

Bigdata Cluster Overview V.1.0

Cubedoop Cluster™

CubePi Cluster™

For more detail
daengky@bicube.co.kr

Cubedoop Cluster™

BI Cube는 빅데이터 머신러닝 플랫폼을 보유한 회사로서 빅데이터 머신러닝 플랫폼 내에 파일을 저장 하고 MapReduce를 실행 해야 하는 요소가 필수였다. 이에 BI Cube는 플랫폼 내에서 안정성과 성능, 그리고 효율성을 위해서 Apache Hadoop을 이용해 **Cubedoop Cluster™**라는 빅데이터 클러스터 플랫폼을 만들었으며, 자체 빅데이터 클러스터를 개발 운영하면서 축적한 노하우를 바탕으로 **Cubedoop Cluster™**를 상품화 했다.

Cluster Appliance

Cubedoop Cluster™ 는 Rack 단위로 확장해 갈 수 있는 S/W & H/W 일체형의 Cluster Appliance이다.

일단 Cluster 전문가에 의해 고객의 문제 해결에 적합한 클러스터 사이즈가 결정되면, 성능 최적화 H/W Spec을 따라 구성된 Cluster 위에 최적화 Parameter를 갖는 Cubedoop이 얹어진 **Cubedoop Cluster™**이 고객에게 제공 된다.



표준 Rack 3대 예시

Cluster S/W 특징

Cubedoop 최적화 파라미터 가이드를 통해 각 클러스터 타입에 맞는 Cubedoop Cluster가 세팅이 되며, 일정 주기 별로 Cubedoop 설정 파라미터에 대한 평가가 클러스터 사용자에게 레포팅되어 분산작업 시간을 최소화 시켜준다.

또한 다른 Hadoop 배포판과 마찬가지로 노드들을 효율적으로 관리 할 수 있는 모니터링 Tool이 별도로 제공이 되며, 분산 Job 처리시에 각 노드로 분산되는 로그를 일괄적으로 편하게 확인 할 수 있는 등 Job 처리 편의성을 높였다.

Cluster H/W 특징

클러스터를 구성하는 노드들의 하드웨어 관리를 편하고 효율적으로 하기 위해서 메인보드와 하드 디스크 그리고 파워서플라이를 분리 시켜 그룹화 했으며, 각 노드들의 상태(Status)와 전력 상태등을 전담 모니터링하는 서버를 각 Rack 마다 하나씩 추가했다. 그 결과 고장난 노드 교체 비용을 최소화 했으며, 복잡한 클러스터 운영을 많은 부분 단순화 시켰다.

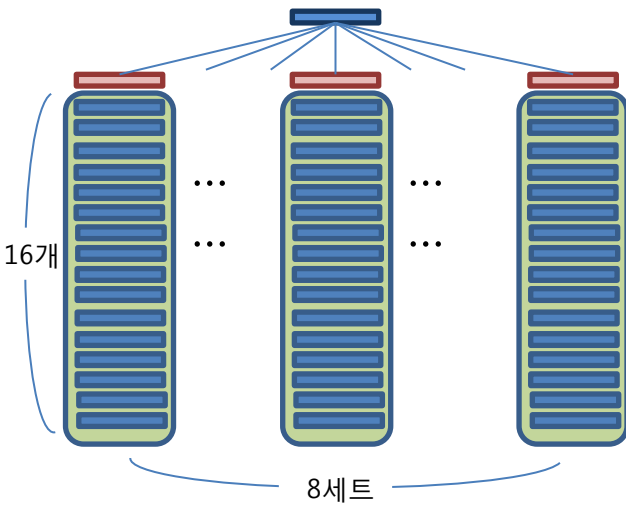


표준 이하 사이즈 Rack 3대 예시

Cluster 종류(노드수)

- Type - A : $16 * 8 = 128$ nodes
- Type - B : $16 * 16 = 256$ nodes
- Type - C : $16 * 24 = 384$ nodes
- Type - D : $24 * 24 = 576$ nodes

Type-A 구조



Cluster 종류(WorkLoad)

표준형 : 실행 잡이 CPU 계산 혹은 파일 I/O 작업 어느 한쪽으로 편중 되지 않고 분산처리 작업의 유형이 다양한 경우 선택

CPU 편중형 : 데이터 마이닝 혹은 파생 데이터를 이용한 계산이 많은 경우 선택(예: 자연어처리, HPPC)

I/O 편중형 : 파일 입/출력이 많은 작업의 비중이 높은 경우 선택
(예: 전형적인 MapReduce job, Sorting)

Appliance + 지원 서비스

거대 규모의 빅데이터 클러스터를 안정적으로 운영하기 위해서는 클러스터 전문가의 기술이 절대적으로 필요하다.

일반적으로 클러스터 운영의 노하우는 오랜 기간에 걸쳐 쌓이기 때문에 빅데이터 클러스터 전문가를 보유하기가 쉽지 않은 고객사는 빅데이터 클러스터 도입을 망설일 수 밖에 없다.

당사는 고객사에 대한 빅데이터 클러스터 전문가의 지속적인 지원 계획을 통해, 빅데이터 클러스터 어플라이언스를 도입한 고객이 안정적인 클러스터 운영을 보장 받을 수 있도록 최대한 노력할 것이다.

Service Layer



For more detail

daengky@bicube.co.kr
godwithks@bicube.co.kr

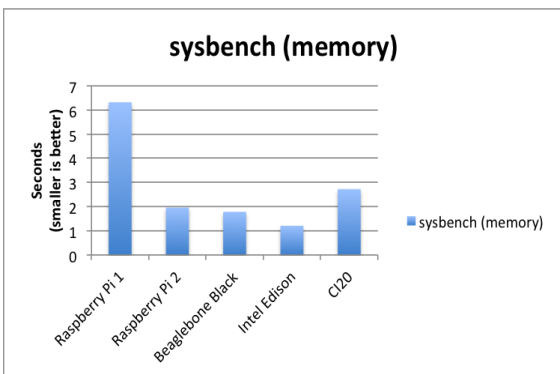
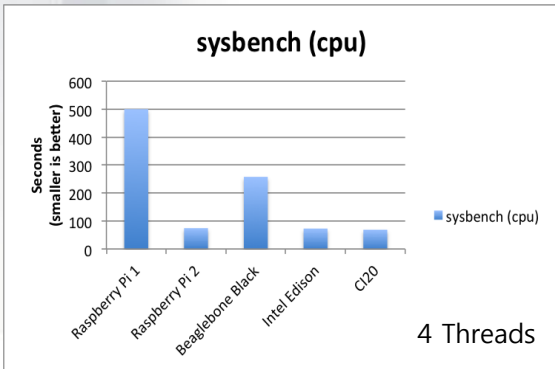
CubePi Cluster™

CubePi Cluster는 분산처리의 성능을 벤치마킹하기 위해 BICube가 디자인한 Raspberry Pi 2 Cluster에서 시작되어 상품화 되었다.

Raspberry Pi 2 각각의 성능은 PC와 비교가 안될 정도로 미약하지만 다수의 Raspberry Pi 2가 클러스터링 됐을 때 분산처리의 성능은 엔터프라이즈 용도로도 손색이 없을 정도로 훌륭하다.

BICube는 **CubePi Cluster™**의 잠재적 고객 니즈를 예상해 제품화했다. (**HPCC** (High-Performance Computing Cluster), or **DAS** (Data Analytics Supercomputer))

Raspberry Pi 2 성능



** Referred from <http://www.davidhunt.ie>

Raspberry Pi 2 Spec

	Pi1 B+	Pi 2B
CPU	Arm11	CortexA7
Cores	1	4
Clock	700MHz	900MHz
GPU	Videocore IV	Videocore IV
Memory	512MB	1G
USB Ports	4	4
Storage	microSD	microSD
Network	10/100	10/100
GPIO	40-pin	40-pin

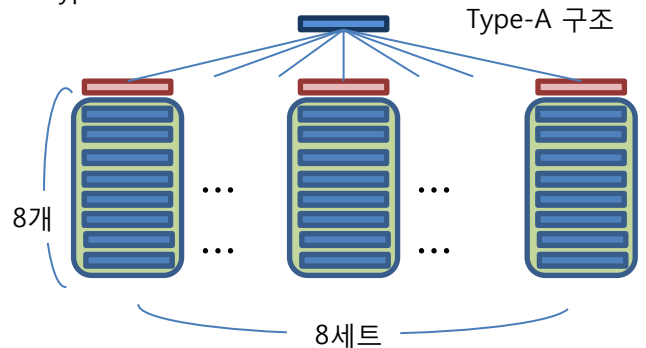
클러스터 종류

Type - A : 8 * 1 = 8 nodes

Type - B : 8 * 8 = 64 nodes

Type - C : 8 * 16 = 128 nodes

Type - D : 8 * 24 = 192 nodes



CubePi Cluster™ 용도

A-Type과 B-Type의 **CubePi Cluster™**는 교육용 hadoop cluster로 사용하기 적합하다. 하둡 클러스터를 학습하는 거의 대부분의 빅데이터 인프라 인력들은 64대가 넘어가는 거대 규모의 하둡 클러스터를 접할 수 있는 기회가 전무하다. C-Type 이상의 **CubePi Cluster™**는 HPCC용 클러스터로 사용될 수 있다.