 결과가 매일 취합되도록 자동화하였다.

(그림 4) 인터넷 분산컴퓨팅에서 QPM의 데이터 처리과정

(Figure 4) QPM Data Processing on Internet-Distributed Computing

(그림 5) 인터넷 분산컴퓨팅 얻은 QPM의 결과 분석 및 수집

(Figure 5) Collection and Analysis of QPM Results from Internet-Distributed Computing

(그림 6) 고해상도 강수량예측 결과

(Figure 6) Results of High Resolution Rainfall Prediction

(그림 4)와 (그림 5)는 인터넷 분산컴퓨팅을 이용한 QPM 데이터 처리 및 분석, 수집 과정을 보여주고 있으며, (그림 6)은 강수량 예측 결과를 보여주고 있다.

4.2.2 성능 통계

(그림 7)과 같이 인터넷 분산컴퓨팅 환경을 통해 한달간 강수진단모델에 대한 성능을 분석하였다. 이를 위해 각 분산되어있는 계산자원에 예측 모델을 수행할 수 있는 경량의 에이전트를 설치하였다. 에이전트는 중앙관리서버로부터 공통파일과 입력파일을 다운로드 받아 작업을 수행하게 된다.

- 공통파일: 응용을 수행하기 위해 각 계산자원에 공통적으로 분배되는 파일(실행파일, 설정파일 등 입력파일을 수행하기 위해 필요)
- 입력파일: 각 계산자원에서 수행되는 유일한 최소 단위의 작업파일

(그림 7) 인터넷 분산컴퓨팅을 이용한 고해상도 강수량예측 개념도

(Figure 7) High Resolution Rainfall Prediction Using Internet-Distributed Computing

에이전트는 각각의 $3 \times 3 \text{ km}^2$ 격자의 단위의 작업을 수행한 후 중앙관리서버에게 결과 값을 보내게 된다.

(그림 8)은 일별 강수량예측을 위한 계산 성능을 보여주고 있는데 실선은 이론성능을 막대그래프는 실측성능을 나타내고 있다. 이론성능은 인터넷 분산컴퓨팅 환경에 구성된 컴퓨팅 자원의 장착된 프로세서수, 클럭주파수, 성능 요소를 고려하여 이론적으로 산출된 최대 Flops 값을 나타내며, 실측성능은 실제 강수량예측 모델을

계산하기 위해 사용된 컴퓨팅 자원의 성능을 나타낸다. 일별 성능을 살펴보면 대략 150~200대의 Agent를 활용하여 평균 150GFlops의 성능을 보이고 있다.

(그림 8) 인터넷 분산컴퓨팅을 활용한강수량예측 성능(30일간)

(Figure 8) Performance of High Resolution Rainfall Prediction Using Internet-Distributed Computing(30days)

4.2.3 강수 모의 분석 및 기대효과

한반도에서 태풍, 폭우 등과 같이 기상 이 사회 및 경제 산업분야 전반에 미치는 영향은 매우 크다. 특히 농업과 산림 등 생태계와 수자원의 변동 등은 온실기체의 증가에 따른 지구온난화에 따라 크게 변화할 것으로 예상되므로 큰 우려를 낳고 있다. 그러므로 지구환경 변화로 인한 피해를 최소화하기 위해서는 기후의 특성을 충분히 알고 제시하는 것이 중요하다. 여기서 제시된 고해상도 강수량 분석 연구 자료는 사회경제, 농업과 산림 등 생태계, 장차 에너지 분야의 수요, 수자원의 관리와 확보 등의 분야에서 지구환경 전반에 대한 장기적 정책을 수립하는 데 중요한 기초 자료가 될 수 있다. 특히 앞으로 예상되는 기상재해의 지속적인 증가 추세에 효율적으로 대응할 수 있는 방안을 마련하여 홍수, 가뭄을 비롯한 직접적인 재해에 능동적으로 대처하는데 기여할 것으로 기대된다.

5. 결 론

지구환경을 예측하기 위한 수치모델을 계산하는 컴퓨팅 자원으로써 슈퍼컴퓨터는 한계를 가

지고 있다. 여러 가지 조건들을 고려한 다양한 기상모델의 계산이 단일 슈퍼컴퓨터에서 현실적으로 불가능한 반면, 대규모 유휴 자원을 바탕으로 방대한 계산력을 갖춘 분산컴퓨팅 환경은 기상예보에 활용하기 위한 컴퓨팅 자원의 공급 부족을 해소할 수 있다[9]. 고해상도 강수량 모델은 이러한 자원을 바탕으로 보다 많이, 보다 빠르게 수행되어 정확한 미래를 예보할 수 있는 가능성을 열어준다.

정량적 강수진단 모델을 $3 \times 3 \text{ km}^2$ 격자에서 $1 \times 1 \text{ km}^2$ 격자까지 해상도를 높여 상세한 강우 예측을 시도할 경우 계산에 필요한 컴퓨팅 리소스는 3GB 이상의 메모리, 파일 저장에 필요한 2.2GB 디스크 용량 정도가 요구되었다. 일반 PC의 스펙과 사양이 고성능화됨에 따라 연산속도와 격자를 더욱 세밀하게 분석이 가능하다. 특히 GPU와 같은 가속기가 설치된 컴퓨팅 자원의 경우에는 GPGPU를 이용하여 병렬처리를 통해 벡터연산을 수행한다면 빠른 시간 안에 상세한 강우 예측이 가능하다[10].

인터넷 분산컴퓨팅 기술은 향후 기상예측분야 뿐만 아니라 다양한 응용문제들을 해결할 수 있는 가능성을 가지고 있으며, 여러 분야에 접목할 수 있도록 활성화 되어야 한다.

References

- [1] Hwang, Kai, Jack Dongarra, and Geoffrey C. Fox. Distributed and cloud computing: from parallel processing to the internet of things. Morgan Kaufmann, 2013.
- [2] Al-Azzoni, I. Down, D.G., "Dynamic scheduling for heterogeneous Desktop Grids", Grid Computing, 2008 9th IEEE/ACM International Conference on Sept. 29 2008-Oct. 1, pp.136-143, 2008.
- [3] Javadi, Bahman, et al. "Mining for statistical models of availability in large-scale distributed systems: An empirical study of seti@home." Modeling, Analysis & Simulation of Computer and Telecommunication Systems, MASCOTS'09, IEEE, 2009.
- [4] Skonnard, Aaron. "Distributed. NET-Learn The ABCs Of Programming Windows Communication Foundation." MSDN Magazine pp.44-55, 2006.

- [5] WOLTMAN, George, et al. GIMPS, The Great Internet Mersenne Prime Search. 2007.
- [6] Pease, Alison. "Folding@ home: Citizen Science Has Led to Advances in Biophysics and Fighting Disease." 2016 AAAS Annual Meeting. 2016.
- [7] Kim, Jin-Young, et al. "Prediction of rainfall with high-resolution QPF model using public-resource distributed computing." Asia-Pacific Journal of Atmospheric Sciences 44.3, pp.287-296, 2008.
- [8] Eom, Hyun-Min, et al. "Development of Coastal Inundation Forecasting System (CIFOS) and Its Application to the Future Climate Change.", Coastal Research 265, 2013.
- [9] Dongarra, Jack, and Michael A. Heroux. "Toward a new metric for ranking high performance computing systems." Sandia Report, SAND2013-4744 312, 2013.
- [10] Banerjee, Poorna, and Amit Dave. "GPGPU Based Parallelized Client-Server Framework for Providing High Performance Computation Support.", 2015.

윤 준 원

2004년 : 고려대학교 대학원 컴퓨터학과(이학석사)
 2011년 : 고려대학교 대학원 컴퓨터학과 박사 수료
 2005년~현재 : KISTI 국가슈퍼컴퓨팅연구소 선임연구원

관심분야: 분산 컴퓨팅, 결합포용시스템, 슈퍼컴퓨팅, 병렬파일시스템, 배치스케줄링

송 의 성

1997년 : 고려대학교 컴퓨터학과(학사)
 1999년: 고려대학교 대학원 컴퓨터학과(이학석사)
 2005년: 고려대학교 대학원 컴퓨터학과(이학박사)

2006년~현재: 부산교육대학교 컴퓨터교육과 교수
 관심분야: 컴퓨터교육, 교육용로봇교육, 컴퓨터네트워크, 스마트러닝