

**ETRI** 한국전자통신연구원

# 유전체 분석용 슈퍼컴퓨팅 시스템(MAHA)

MAHA : MAnycore Hpc for bio Application



## Summary

Lab 소개	<ul style="list-style-type: none"><li>유전체 분석용 슈퍼컴퓨팅 기술 개발</li></ul>
기술 소개	<ul style="list-style-type: none"><li>본 기술은 유전체 분석에 특화된 초당 <math>10^{15}</math> 계산이 가능한 페타급 슈퍼컴퓨터 시스템(MAHA)</li></ul>
기술 동향	<ul style="list-style-type: none"><li>슈퍼컴퓨터 기술은 최근에 출원 건수가 다소 하락하는 추세지만, 전체적으로는 연구개발이 활발한 성장기 단계의 기술로 분석됨</li><li>국내에서는 슈퍼컴퓨터 자체구축을 시도했지만, 시스템 핵심기술에 대한 연구개발이 부족하여 선진국과 기술격차를 줄이지 못함</li></ul>
시장 동향	<ul style="list-style-type: none"><li>유전체 분석용 슈퍼컴퓨터는 전체 슈퍼컴퓨터 시장의 약 5.9%를 차지할 것으로 예측되며, 2015년 4억불에서 2019년 5.9억불로 성장할 전망</li><li>미국 오바마 정부는 맞춤의학 연구개발 강화로, 2억 1000만 달러를 투입 발표('15.1.), 100만개 이상의 유전자 DB 구축 사업 추진</li></ul>
협력 사업	<ul style="list-style-type: none"><li>본 기술의 공급자인 한국정보통신연구원은 컴퓨팅 R&amp;D 인력을 기반으로 관련 사업 수행 경험이 있는 기술 수요자들을 대상으로 기술제공 및 공동연구개발 등을 통해 협력사업 추진이 필요할 것으로 예상</li></ul>

1. Lab 소개
2. 기술소개
3. 환경분석
4. 사업화 전략
5. 비즈니스 모델
6. 협력방안

# ETRI SW·콘텐츠연구소 SW기반기술연구본부 클라우드컴퓨팅연구부

VISION : 신뢰받는 국가 SW·콘텐츠 연구소

목표

창의적인 SW 융합 R&D를 통한 신시장 창출

- 유전체 분석용 슈퍼컴퓨팅 시스템 개발
- 클라우드 DaaS 시스템 및 단말기술 개발
- 가상 데스크탑 서비스용 클라우드 파일 시스템 기술
- IT 에너지 30% 절감을 위한 에너지 인지 범용 운용체제 핵심원천기술 개발
- 기가 미디어 기반 Tele-Experince 서비스 SW플랫폼 기술 개발
- 다수의 이종 클라우드 자원을 통합 관리하는 클라우드 서비스 브로커 기술 개발



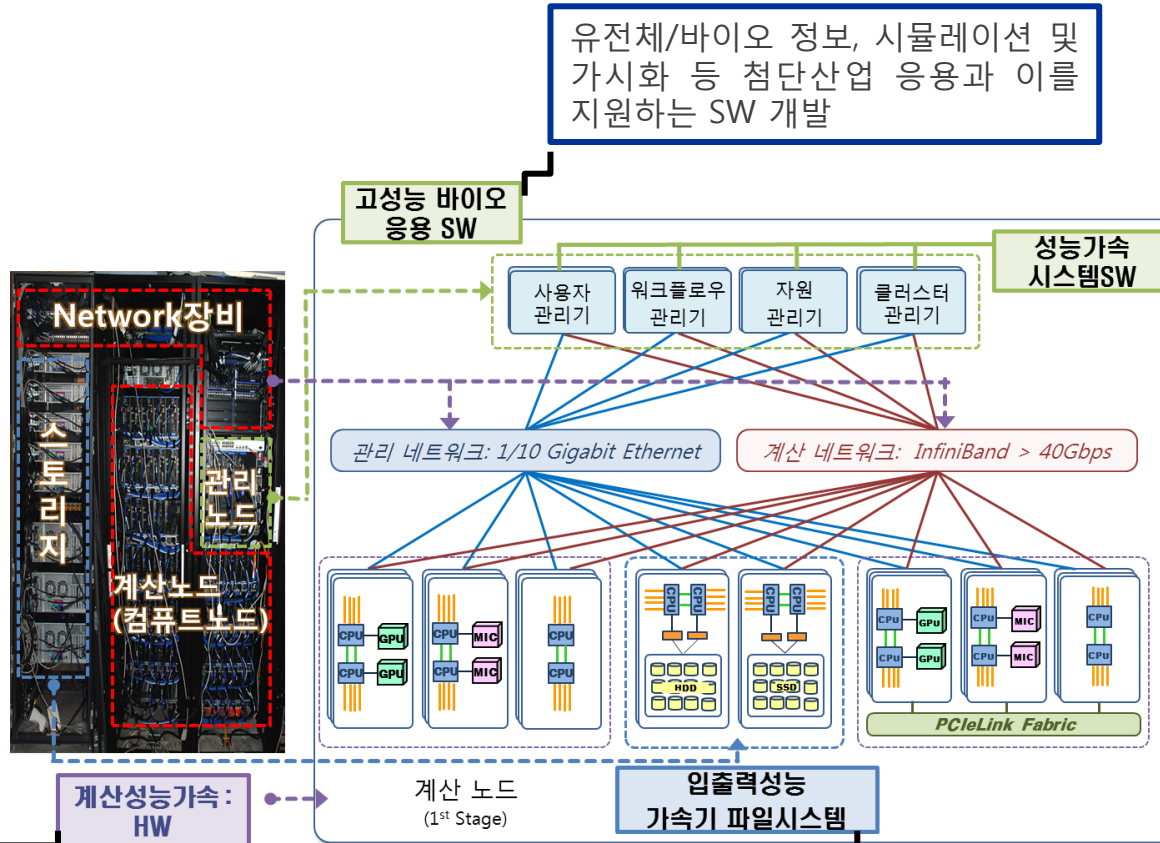
## □ 기술의 간략한 소개

- 본 기술은 유전체 분석에 특화된 초당  $10^{15}$  계산이 가능한 페타급 슈퍼컴퓨터 시스템(MAHA)
  - 매니코어 연산장치 기반의 이중 컴퓨팅 자원
  - PCI Express 기반한 고속 시스템 네트워크 연결망
  - 입출력 성능한계를 극복하는 파일 시스템 및 저전력 스토리지 기능
  - 이중 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

□ 기술 구성도



이종 자원을 관리하고 관리 일원화를 위한 시스템 통합 관리 환경을 제공하는 성능가속 시스템 SW 기술

고확장 시스템 버스에 기반한 고성능 연산가속장치(GPGPU+MIC, 500 코어 이상)를 동시에 활용하는 계산성능 가속 기술

SSD에 기반하여 저장장치의 입출력 성능 한계를 극복하는 입출력 성능가속 기술과 저장장치 전력제어 기술(MAID)이 융합된 저전력 스토리지 기술

### □ Feature / Benefit 분석

#### ▪ 특징

- 이기종 연산 가속 HW
- 고성능, 저비용, 저전력 화일시스템 SW
- 유전체 분석 응용에 특화된 HPC 시스템 SW
- 고성능 휴먼 게놈 분석 응용

#### ▪ 장점

- 분산 병렬 파일 시스템 : 구축 ½비용, SSD+HDD 융합, 1.3GBps/서버 (28% 개선)
- 바이오 워크플로우 및 시스템 사용률 개선, 게놈 분석 성능 개선 (30% 개선)

유전자 분석을 위한 슈퍼컴퓨터 시스템

## □ 목표제품/서비스

분야	유전체 분석 슈퍼컴퓨터 시스템	슈퍼컴퓨팅 시스템
제품	<ul style="list-style-type: none"> <li>슈퍼컴퓨터 시스템-</li> <li>유전체 분석 슈퍼컴퓨터 시스템</li> </ul>	<p>기능</p> <ul style="list-style-type: none"> <li>클러스터 통합관리</li> <li>메모리 확장</li> <li>분산 파일 시스템</li> </ul>

서비스	<b>대규모 유전체</b> 데이터	유전체 분석 <b>시간 단축</b>	<b>동시</b> 유전체 분석 서비스	사용자 <b>편의성</b> 제공	저비용
	데이터 규모 1인 1TeraByte	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



## □ 기술완성도(TRL 단계)

- 본 기술은 시작품 인증 / 표준화가 완료된 TRL 6단계

TRL 9	사업화	<ul style="list-style-type: none"> <li>▪ 본격적인 양산 및 사업화 단계</li> </ul>
TRL 8	시작품 인증/ 표준화	<ul style="list-style-type: none"> <li>▪ 일부 시제품의 인증 및 인허가 취득 단계 - 조선 기자재의 경우 선급기관 인증, 의약품의 경우 식약청의 품목 허가 등</li> </ul>
TRL 7	Pilot 단계 시작품 신뢰성 평가	<ul style="list-style-type: none"> <li>▪ 시작품의 신뢰성 평가</li> <li>▪ 실제 환경(수요기업)에서 성능 검증이 이루어지는 단계</li> </ul>
TRL 6	Pilot 단계 시작품 성능 평가	<ul style="list-style-type: none"> <li>▪ 경제성(생산성)을 고려한, 파일럿 규모의 시작품 제작 및 평가</li> <li>▪ 시작품 성능평가</li> </ul>
TRL 5	시제품 제작/ 성능평가	<ul style="list-style-type: none"> <li>▪ 개발한 부품/시스템의 시작품(Prototype) 제작 및 성능 평가</li> <li>▪ 경제성(생산성)을 고려하지 않고, 우수한 시작품을 1개~수개 미만으로 개발</li> </ul>
TRL 4	연구실 규모의 부품/시스템 성능평가	<ul style="list-style-type: none"> <li>▪ 연구실 규모의 부품/시스템 성능 평가가 완료된 단계</li> <li>▪ 실용화를 위한 핵심요소기술 확보</li> </ul>
TRL 3	연구실 규모의 성능 검증	<ul style="list-style-type: none"> <li>▪ 연구실/실험실 규모의 환경에서 기본 성능이 검증될 수 있는 단계</li> <li>▪ 개발하려는 시스템/부품의 기본 설계도면을 확보하는 단계</li> <li>▪ 모델링 / 설계기술 확보</li> </ul>
TRL 2	실용 목적의 아이디어/특허 등 개념 정립	<ul style="list-style-type: none"> <li>▪ 실용 목적의 아이디어, 특허 등 개념 정립</li> </ul>
TRL 1	기초 이론/실험	<ul style="list-style-type: none"> <li>▪ 연구과제 탐색 및 기회 발굴 단계</li> </ul>

## □ 지식재산권 현황

연번	특허종류	등록번호	제목
1	등록특허	10-1010219	비대칭 분산 파일 시스템과 그의 점진적 일관성 오류 감지 및 복구 방법
2	등록특허	10-0956637	분산파일 시스템에서의 비동기식 데이터 복제 방법 및 그에따른 분산파일 시스템
3	등록특허	10-0936238	파일 입출력과 복제의 균형적 수행을 위한 지연복제 시스템 및 방법
4	등록특허	10-2013-0004519	빅 데이터를 위한 병렬 정렬 방법 및 장치

## □ 지식재산권 현황

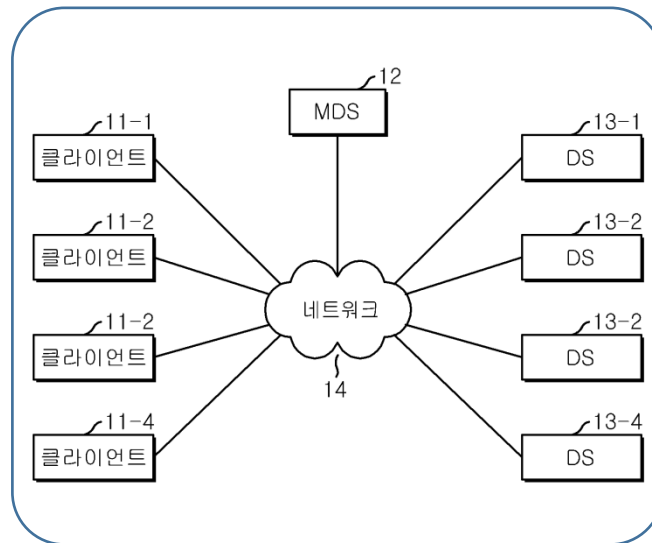
1

특허명 : 비대칭 분산 파일 시스템과 그의 점진적 일관성 오류 감지 및 복구 방법

출원번호 : 10-2008-0127872

등록번호 : 10-1010219

초록 : 본 발명은 네트워크 기반의 비대칭 분산 파일 시스템에서 파일 시스템의 구조적 일관성을 유지할 수 있는 점진적 일관성 오류 감지 및 복구 방법을 제공한다.



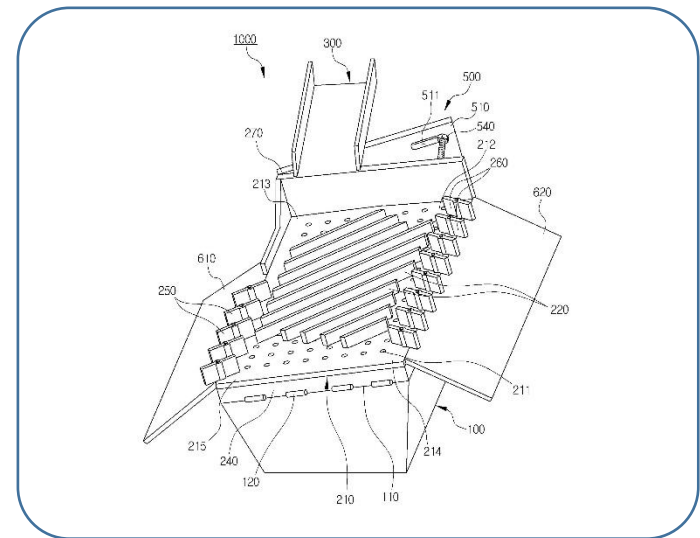
2

특허명 : 분산파일 시스템에서의 비동기식 데이터 복제 방법 및 그에따른 분산파일 시스템

출원번호 : 10-2007-0126926

등록번호 : 10-0956637

초록 : 본 발명은 모든 복제본이 동일한 데이터를 유지할 수 있도록 하는 분산파일 시스템에서의 비동기식 데이터 복제 방법 및 그에 따른 분산파일 시스템에 관한 것이다.



## □ 지식재산권 현황

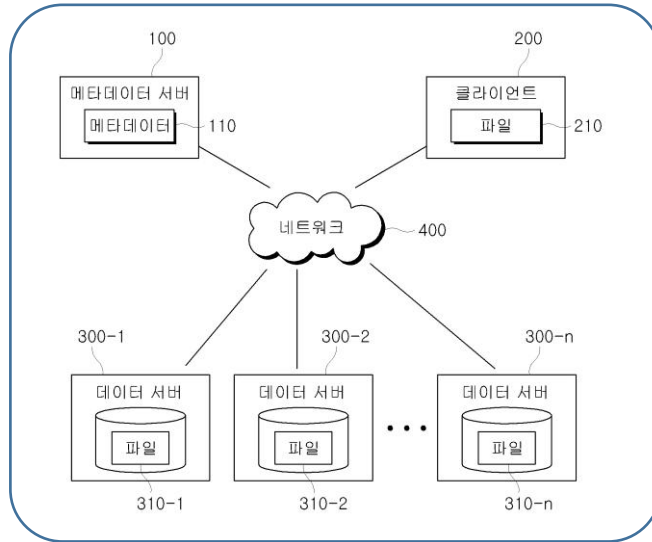
3

특허명 : 파일 입출력과 복제의 균형적 수행을 위한  
지연복제 시스템 및 방법

출원번호 : 10-2007-0129209

등록번호 : 10-0936238

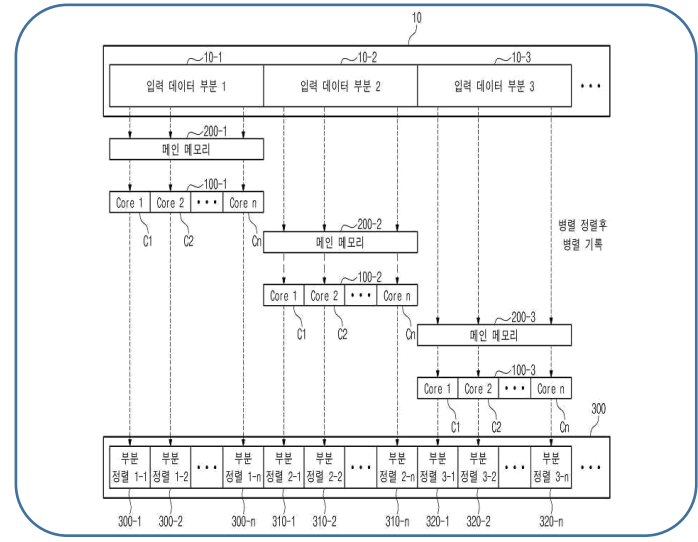
초록 : 본 발명은 파일 입출력과 복제의 균형적 수행을 위한  
지연 복제 시스템 및 방법을 제공하는 것이다.



4

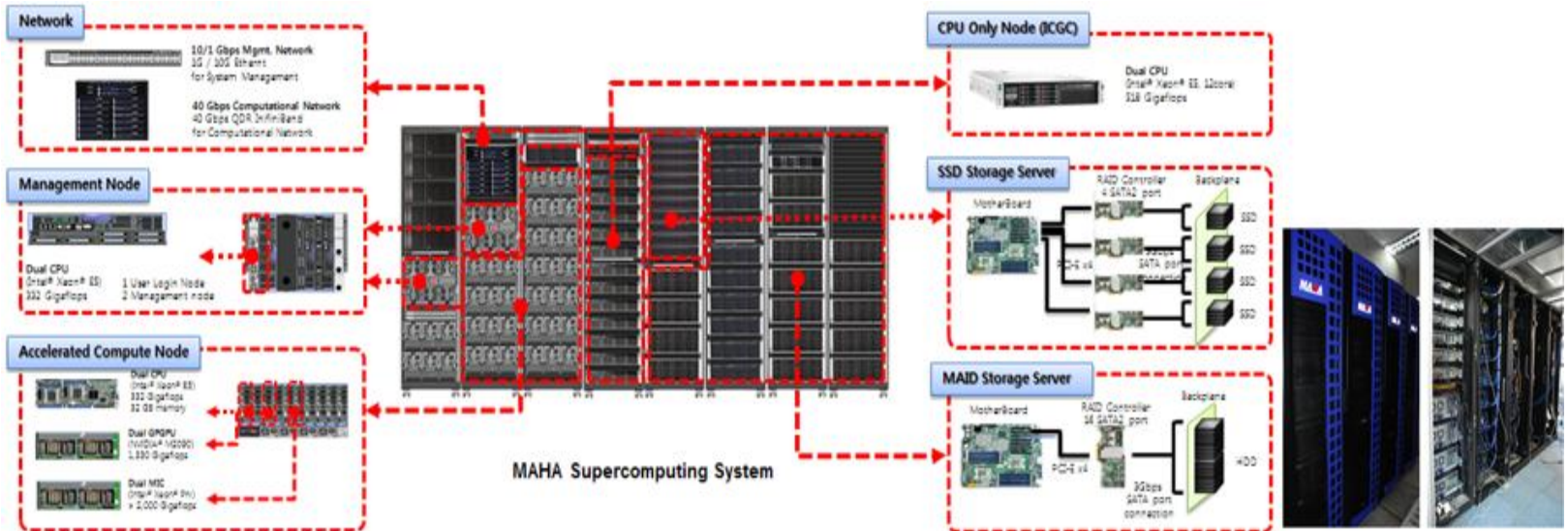
특허명 : 빅 데이터를 위한 병렬 정렬 방법 및 장치  
출원번호 : 10-2013-0004519

초록 : 본 발명은 정렬되지 않는 모든 데이터 개체(element)를  
시스템이 가진 메모리에 올려서 정렬할 수 없는 문제를  
해결하기 위한 것이다.



## □ 응용분야(SFN 분석)

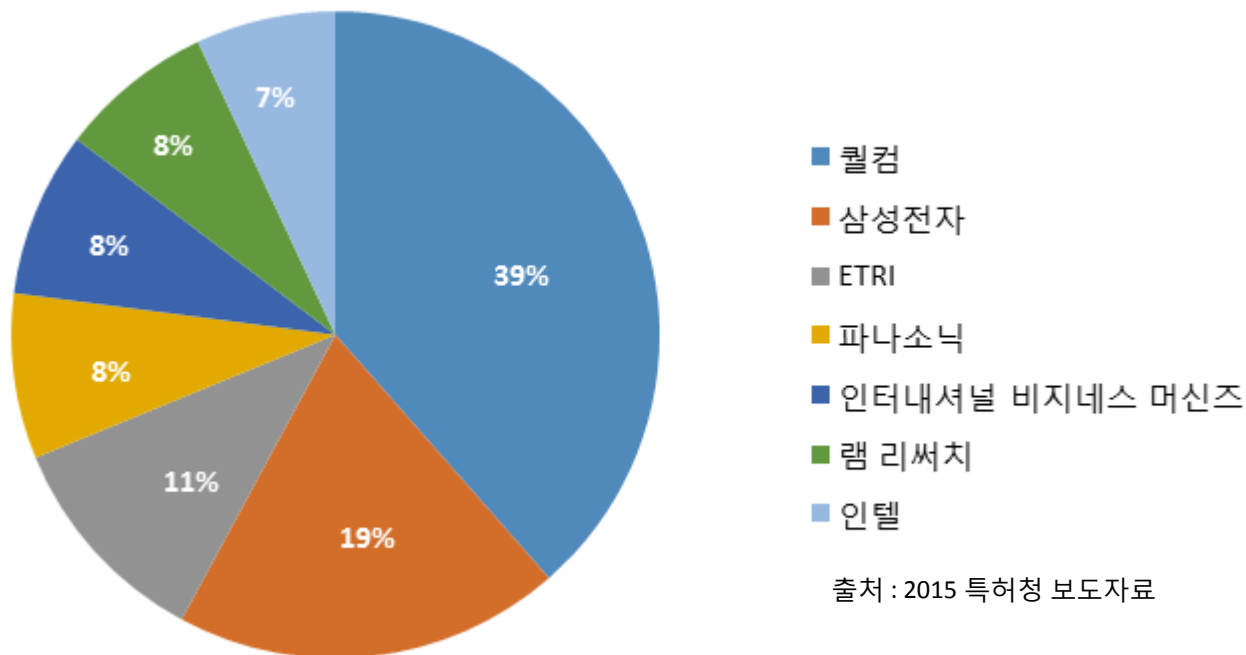
- Seed
  - 비대칭 분산 파일 시스템과 그의 점진적 일관성 오류 감지 및 복구 방법 외 3건 (등록특허 10-1010219)
- Function
  - 이기종 연산 가속속도 향상
  - 고성능, 저비용, 저전력 화일시스템
  - 유전체 분석 응용에 특화된 HPC 시스템
- Need
  - 고성능 휴먼 게놈 분석을 위한 HPC 시스템(슈퍼컴퓨터 시스템)



<MAHA 시스템>

## □ 국내 특허 출원 동향

- 국내 주요 출원인은 퀄컴이 39%로 가장 많았으며, 그 다음은 삼성전자, ETRI 순으로 나타남
- 국내에서는 슈퍼컴퓨터 자체구축을 시도했지만, 시스템 핵심기술에 대한 연구 개발이 부족하여 선진국과 기술격차를 줄이지 못함



< 국내 주요 출원인 >

□ 경쟁기술 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/Samttools)
(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

□ 목표시장

- 대규모 유전체 분석 동시 처리 시에 성능을 가속
- 저비용으로 대규모 유전체데이터 저장

● 적용 제품 및 서비스

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

● 기술 적용 분야

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



□ 유전체 분석용 슈퍼컴퓨터 시장 동향

- 유전체 분석용 슈퍼컴퓨터는 전체 슈퍼컴퓨터 시장의 약 5.9%를 차지할 것으로 예측되며, 2015년 4억 달러에서 2019년 5.9억 달러로 성장할 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%

(\*) 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), 2018년 이후는 과거 GAGR 적용 추정,

(2)IDC, “A Long-Term View of the HPC Technical Server Market, 1996-2013

유전체 분석을 위한 슈퍼컴퓨터 시장 성장 전망

## □ 국내외 주요 사업자 및 산업동향

### <고성능컴퓨팅 및 슈퍼컴퓨팅 기술 동향>

- (미국) '91년 HPC Act. 제정이후 연간 약 3,000억 규모를 고성능컴퓨팅 기술에 집중 투자하여 첨단산업의 경쟁우위 확보와 글로벌 리더의 독점적 지위를 확보
- (일본) 총무성산하, 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 세계최고 대비 67.0% 수준으로 미국과 4년 이상의 기술격차가 존재



### □ 국내외 주요 사업자 및 산업동향

#### <유전체 분석 관련 동향>

- (미국) 오바마 정부는 맞춤의학 연구개발 강화로, 2억 1000만 달러를 투입 발표('15.1.), 100만개 이상의 유전자 DB 구축 사업 추진
- (영국) 커메론 수상, 10만명 환자유전체 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) (주)신테크바이오는 2014. 4., 한국전자통신연구원으로부터 연구소기업 출자기술을 이전받아 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

유전체 분석을 위한 DB 구축 및 컴퓨팅 시스템 구축 시장 확대

□ 주요 시장 참여자

슈퍼컴퓨터 서버 솔루션

기업명	URL	대표자	주요 수행 사업
(주)클루닉스	www.clunix.com	권대석	아르콘서버, 소프트웨어고속메시징기술, 웹가속기 프로토타입 클러스터솔루션엔클러스터
(주)제노닉스	www.zeronex.co.kr	김유환	무선통신장비 제조/소프트웨어, 하드웨어 개발
(주)LG U+	www.uplus.co.kr	이상철	개인휴대통신서비스, 음성서비스, 데이터서비스, 프로그램 개발
삼성 SDS(주)	www.sds.samsung.co.kr	전동수	컴퓨터 프로그래밍, 시스템 통합, 관리/소프트웨어 개발
(주)KT	www.kt.com	황창규	유선통신사업, 인터넷, 전자상거래, 네트워크
(주)이노그리드	www.innogrid.com	조호건	시스템소프트웨어 개발, 공급

## □ STP 분석

### Segmentation

대분류	중분류	소분류
출판업	소프트웨어 개발 및 공급업	시스템 응용 소프트웨어 개발 및 공급업
컴퓨터 프로그래밍, 시스템 통합 및 관리업	컴퓨터 프로그래밍, 시스템 통합 및 관리업	기타 정보기술 및 컴퓨터운영 관련 서비스업

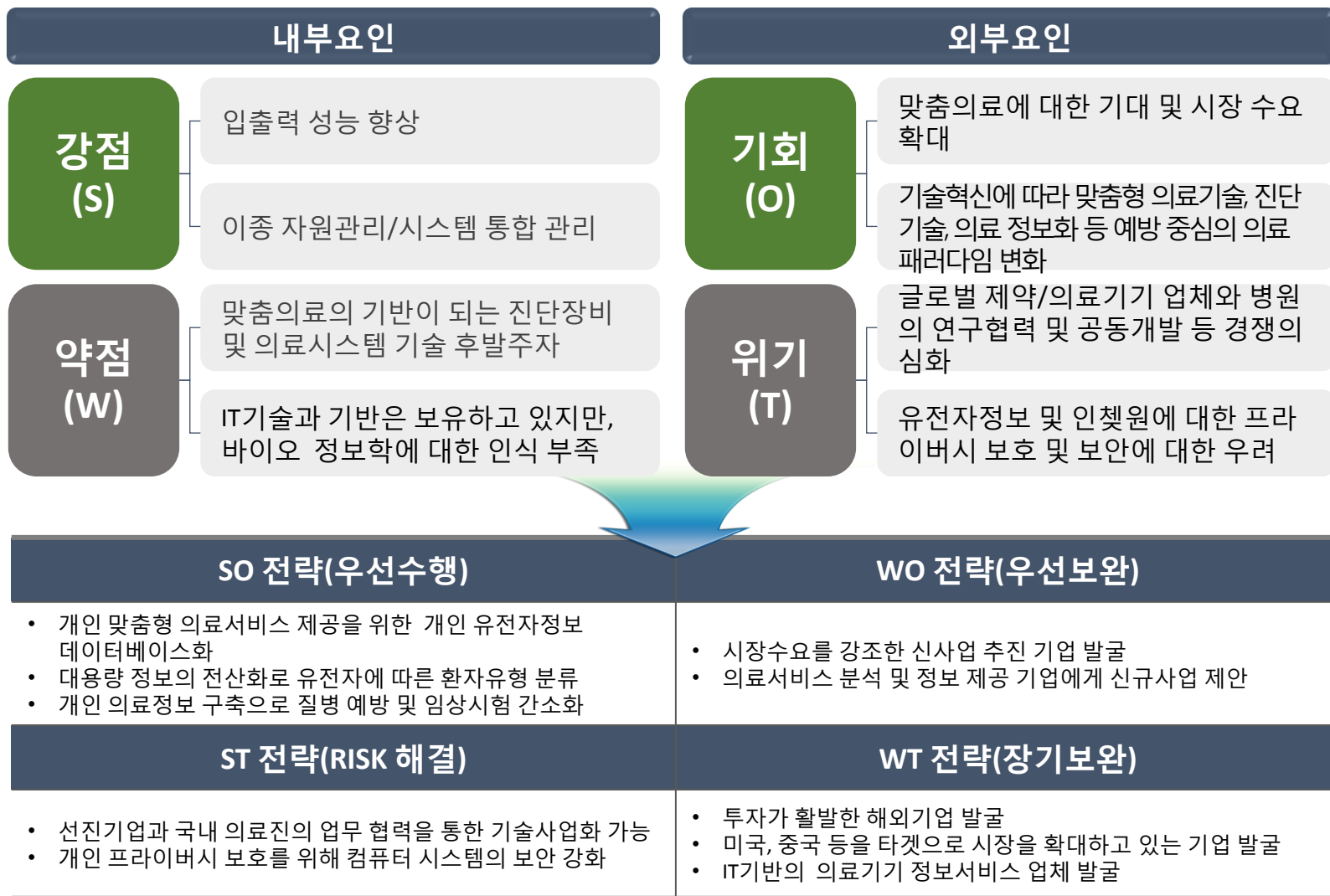
### Targeting

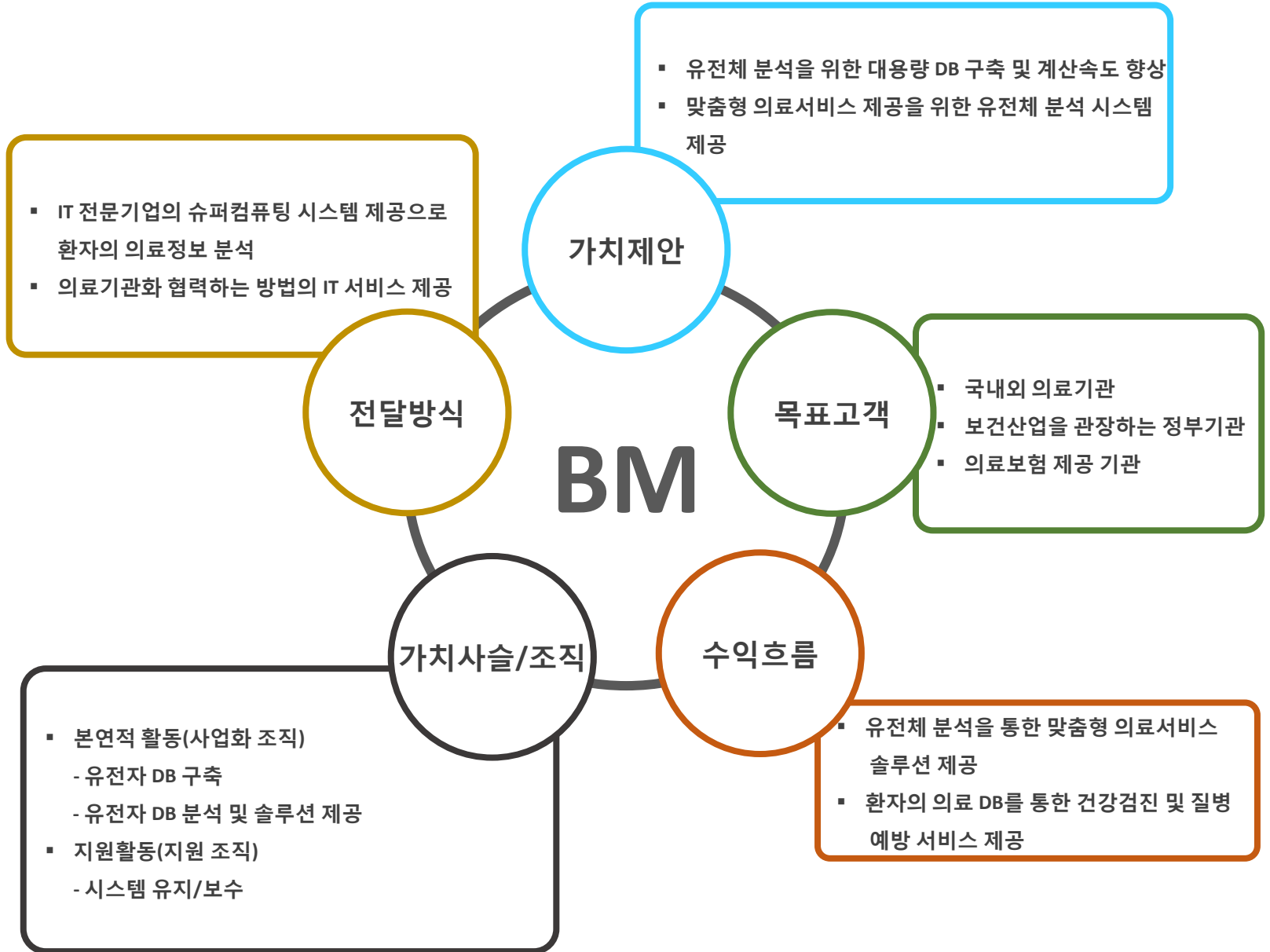
적용분야	산업 요구사항	타겟기업	
슈퍼컴퓨팅 서버 솔루션	<ul style="list-style-type: none"> <li>대용량 데이터의 신속한 산술처리</li> </ul>	대 / 중견기업	(주)KT, 삼성 SDS(주), (주)LG U+
		중소기업	(주)클루닉스, (주)제노닉스, (주)이노그리드

### Positioning

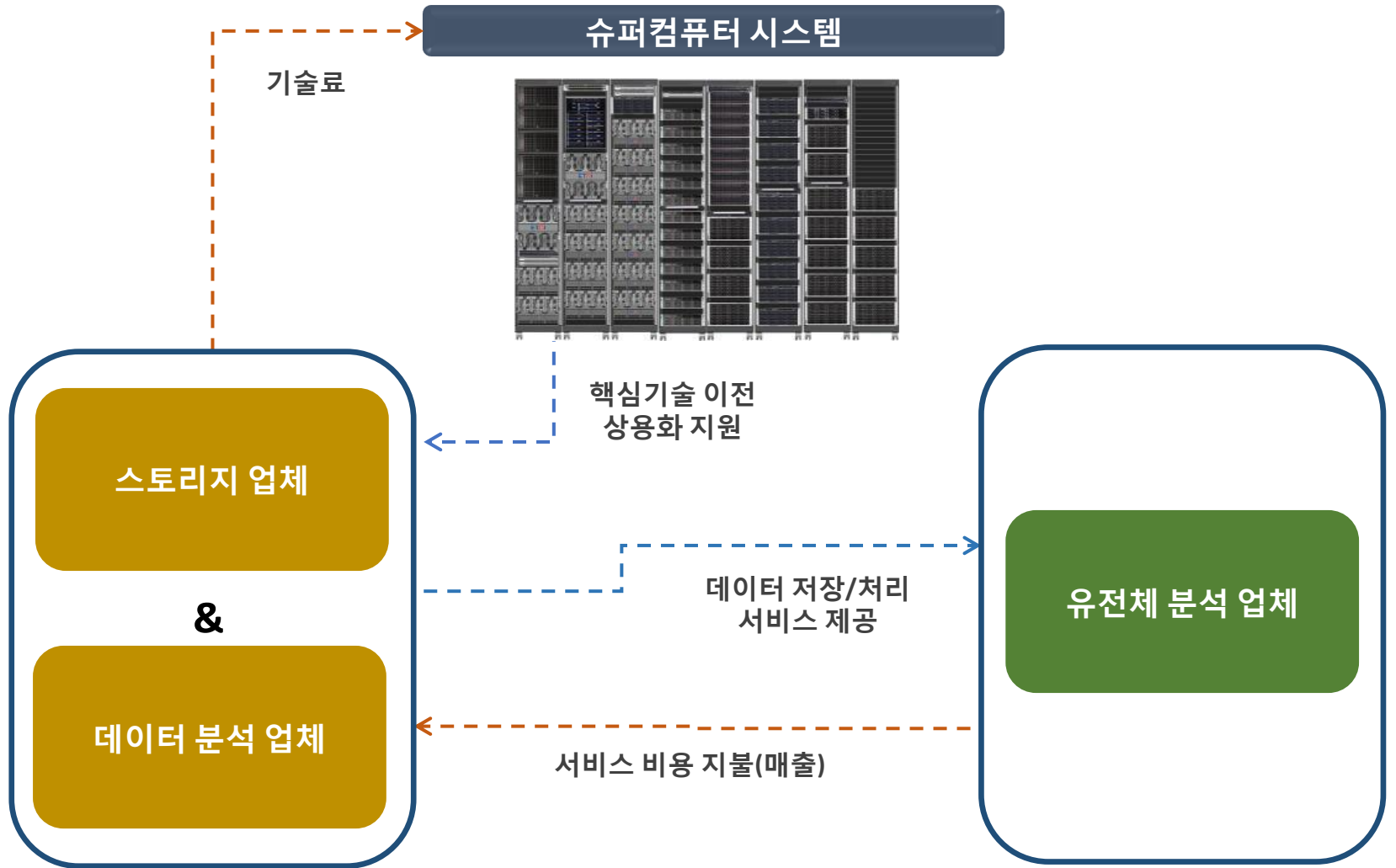
경쟁요소	DB 저장/분석 속도	유전자 분석	타겟기업	Selling Point
포지셔닝	높음	높음	삼성SDS	<ul style="list-style-type: none"> <li>삼성의료원 등과 라이프 테크노로지스를 구현하고 있어 해당기술의 구현에 성능향상 기술 거래 제의</li> </ul>
	높음	보통	클루닉스	<ul style="list-style-type: none"> <li>본 기술과 유사한 기술을 보유하고 있어 기술의 포트폴리오 확보 제안</li> </ul>
	높음	낮음	이노그리드	<ul style="list-style-type: none"> <li>클라우드와 바이오기술이 결합한 기술에 대해 관심이 높아 기술 관심도 높을 것으로 예상</li> </ul>
	높음	보통	마크로젠	<ul style="list-style-type: none"> <li>본 기술과 유사한 사업을 준비 중으로 기술에 대한 수요가 높을 것으로 예상</li> </ul>

□ SWOT 분석





□ 기술사업화 수익구조





□ 협업 사항

