

효율적인 생물정보 서열검색을 위한 PC-클러스터 시스템 구현

공재근, 좌용권, 박정선, 유선주, 이문상*
한국보건산업진흥원
(주)이파워게이트*

{jkkong, left, pjs88, sjyoo}@khidi.or.kr, mslee@epowergate.co.kr*

Implementation of PC-Cluster System for Efficient Bioinformatics Sequence Analysis

Jae Keun Kong, Yong Gweon Jwa, Jeong Seon Park, Seon Ju Yoo, Moon-Sang Lee*
Korea Health Industry Development Institute
E-powergate, Inc.*

요 약

최근 들어 유전자 서열의 생산량 증가에 비례하여 유전자 발현 마이크로 칩과 같은 새로운 분석방법과 기술들이 도입되면서 연구자들이 매일 수천개의 서열을 효율적으로 분석해야 할 필요성이 증대되고 있다. 이러한 생명공학분야의 급속한 발전은 대용량 유전자 서열에 대한 빠른 분석이 가능한 컴퓨팅 자원을 요구하고 있으나 IT 인프라에 대한 막대한 투지비용으로 인해 관련 연구기관에서 쉽게 이들 컴퓨팅 자원을 도입하지 못하고 있는 실정이다.

본 연구에서는 저가의 PC 서버를 고속의 네트워크로 연결한 PC 클러스터를 활용하여 시스템의 안정성과 신뢰성을 보장함과 동시에 범용성을 지닌 생물정보 서열검색 시스템을 구축하였다. 이러한 효율적인 시스템 구축을 통해 생물정보 데이터베이스 및 서열 검색 시스템을 제공하고, 대용량 서열 데이터베이스의 검색 시간을 단축하였다.

1. 서론

Watson과 Crick의 염기구조 발견이후 생화학 및 분자생물학의 발달과 1990년대의 인간 유전체 프로젝트(Human Genome Project)로 인하여 생물학적 정보의 양은 급속도로 증가하게 되었다. 특히 1970년대 이후 염기나 단백질의 서열을 자동으로 분석할 수 있는 각종 기기(automatic DNA sequencer, DNA microarray, image analyser, mass spectroscopy)와 고속처리 탐색기술(high throughput screening)의 개발로 생물학적 정보의 양은 기하급수적으로 증가하고 있다. 최근에는 다양한 방식을 통해 데이터를 얻는 작업 자체를 위해서도 여러 가지 전산적인 도구들이 필요하게 되었다. PC 클러스터 시스템은 저가의 고성능 PC 및 워크스테이션을 고속 네트워크로 연결하여 고성능 서버의 성능을 발휘하는 방식으로, 저가의 상용제품을 사용함으로써 기존의 고성능 서버보다 수배에서 수십배 적은 비용으로 동일한 성능의 시스템을 구현할 수 있어 가격대 성능비가 높은 장점을 가진다. 또한, 사용자가 직접 상용 부품을 사용하여 손쉽게 확장할 수 있기 때문에 시스템 유지비용이 감소하고, 사용이 편리한 PC 및 워크스테이션의 환경을 그대로 사용할 수 있기 때문에 병렬 프로그램의 연구, 개발 및 운영이 용이하다. 이미 전세계 슈퍼컴퓨터 상위 500위중 43.8%에 해당하는 219개 컴퓨터가 클러스터 시스템 방식으로 구축되어 있으며[1], 클러스터 시스템의 장점을 활용한 효율적인 생물정보 검색 시스템은 유전체 연구기관간 유기적 네트워크 형성과 공동연구의 초석으로 활용될 수 있다. 본 연구에서는 생물정보 서열검색 시스템을 위한 클러스터 시스템의 효율적인 구조를 제안하고, 구축된 생물정보 서열검색 시스템에 대한 성능 평가를 통해 향후 개선 방향을 살펴본다.

2. 관련연구

2.1 클러스터 기반 BLAST의 단계별 병렬화

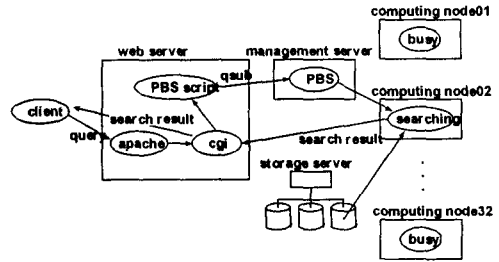
생물정보학(bioinformatics)에서 가장 기본적인 작업 중 하나는 새롭게 서열을 알게된 염기와 이미 기록되어 있는 염기서열 간의 유사성(similarity)이나 상동성(homology)을 검색하는 것이다. 이러한 작업을 통해서 연구자들은 새롭게 얻어진 유전자 서열의 단백질 종류를 예측할 수 있다. BLAST(Basic Local Alignment Search

Tool)[2]는 웹 상에서 가장 대중화되고 사용자에게 친숙한 서열 유사성 검색도구로서, 미국생명공학연구소(NCBI : National Center for Biotechnology Information)의 유전자 서열 데이터베이스(GenBank) 뿐만 아니라 SWISS-PROT이나 PDB 등과 같은 모든 주요 서열 데이터베이스를 검색할 수 있는 도구이다. 1990년대 인간 유전체 프로젝트가 본격화된 후 유전자 서열 데이터베이스 크기의 급격한 증가[3]때문에 BLAST 검색 시간의 단축이 생물정보 연구에서 매우 중요한 요소로 자리 잡게 되었다. 따라서 BLAST 검색 시간의 단축을 위해서 SMP(Symmetric Multiprocessing) 호스트 기반의 병렬화가 주로 진행되었으며, 그 결과 SMP 호스트의 쓰레드에 기반한 NCBI BLAST가 개발되었다. 하지만, 프로세서 개수가 제한되는 SMP 구조의 특성상 제한된 확장성(scalability)을 가지는 문제점이 존재한다.

R.C.Braun 등이 수행한 클러스터 시스템 상에서의 BLAST 병렬화 연구([4]-[7])에서는 BLAST 병렬화를 세 가지 방식으로 나눈다. 세립형(fine grained) 병렬화에서 다중 정렬 비교는 상호 독립적으로 이루어지기 때문에 매우 효과적이며, 특정 하드웨어를 사용할 경우 더욱 효과적이다. 중립형(medium grained) 병렬화는 계산 노드별로 데이터베이스를 분할하는 것이다. 이렇게 함으로써 데이터베이스에서 요구하는 메모리량을 줄일 수 있다. 중립형 병렬화에서 서버 노드는 사용자 질의를 분할하고 각 계산 노드에 분배하며, 검색 완료 후 각각의 결과를 취합한다. 이러한 구조는 중복성이 배제됨으로써 데이터베이스양 증가에 따라 컴퓨팅 파워를 추가할 수 있는 장점이 있다. 조립형(coarse grained) 병렬화는 BLAST 엔진 소스의 수정없이 각 계산노드에 데이터베이스 모두를 저장해 두고, 사용자 질의를 분할하여 배치처리(batch processing)하는 것이다. 이 방식은 단일 사용자 질의의 검색 시간은 동일하나 웹 서비스와 같이 동시다발적인 다수의 사용자 질의 검색 서비스에 유용하다. 하지만, 전체 데이터베이스가 각 계산 노드에 적재되기 위해서는 많은 양의 메모리가 필요하고, 사용자 질의 서열 비교시 응답 시간의 향상이 없고, 복제 데이터베이스의 일관성을 관리해야 하는 단점이 있다.

[표 1] BLAST의 단계별 병렬화

	Fine grained	Medium grained	Coarse grained
Subject(s)	1 sequence	1 sequence	N sequence
Target(s)	1 sequence	M sequence(in database)	M sequence(in database)
Parallelism	Multiple alignments on single sequence pair	Partitioned database Multiple targets examined at once	Replicated database Partition input sets



[그림 1] 클러스터 시스템 기반 생물정보 서열 검색 시스템의 질의 수행 과정

2.2 mpiBLAST

mpiBLAST[8]는 [표 1]의 구분에 따르면 증렬형 병렬화에 해당하며, BLAST 알고리즘 자체는 수정하지 않고, 프로세스 제어와 생성, 데이터 통신에만 수정을 가한 BLAST 프로그램이다. mpiBLAST의 사용은 데이터베이스가 분할되어 공유 저장장치에 저장되는 단계와 mpiBLAST 질의가 각 계산 노드에서 검색되는 두 단계로 이루어진다. mpiBLAST의 검색 단계는 검색에 필요한 데이터베이스 파티션 복사 단계와 실제 검색 단계, 그리고 결과 취합 단계로 다시 나눌 수 있다. 즉, 주어진 사용자 질의에 대해 필요한 데이터베이스 파티션을 각 계산 노드로 복사하고, 복사된 각 파티션 내에서 서열 유사성 검색을 수행한 후 각각의 파티션에 대한 검색 결과를 취합하는 과정을 거친다.

3. 생물정보 서열검색을 위한 PC-클러스터 시스템 구축

3.1 시스템 구성

본 연구에서 구축한 생물정보 서열검색 시스템은 베오울프(Beowulf) 계열의 리눅스 클러스터 시스템을 기반으로 하였다. 플랫폼에 관한 자세한 사항은 [표 2]에 나타나 있으며, 병렬화된 BLAST를 수행하기 위해 MPI(Message Passing Interface)를 사용한다. 총 4대의 역할 분담 서버, 32대의 계산 노드, 그리고 1대의 스토리지 서버가 100Mbps 이더넷(Ethernet)으로 연결되어 있으며, 역할 분담 서버는 역할에 따라 웹 서버, 관리 서버, FTP 서버, DNS 및 메일 서버로 나뉜다.

[표 2] 생물정보 서열검색 시스템의 하드웨어와 운영체제 구성

역할 분담	운영 체제	H/W 명세
서버 노드	RedHat 7.3 (커널 버전 2.4.18)	2-way P-III 1.2GHz 1GB memory Intel Ethernet Pro100
계산 노드	RedHat 7.3 (커널 버전 2.4.18)	P-IV 2GHz 1GB memory Intel Ethernet Pro100
스토리지 서버	SUN-Fire 워크스테이션	SunOS 5.8 sun4u

본 연구에서 구축한 클러스터 시스템은 각 서버와 계산 노드의 상태 확인 및 관리를 위해 ezCon[9] 클러스터 관리 도구를 사용하며, 작업 스케줄링을 위해 PBSpro[10] 큐잉 시스템을 사용한다. 정확성 보장을 위해 BLAST 알고리즘의 수정을 배제하고 배치작업의 처리율을 높이는 데 기본적인 초점을 맞추고 있다. 사용자의 질의는 웹 페이지를 통해 클러스터 시스템의 전단에 위치하는 웹 서버로 전달되고, CGI 프로그램에 의해 생성된 PBS 스크립트가 PBS 큐에 추가된다. PBS 스케줄러는 32대의 계산 노드들 중 가장 부하가 적은 노드를 선택하여 사용자 질의를 수행한다. [그림 1]은 일련의 질의 검색 과정을 도식화 한 것이다.

3.2 시스템 설정 최적화

구축한 클러스터 시스템은 사용자 질의의 폭발적 증가에 대비하기 위해 다음과 같은 최적화를 수행하였다. 첫째, TCP 연결 관리에 드는 부하를 줄이기 위해 TCP TIME_WAIT 시간을 단축하고, 소켓당 백로그(backlog) 항목수를 늘려주었다. 둘째, 원활한 NFS 동작을 위해 모든 계산 노드 및 서버 노드에 시간 동기화를 위한 NTP(Network Time Protocol)을 설치하고, 웹 서버를 NTP 서버로 활용하였다. 셋째, 작업 분배에 TCP를 사용하는 OpenPBS 사용을 배제하고 UDP를 사용하는 PBSpro를 사용하여 TCP 연결 관리 부하를 감소시켰다. 넷째, 리눅스 커널의 설정 최적화와 인터넷 데몬(inet daemon)의 설정을 최적화하였다. 이와 같은 설정 최적화를 통해 단위 시간당 더 많은 검색 서비스를 제공할 수 있으며, 사용자 질의가 특정 시간에 폭발적으로 증가하더라도 견딜 수 있는 내구성을 갖추었다.

4. 생물정보 서열검색을 위한 PC-클러스터 시스템의 성능 평가

4.1 성능평가 환경

시스템의 성능을 평가하기 위해 3절에서 언급된 클러스터 시스템 상에서 [표 3]과 같은 생물정보학 데이터베이스를 사용하였다. 비교적 작은 용량의 ecoli 데이터베이스와 대용량의 nt 데이터베이스에 대해 9226C 염기 서열과 1461C 단백질서열을 사용하여 질의검색시 걸린 시간을 측정하였다. 성능 측정 결과에서 나타난 수행 시간은 모든 사용자 질의가 수행 완료된 시간을 의미한다.

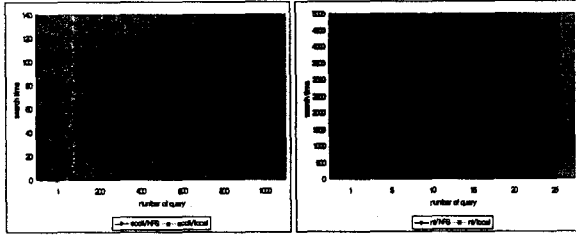
[표 3] 데이터베이스 종류와 크기

데이터베이스 종류	ecoli	nt
데이터베이스 크기	1.34MB	2.61GB

4.2 성능평가 결과

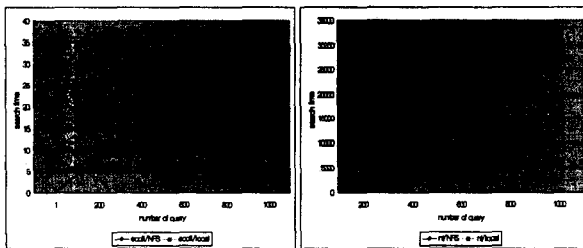
[그림 2]는 단일 계산 노드에서 서열검색을 수행할 경우 유전자 서열 데이터베이스가 계산 노드의 지역 저장 장치에 있을 경우(DB/local)와 NFS로 공유되는 저장 서버의 공유 저장 장치에 있을 경우(DB/NFS)의 검색 시간 차이를 나타낸다. 작은 용량의 ecoli 데이터베이스에 대한 검색의 경우 ecoli/local과 ecoli/NFS의 경우에 대해 검색 시간이 크지 않고, 동시에 수행되는 사용자 질의의 개수가 증가해도 검색 시간은 선형적으로 증가한다. 이러한 현상은 두 경우 모두에 대해 운영체제의 파일 캐시가 ecoli 데이터베이스를 모두 캐싱하기 때문이다. 또한, 데이터베이스 크기가 크지 않기에 스와핑(swapping)이 발생하지 않고, 이로 인해 사용자 질의의 개수가 증가해도 검색 시간은 선형적으로 증가한다. 하지만, 대용량 nt 데이터베이스에 대한 검색의 경우, 단일 사용자 질의를 검색하는데 nt/local의 경우 178.79초가 걸린 반면, nt/NFS의 경우 321.11초로 검색 시간이 약 80%정도 증가한다. 대용량의 nt 데이터베이스는 운영체제의 파일 캐시에 모두 캐칭될 수 없고, 스와핑이 빈번하게 발생하기 때문에 동시에 수행되는 사용자 질의의 개수가 증가하면 시스템 멈춤(system hanging) 현상이 발생한다. 하지만 적당한 개수의 사용자 질의가 동시에 수행될 경우 각 작업이 사용하

는 데이터베이스의 파일 캐쉬 적중으로 인한 이득이 있으며, 순차적 검색에 비해 검색 시간이 단축되는 슈퍼리니어(super-linear) 효과도 나타난다.

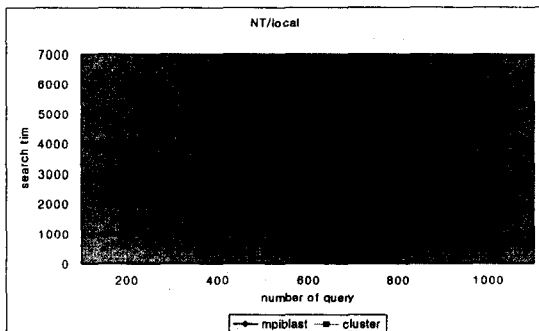


[그림 2] 단일 노드에서의 생물정보 서열검색

[그림 3]은 32대의 계산 노드로 구성된 클러스터 시스템에서 생물정보 서열검색을 수행했을 때의 검색 시간을 나타낸다. R.C.Braun의 분류에 따르면 조립형 병렬화에 해당하는 이 실험에서 사용자 질의는 각 계산 노드로 분배되어 검색된다. 단일 노드에서 단일 사용자 질의를 ecoli/local 데이터베이스에서 검색할 경우 0.12초가 걸리는데 반해, 클러스터 시스템에서는 0.36초로 3배 가량 검색 시간이 증가한다. 이는 큐잉 시스템에 검색 작업을 삽입하고, 검색 결과를 복사하는 과정에서 발생하는 오버헤드가 있기 때문이다. 하지만 1,000개의 사용자 질의에 대한 ecoli/local 검색 시간은 단일 노드와 클러스터 시스템의 경우 각각 127.15초와 37.21초로 클러스터 시스템이 우수하다. 클러스터 시스템을 사용한 검색의 경우에서도 ecoli/local과 ecoli/NFS의 검색 시간 차이는 크지 않으나, 대용량의 nt 데이터베이스에 대한 검색은 높은 I/O 트래픽으로 인한 NFS 서버의 병목 현상으로 인해 현격한 검색 시간 차이를 보인다.



[그림 3] 클러스터 시스템에서의 생물정보 서열검색



[그림 4] mpiBLAST를 사용한 생물정보 서열검색

[그림 4]는 BLAST의 증립형 병렬화 버전인 mpiBLAST를 사용하여 nt 데이터베이스를 검색했을 경우의 검색 시간을 나타낸다. 동일한 단일 사용자 질의에 대해 클러스터 시스템 상에서 조립형 병

렬화 형태로 수행했을 경우 $T_c=179.68$ 초가 걸리는데 반해, 단일 사용자 질의를 계산노드 수에 따라 분리하여 증립형 병렬화 형태로 수행할 경우 $T_m=5.19$ 초가 걸린다. 단일 사용자 질의를 증립형 병렬화를 통해 수행한 경우(T_m)가 순차적으로 수행한 경우(T_c)에 비해 32배 이상 빠르다는 사실을 알 수 있다. 이러한 슈퍼리니어(super-linear) 현상[8]은 분할된 데이터베이스가 파일 캐쉬에 모두 포함될 수 있어서 저장 장치로의 I/O 트래피이 줄어들었기 때문이다. 사용자 질의의 개수가 일정 개수이상 증가하면 증립형 병렬화 BLAST 검색과 조립형 병렬화 BLAST 검색의 검색 시간 차이가 227.1초, 328.7초, 414.8초, 563.6초, 745.8초로 꾸준히 증가하는 점을 관찰할 수 있다. 이는 사용자 질의의 개수와 상관없이 증립형 병렬화 BLAST 검색이 항상 좋은 성능을 보임을 의미한다. 전체적으로 동일한 개수의 CPU와 메모리를 사용하지만, 검색을 수행하는 계산노드의 입장에서는 증립형 병렬화 검색이 조립형 병렬화 검색보다 더 작은 크기의 분할된 데이터베이스를 검색한다. 따라서, 요구되는 데이터베이스 조각들이 운영체제의 파일 캐쉬에 포함될 확률이 높아지고, 그로 인해 저장 장치로의 I/O 트래피이 감소하기 때문이다.

5. 결론

NCBI BLAST는 SMP 호스트의 쓰레드 기전을 이용하여 병렬화되어 있다. SMP 시스템은 시스템 자체의 구조적 특성에 의해 일정수의 프로세서 이상은 확장되지 않기 때문에 제한적인 속도향상 밖에 달성하지 못한다. 반면에, 155M 초고속선도망 기반 저비용 고효율의 생물정보 서열검색을 위한 PC-클러스터 시스템은 저가의 하드웨어를 이용함으로써 SMP 시스템에 비하여 비용 대비 성능에서 효과적인 병렬 프로세서 시스템을 구현할 수 있다. 본 연구에서는 클러스터 시스템 구축시 운영체제의 최적화 방안을 제안하였고, 대용량 서열 데이터베이스를 이용한 성능평가를 통하여 우수하고, 향상된 성능과 웹을 통한 서비스[11]를 제공하였다. 이를 통해 유전체 연구기관간 유기적 네트워크 및 공동연구의 기반으로 활용될 수 있을 것으로 예상된다. 향후 연구에서는 클러스터 시스템내의 효과적인 자원 할당, 데이터베이스 크기에 따른 검색 작업 분배 방법, 다양한 생물정보 응용 프로그램의 병렬화 등에 관해 진행할 예정이다.

참고문헌

- [1] <http://www.top500.org>
- [2] <http://www.ncbi.nlm.nih.gov/BLAST/>
- [3] <http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>
- [4] R.C. Braun, K.T.Pedretti, T.L.Casavant, T.E.Scheetz, C.L.Bitkett, C.A.Roberts, "Parallelization of local BLAST service on workstations clusters", FCGS, Vol.17, pp. 745-754, 2001.
- [5] Hong Soog Kim, "Hyper-BLAST: A Parallelized BLAST on Cluster Systems", School of Engineering Information and Communication University, pp. 15-16, 2003.
- [6] David R.Mathog, "Parallel BLAST on split databases", Bioinformatics, Vol.19, No.14, pp.1865-1866, 2003.
- [7] Anne Julich, "Implementations BLAST for Parallel Computers, CABIOS, Vol.11, No.1, pp.3-6, 1995.
- [8] A.E.Darling, Lucas Carey, Wu-chun Feng, "The Design, Implementation and Evaluation of mpiBLAST", ClusterWorld 2003 conference, 2003
- [9] http://www.epowergate.co.kr/biz/clu_ezcon.html
- [10] <http://www.altair.com/software/pbspro.htm>
- [11] <http://www.biohealthinfo.org>