

大数据集群概 观

Cubedoop Cluster™

CubePi Cluster™

Cubedoop Cluster™

BI**Cube**™ 作为一个持有大数据机器学习平台的公司，在大数据机器学习平台储存数据的同时，必需运行 **MapRedue**。在此，BI**Cube**™，为了平台内的稳定性和性能的效率性，利用**Apache Hadoop**制造了大数据集群平台，**Cubedoop Cluster™**。并且通过自制研发大数据集群积累的经验，实现了 **Cubedoop Cluster™** 的商品化。

Cluster Appliance

Cubedoop Cluster™ 是一个以Rack为扩张单位的 S/W & H/W 一体化Cluster Appliance。

Cluster专家，针对顾客的解决方案，确定Cluster大小。在以最佳性能的H/W Spec组成的 Cluster之上，安装已设置完最佳Parameter的Cubedoop。随后，将这款 **Cubedoop Cluster™** 提供给客户



图1. 3台标准大小的 Rack

Cluster S/W 特点

通过Cubedoop 最佳Parameter的导航，设定符合每个Cluster类型的Cubedoop Cluster，针对Cubedoop设定Parameter的评价，会以一定的周期 为单位，发送到Cluster使用者手里，实现分散作业时间的最小化。

另外，将提供 像Hadoop 配布版一样，能有效管理Node的监视工具 (Monitoring Tool)。在处理分散作业时，通过此工具，使用者可以很容易观察到从各个Node分散出的统一记录。如此，大大提高了作业处理的简便性。

Cluster H/W 特点

为了有效而简便的管理Node(构成Cluster)的H/W，分离了Main Board, Hard disk 和 Power Supply的同时，建立了集体化。而且在每个Rack增加了一个可专门监视每个Node的状态(Status)和电力状态等的服务器。如此，实现了更换故障Node费用的最小化，而且大部分简化了复杂的Cluster的运营。



图2. 3台标准大小以下的 Rack

Cluster 种类 (Node数)

Type - A : $16 * 8 = 128$ nodes

Type - B : $16 * 16 = 256$ nodes

Type - C : $16 * 24 = 384$ nodes

Type - D : $24 * 24 = 576$ nodes

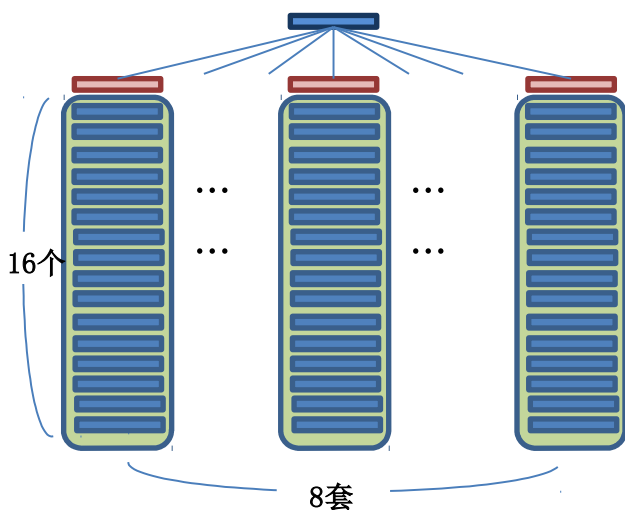


图3. Type-A 结构

Cluster 种类 (Work Load)

标准型 : 实行作业, 不偏向于任何CPU 计算或文件 I/O作业, 适用于多种分散处理作业。

CPU 偏重型 : 适用于利用数据挖掘技术或派生数据的计算 (例如: 自然语言处理, HPPC)

I/O 偏重型: 适用于高频度文件输入/输出作业 (例如: 典型的 MapReduce Job, Sorting)

Appliance + 援助服务

为了稳定的运营巨大规模的大数据集群 (Cluster), 迫切的需要Cluster专家的技术。

总所周知, Cluster运营经验是需要长时间的积累。因而, 无法拥有大数据集群专家的客户, 很自然的对引进大数据集群, 有所顾忌。

敝司对客户的大数据集群专家, 将提供可持续援助计划, 以及确保客户引进的大数据集群 Appliance的稳定运行, 做出最大的努力。

Service Layer

稳定的集群 (Cluster) 运营

障碍措施

Audit Report

集群监视

定期检查

集群维持补修

For more detail

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

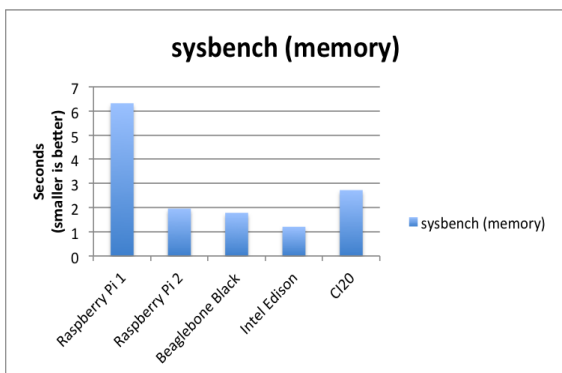
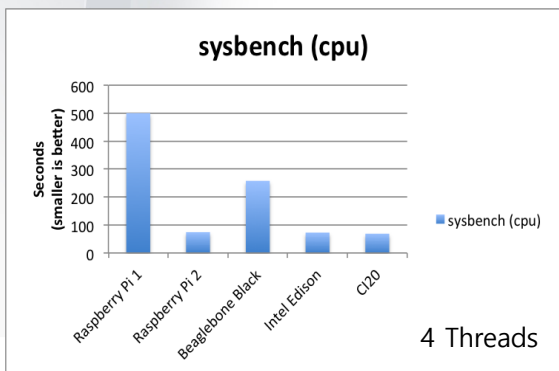
CubePi Cluster™

CubePi Cluster 是为了分散处理性能的 Bench Marking. 从 BICube™ 设计的 Raspberry Pi 2 Cluster 发展并实现了商品化。

虽然 Raspberry Pi 2 各个性能无法与 PC 相比, 但聚集了多数的 Raspberry Pi 2 的时候, 分散处理性能的优越性一点都不逊色于 Enterprise 用途。

BICube™ 预测了 CubePi Cluster™ 的潜在性客户需求, 完成了产品化。(HPCC (High-EnterPerformance Computing Cluster), or DAS (Data Analytics Supercomputer))

Raspberry Pi 2 性能



** 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

Type - A : $8 * 8 = 64$ nodes

Type - B : $8 * 16 = 128$ nodes

Type - C : $8 * 24 = 192$ nodes

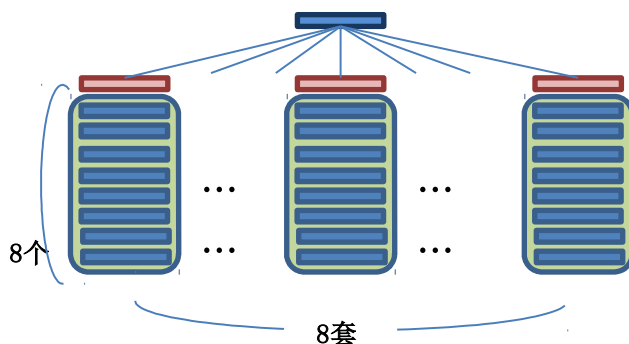


图4. Type-A 结构

CubePi Cluster™ 应用

属于A-Type的 CubePi Cluster™是适用于教育型 Hadoop Cluster。学习 Hadoop Cluster 的大部分大数据基础人力, 毫无机会接触到超过64台的巨大规模的 Hadoop Cluster。属于C-Type 以上的 CubePi Cluster™是适用于 HPCC型 Cluster。