



1-8. 유전체분석용 슈퍼컴퓨팅 시스템 기술

클라우드컴퓨팅연구부

ETRI

Electronics and Telecommunications
Research Institute

CONTENTS

- I 기술 개요
- II 개발기술의 주요내용
- III 기술적용 분야 및 기술의 시장성
- IV 기대효과



기술 개요(1)

1. 기술개발의 필요성

● 고객 및 시장의 니즈

- (필요성)유전체 데이터 분석(1인 분석 시 누적 1TB)에서 컴퓨팅 비용은 80% 이상을 차지하므로 고성능 컴퓨팅 시스템을 필요로 함

대규모 유전체 데이터	유전체 분석 시간 단축	동시 유전체 분석 서비스	사용자 편의성 제공	저비용
데이터 규모 1인 1 TeraByte	유전체 분석시간 현재 3일	유전체 분석수요 건강검진 1,600만	전문가만을 위한 복잡한 사용환경	고가 장비, 높은 유지비용
16 ExaBytes (1,600만명)	1일 이내 유전체 분석 처리	1일 100명 이상 (1개 검진기관)	비전문가도 가능한 편리한 사용환경	낮은 도입비용 낮은 유지비용

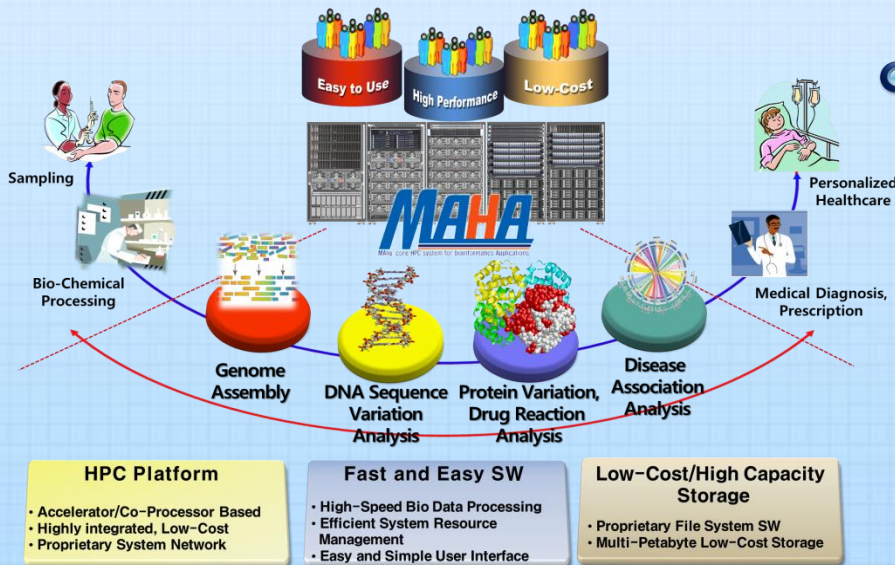
2. 기술의 개념 및 구성

기술의 개념

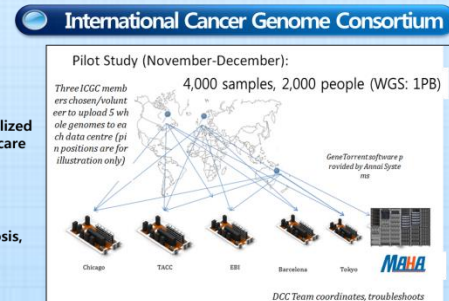
유전체 분석에 특화된 초당 10^{15} 계산이 가능한 페타급 슈퍼컴퓨터 시스템(MAHA)

* MAHA: MAny-core Hpc system for bio Application

- 매니코어 연산장치 기반의 이종 컴퓨팅 자원
- PCI Express 기반 한 고속 시스템 네트워크 연결망
- 입출력 성능한계를 극복하는 파일 시스템 및 저전력 스토리지 기능
- 이종 자원 관리 및 관리일원화를 위한 시스템 통합 관리 환경 제공
- 유전체 분석 파이프라인 및 응용 SW의 병렬화로 고성능 유전체 분석

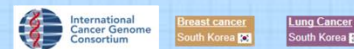


Supercomputing System for Genome Analysis



○ ETRI MAHA was selected as 6th Data Center for Cancer Genome Analysis (2013.12.)

※ Chicago Univ., TACC, EBI(Europe),
Barcelona Supercom, Tokyo(Japan), ETRI(Korea)

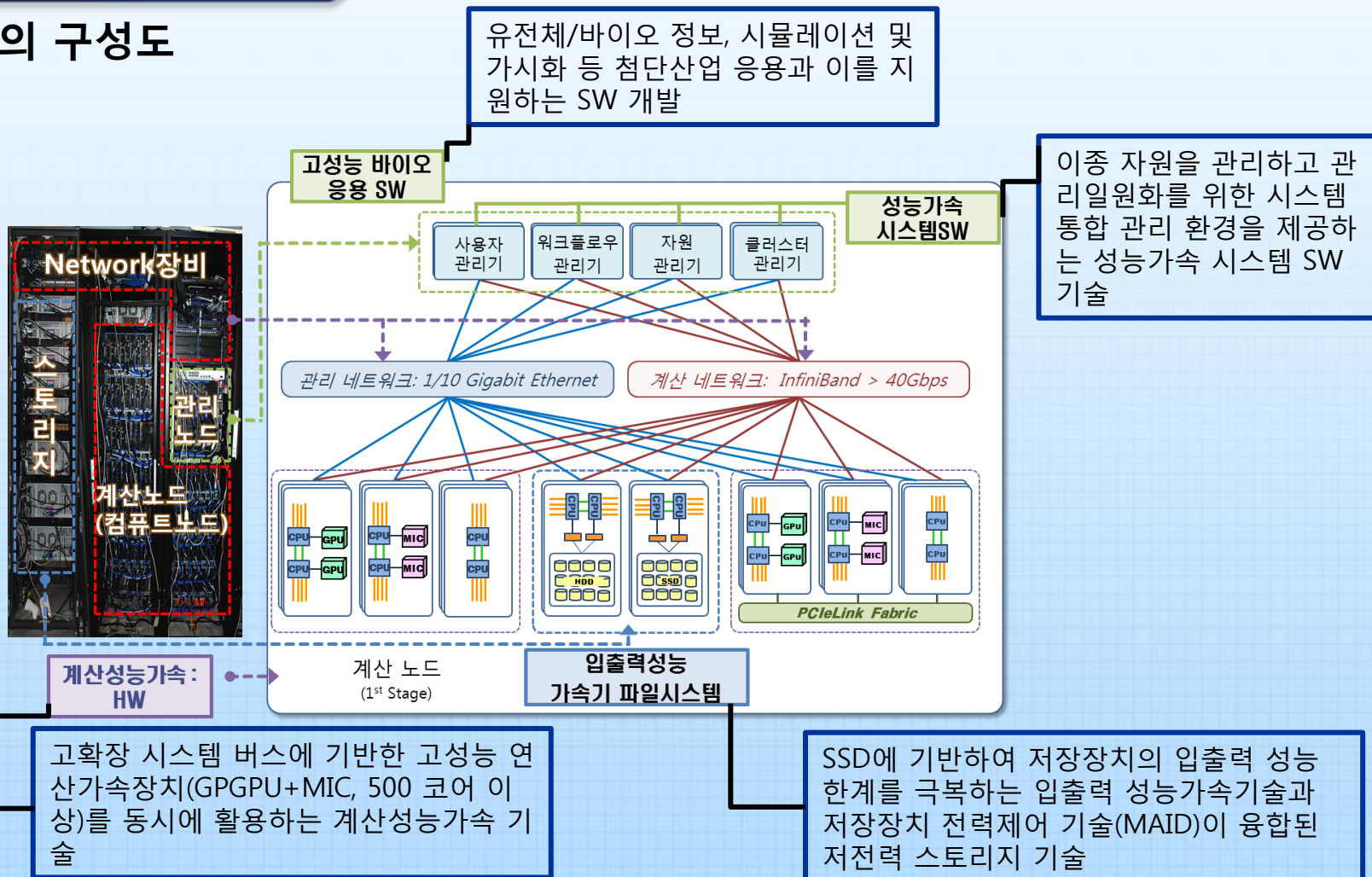




기술 개요(2)

2. 기술의 개념 및 구성

기술의 구성도





개발기술의 주요내용(1)

1. 기술의 특징

● 고객/시장의 니즈를 충족시키는 독특한 점



계산성능
가속기술
(HW)

- 매니코어 계산 노드 및
고속 시스템 네트워크에 기반한 계산 성능 가속 지원
- 매니코어 코프로세서 활용한 시스템(GPGPU/MIC)
- 저지연 고속 네트워크를 통한 시스템 성능향상
(PCIeLINK, 128/256Gbps)

입출력성능
가속기술
(스토리지
시스템)

- 기성품 스토리지 서버 기반 도입 비용 절감
- 고성능과 대용량 지원을 위한 SSD+HDD 융합 스토리지
- 미접근 HDD 및 스토리지의 전력 차단으로 소모 전력 절감

성능가속
SW기술
(시스템SW)

- 유전체 분석 성능가속을 위한 자원 관리(실행 지원)
- 쉽게 사용가능한 바이오워크플로우 기능 지원
- 사용자 친화적 시스템 사용환경 제공

바이오 응용
SW기술
(응용SW)

- 성능향상을 위한 유전체 데이터 병렬 분석 구현
- 유전체 데이터 분석 파이프라인 최적화 구현



개발기술의 주요내용(1)

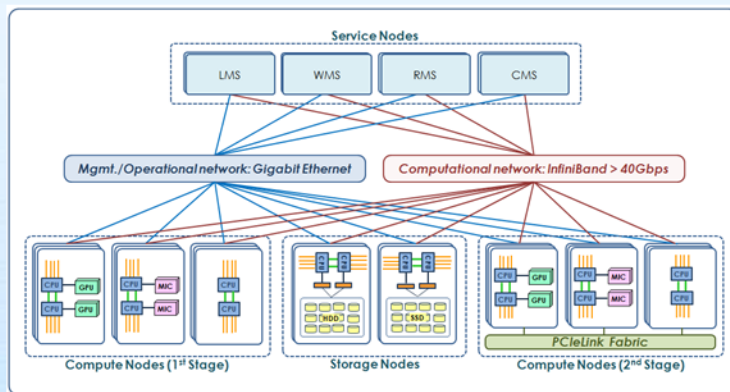
1. 기술의 특징

계산성능가속

● MAHA 시스템 HW

- 이론최대성능 : 110 TFs
 - . GPGPU / MIC 기반 계산성능가속
 - . GPGPU노드/MIC 노드/메모리 노드
- 네트워크
 - . 관리네트워크 : Ethernet 1Gbps
 - . 연산네트워크 : InfiniBand 40Gbps
 - . 연산네트워크 : PCIeLink 32/64 Gbps

● 기술의 상세 사양 유전체 분석용 슈퍼컴퓨팅 시스템



<시스템 구조도>

입출력가속

● SSD+HDD 융합 파일시스템SW

- 입출력 효율: 10 GBps/서버
- 전력 절감
 - . 미접근 HDD/서버 전원 동적 차단 (60 % 절감)



<시스템 형상>

바이오 응용SW

● 유전체 분석 파이프라인 SW

- 병렬기반 유전체 분석 파이프라인 구축
- . 유전체 정보 인덱싱 기술 최적화
- . 병렬기반 서열 정보 매핑 기술
- . 변이 정보 추출 병렬화 및 분석 기술
- . 유전체 가시화 및 검색 브라우저 구축

● 단백질 질 분석 파이프라인

- CPU/GPGPU 기반 병렬화 성능 개선
- 단백질 구조 변화율 구조 예측

시스템 SW

● 바이오 워크플로우 관리 SW

- 바이오 응용에 특화된 워크플로우 관리
- . 사용편의성과 생산성 향상

● 이종(CPU/GPGPU/MIC) 자원관리 SW

- 워크플로우 인지형 이종자원 스케줄러
- . 응용 특성에 따른 최적 자원 할당
- 시스템 통합 실행환경 관리
- . 실행환경관리 & 이종자원관리 일원화



개발기술의 주요내용(2)

2. 경쟁기술대비 우수성

● 경쟁기술/대체기술 현황

구분	경쟁기술	현황(SW 혹은 기관)
(성능가속) 시스템 SW	클러스터 통합관리	공개 SW, xCAT
	유전체분석파이프라인 성능가속 유전체 분석 실행지원	워크플로우SW(공개SW) 공개 SW (SOAP3-DP/Samtools)
(입출력성능가속) 분산 파일시스템	페타바이트급 분산 파일 시스템	미국, Lustre
시스템 구조	매니코아 기반 슈퍼컴퓨팅 구조	* 미국, 인텔 Xeon서버 * 게놈분석플랫폼: 아마존 * 게놈분석서비스: BGI (분석장비+시스템+분석)
시스템 네트워크	PCIe 기반 네트워크	(경쟁)Infiniband 40Gbps



개발기술의 주요내용(2)

2. 경쟁기술대비 우수성

● 경쟁기술/대체기술 대비 우수한 점

구분	경쟁기술	본 기술의 우수성
(성능가속) 시스템 SW	클러스터 통합관리	<ul style="list-style-type: none"> • 최대 100,000 노드 관리 가능 • non-stop 슈퍼컴퓨터 시스템 관리 지원
	유전체 분석 실행지원	<ul style="list-style-type: none"> • 유전체 분석 시간 30% 단축 (유전체 분석 파이프라인 실행 최적화, 204분 -> 138분)
(입출력성능가속) 분산 파일시스템	페타바이트급 분산 파일 시스템	<ul style="list-style-type: none"> • 구축 비용 1/2 축소 (1GBps당 2000만원 -> 1000만원) • 유전체 분석 시간 28% 축소 (GATK 87분 -> 63분) (세계 최고 입출력 성능 (1.0GBps/서버))
시스템 구조	매니코아 기반 슈퍼컴퓨팅 구조	<ul style="list-style-type: none"> • 소비전력 1/3 축소 692W/TFs -> 237W/TFs • 사용 공간 1/7 축소 3nodes/TFs -> 0.5node/TFs • 구축 비용 1/2 축소 \$8024/TFs -> \$3636/TFs • 노드당 성능 7배 332GFs/노드 -> 2332GFs/노드
시스템 네트워크	PCIe 기반 네트워크	<ul style="list-style-type: none"> • MAHA PCIe 64Gbps (국내, PCIe Gen1.1, 5 Gbps 급 수준)



개발기술의 주요내용(3)

3. 기술의 완성도

● 기술개발 완료시기 및 완성도

- 기술개발 완료시기: 2015. 9.
- 완성도: 기술이전 가능(참고: 연구소기업기술출자-2014.4.)

● 기술이전 범위 및 내용

- 유전체 분석에 특화된 슈퍼컴퓨팅 시스템(이하 MAHA) 기술
 - ✓ 고성능 시스템 구조 기술
 - ✓ 시스템 네트워크 기술
- MAHA 시스템 SW 기술
 - ✓ 바이오 응용 특성 인지형 실행지원 기술
 - ✓ MAHA 시스템을 위한 클러스터 통합 관리 기술
- MAHA 파일 시스템 기술
 - ✓ 페타바이트급 분산 파일 시스템 기술
 - ✓ 분산 파일 시스템 입출력 성능 가속 기술
- 고성능 유전체 응용SW기술
- 설계서, SW, 특허실시권



개발기술의 주요내용(4)

4. 특허

● 보유 특허

출원/ 등록 구분	특허명	출원국 (등록)	출원(등록)번호	출원(등록) 년도
출원	METHOD AND APPARATUS FOR MEMORY MANAGEMENT		13/942328	2013
출원	APPARATUS AND METHOD FOR ALLOCATING RESOURCES	중국	2013062100834890	2013
출원	프로비저닝 이미지 리비전 관리 장치, 프로비저닝 이미지 리비전 관리 방법 및 이를 이용한 프로비저닝 시스템		10-2013-0005685	2013
출원	적응형 자원 할당 및 관리 방법		10-2012-0146892	2012
등록	Method and system for managing file metadata transparent about address changes of data servers and movements of their disks		8046345	2011
등록	파일 저장 시스템 및 그 시스템에서의 파일 저장 및 검색 방법		1008345740000	2008
등록	비대칭 분산 파일 시스템과 그의 점진적 일관성 오류 감지 및 복구 방법		1010102190000	2011
등록	분산 파일 시스템에서의 비동기식 데이터 복제 방법 및 그에 따른 분산 파일 시스템		1009566370000	1010
등록	파일 입출력과 복제의 균형적 수행을 위한 지연복제 시스템 및 방법		1009362380000	1010
출원	빅데이터를 위한 병렬 정렬 장치 및 방법		10-2013-0004519	2013



기술적용 분야 및 기술의 시장성(1)

1. 기술이 적용되는 제품/서비스

● 적용 제품 및 서비스

활용분야(제품/서비스)		제품 및 활용분야 세부내용
1	(제품)유전체 분석용 슈퍼컴퓨터 시스템	<ul style="list-style-type: none"> ○ 대용량 데이터로 이루어진 유전체 분석을 위해 특화된 고성능 컴퓨팅 시스템 ○ 대규모 유전체 데이터 저장 스토리지 시스템
2	(서비스) 유전체 데이터 분석 서비스	<ul style="list-style-type: none"> ○ 대규모 유전체 데이터를 고성능 컴퓨팅 방식으로 분석하는 서비스

● 기술 적용 분야

- 고성능 휴먼 게놈 분석을 위한 HPC 시스템
- 대규모 메모리 기반 게놈 분석
- 고성능, 저비용, 저전력 스토리지

● 제품 및 서비스의 장점과 비중

- 대규모 유전체 분석 **동시 처리 시에 성능**을 가속
- **저비용**으로 대규모 유전체데이터 저장
- 유전체분석 관련하여 "시퀀싱 장비 및 재료", "재료시약", "**시퀀싱 시장**", "**생명정보학**"에서 컴퓨팅 인프라 관련(시퀀싱, 생명정보학)은 60% 차지



기술적용 분야 및 기술의 시장성(2)

2. 해당 제품/서비스 시장 규모 및 국내외 동향

● 시장 규모 및 향후 전망

활용분야(제품/서비스)		관련 시장 규모					제품기여도 (%)
		2015	2016	2017	2018	2019	
1	NGS 시장(*) (억불) (유전체 데이터 처리 및 해독)	20	24.3	29.7	36.2	44.1	60%
2	유전체 분석용 슈퍼컴퓨터(**) (억불)	4	4.3	4.8	5.3	5.9	60%
예상 시장점유율(%)						10%	

(*) 시퀀싱 장비 및 자료시장, 시퀀싱 시장, 생명정보학 시장, 재료 및 시약 시장으로 구분

(**) 슈퍼컴퓨터 시장의 5.9%를 적용함

(근거: (1)Frost&Sullivan 'European Next Generation Sequencing Market'(2011.1), 2018sus 이후는 과거 GAGR 적용 추정, (2)IDC, "A Long-Term View of the HPC Technical Server Market, 1996-2013



기술적용 분야 및 기술의 시장성(2)

2. 해당 제품/서비스 시장 규모 및 국내외 동향

● 국내외 주요 사업자 및 시장동향

<고성능컴퓨팅 및 슈퍼컴퓨팅 기술 동향>

- (미국) '91년 HPC Act. 제정 이후 연간 약 3,000억원 규모를 고성능컴퓨팅 기술에 집중 투자하여 첨단산업의 경쟁우위 확보와 글로벌 리더의 독점적 지위를 확보
- (일본) 총무성 산하, 국가과학기술회 주관으로 '06년부터 차세대 슈퍼컴 개발 추진
- (중국) 정부 주도하에 기술 독립화 추진, 2013.6월 세계 1위를 달성한 중국의 Tianhe(天河)-2는 Intel Xeon 프로세서 및 Intel Xeon Phi(MIC-2) 칩을 사용함
- (국내) 고성능컴퓨팅 기술은 컴퓨팅 기술 전반에 걸쳐 세계 최고 대비 67.0% 수준으로 미국과 4년 이상의 기술격차가 존재

<유전체분석관련 동향>

- (미국) 오바마 정부는 맞춤의학 연구개발 강화로, 2억 1000만달러를 투입 발표('15.1.), 100만개 이상의 유전자 DB 구축 사업 추진
- (영국) 커메론 수상, 10만명 환자 유전체 분석 계획('17년까지 5700억원 투자 발표, '14.8.), 암 희귀병 환자의 유전체 분석 후 항암제 맞춤치료 등 적용을 위해 시퀀싱, 정보저장분석 등 투자
- (중국) 베이징 게놈 연구소는 인간 유전체 다양성 연구, 1000 plant genome project 수행
- (일본) RIKEN(이화학연구소) '혁신적 세포해석 프로그램('09년 693억원 지원) 질병 및 유전체 변이 연구
- (국내) ㈜신테카바이오는 2014.4., 한국전자통신연구원으로부터 연구소기업 출자기술을 이전 받아 맞춤형 뉴트리션, 질병예측 솔루션을 출시하여 병원 등 비즈니스 추진 중
- (국내) 마크로젠은 유전체분석 기술과 인프라를 활용하여 인간 게놈 정보의 실용화를 가능케 할 '개인별 1000불 유전체분석 시장'을 준비하고 차세대 유전체 분석 기술을 바탕으로 맞춤의학 준비 중

기술 도입 효과

● 기술 도입으로 인한 경제적 효과

- 경제적 파급효과는 '17년~'20년까지 약 5조 7,790억원의 생산유발, 2조 4,141억원의 부가가치, 3만 4,193명의 고용유발 효과 창출 전망
(*한국은행, 2011년 산업연관표(2013.7 발표)를 활용하여 관련 계수 산출, 중립적 전망에 대한 경제적 파급효과)
- 미래 맞춤형 의료 실현으로 불필요한 치료·약물 부작용을 줄이고 예방의학 지원으로 보건 지출 비용 및 의료비 절감

● 기술사업화로 인한 파급효과

- 2020년 개인 맞춤형 의료 서비스 견인 – 유전체 시퀀싱 및 생명정보학 시장은 2010년 43%에서 2016년 이후 60% 이상이 예상되므로 “고성능컴퓨팅 인프라” 역할은 생산성 증대로 직결됨

〈분야별 시퀀싱 관련 시장 규모의 성장 전망〉

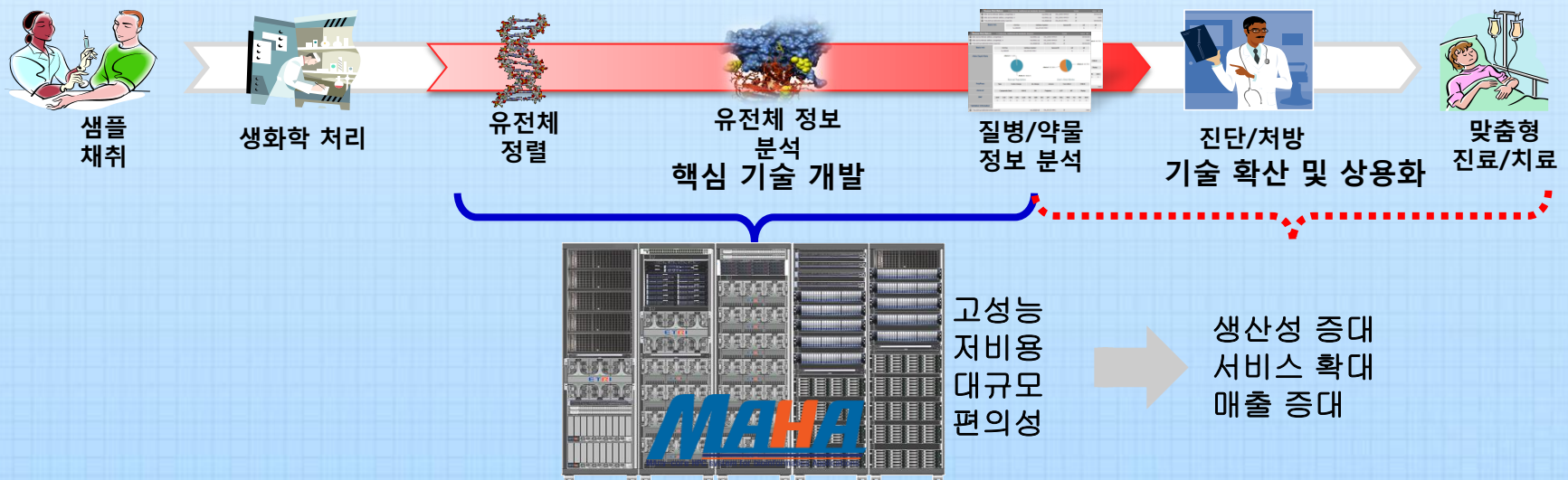
(단위 : 백만 달러)

구분	2010년	2011년	2016년	연평균 성장률(%)
시퀀싱 장비 및 재료 시장	1,245.6	1,570.9	2,229.5	7.3
시퀀싱 시장	837.6	987.6	3,528.6	29.0
생명정보학 시장	196.5	260.7	573.3	17.1
샘플 준비 관련 재료 및 시약	99.7	149.2	310.3	15.8
합계	2,379.4	2,968.4	6,641.7	17.5

출처 : BOC Research

기술 도입 효과

- 기술 도입 환경
 - ✓ 유전체 분석 산업은 특성상 경쟁력 유지를 위해 컴퓨팅 인프라와 같은 투자가 요구되는 분야이며, 국내는 시장 태동기에 해당함
 - ✓ 유전체 전문 기업이 국내 의료연구기관과 연계하여 개인 유전체 분석 서비스를 진행할 경우 해외 유전정보와 차별화된 분석서비스 시장 가능성이 큼
- 기술적 및 경제적 기대효과
 - ✓ 유전체 분석 기업은 차세대 유전체 해독 기술(NGS)이 요하는 높은 컴퓨팅 성능을 통하여 전체 유전체 분석 성능을 약 30% 개선할 수 있어, 그 결과 생산성 증대를 가져올 수 있고,
 - ✓ 자사 유전체 분석 솔루션을 통한 서비스 확대로 매출 증대에 기여 가능함



감사합니다

최 완 부장, 042-860-6315 wchoi@etri.re.kr

