BeeGFS

BeeGFS是基于轻量级的架构并且可以在每个作业的基础上创建BeeOND（BeeGF On Demand）应用程序。 BeeOND旨在给跨计算节点的任务提供一个BeeOND性能加速实例。BeeOND聚合计算节点内部的性能和SSD容量，并在计算任务计算期间以Burst Buffer的方式持续增加性能，这在云环境或者scratch数据场景是非常有用的。配额支持

介绍

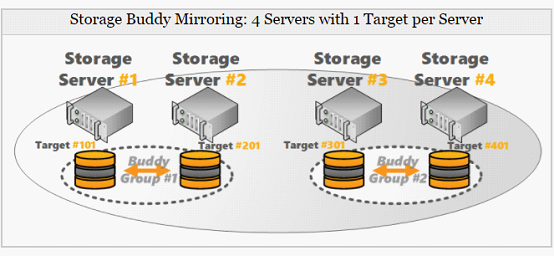
BeeGFS配额支持基于用户和组的配额，通过两个步骤实现。第一步查询用户和组在BeeGFS上使用的数据量和块文件数,第二步实施配额，允许定义用户和组的配额限制。目前只支持基础zfs使用的空间配额，而不支持用户和组使用文件数配额。  
  
BeeGFS配额系统需要使用存储服务器Target目标上的底层本地文件系统(例如xfs 和 ext4)的配额信息。因此，存储Target目标的底层文件系统也需要支持配额跟踪。在元数据Target目标上不需要配额支持。当启用配额执行时，BeeGFS管理守护程序会定期从BeeGFS存储Target目标中收集配额信息，检查是否超出的配额限制。

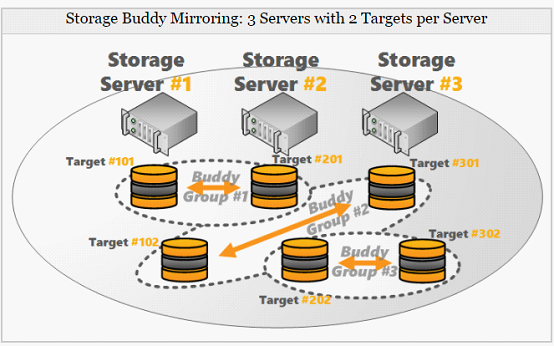
项目目录配额跟踪

BeeGFS的内置配额功能跟踪和执行基于用户和组配额。如果要跟踪使用的空间或基于子目录树的文件数量，您可能需要查看Robinhood策略引擎，Robinhood可以定期对文件系统进行并行扫描，并将发现的文件和目录信息存储在SQL数据库中。一方面采用数据库的各种查询可以快速检索结果。另一方面可以在Robinhood中定义某些事件的自动操作。  
  
由于BeeGFS在元数据服务器（通常为闪存存储）上保留所有扫描元数据，为了确保Robinhood的SQL数据库不会降低扫描速度，建议将Robinhood数据库还存储在闪存上。

**BeeGFS Buddy 镜像**

从2012.10发行版开始，BeeGFS提供对元数据和文件镜像的支持。镜像功能集成到BeeGFS服务中，因此该功能不需要单独的服务或第三方工具支持。元数据镜像和文件内容镜像可以彼此独立使用。  
  
从BeeGFS v6.0开始，元数据镜像也以高可用性功能扩展，高可用性和据镜像基于所谓的Buddy组实现。一般来说，Buddy组是一对内部管理数据副本的两个Target目标。



存储服务器Buddy Mirroring也可以与奇数个存储服务器一起使用(偶数当时是可以的)，这是因为BeeGFS的Buddy组由独立的存储Target目标组成，独立于服务器的，如以下示例图所示， 采用3个服务器和2个Target存储目标，存储Buddy组由不同服务器提供的Target目标组成，甚至可以由连接到同一服务器的2个Target目标组成。  


需要注意的是元数据服务器要求是偶数，在正常操作中，Buddy组中的一个Target存储目标（或元数据服务器）被认为是primary 的，而另一个是secondary的。更新操作将始终发送到primary 服务器负责镜像过程。文件和元数据采用同步镜像方式完成，数据在两个Target存储目前完成后才返回客户端。  
  
如果Buddy组的primary 存储目标或primary 元数据服务器无法访问，它将被标记为offline ，并切换到secondary上。 Secondary将变成新的primary，故障透明切换，改变的数据将同步到所有节点以证系统的一致性。

启用和禁用镜像

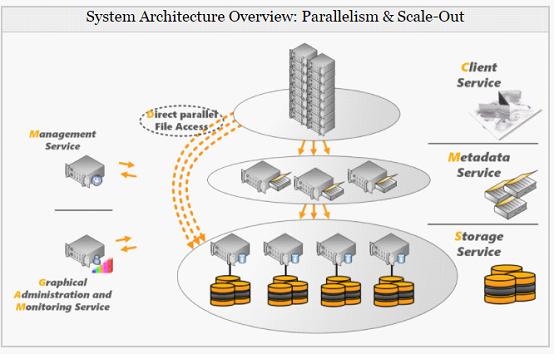
默认情况下，新文件系统实例的镜像功能是被禁用。可以使用命令行工具启用镜像功能。在启用元数据或存储镜像之前，需要定义Buddy组，因为Buddy组是镜像的基础。

**BeeGFS API概述**

除了POSIX接口外，BeeGFS还提供了其他API来更好地控制数据存放位置，查询附加信息或通过其他接口访问数据。  
  
1）条带化API允许应用程序开发人员创建具有单独条带模式的文件，这些文件将针对特定文件的访问模式进行调整或调优。  
  
2）缓存API提供了在BeeGFS缓存文件系统（通常为Beeond实例）和全局BeeGFS之间复制数据的功能。从缓存文件系统预取和刷新数据支持同步和异步方式。  
  
3）Hadoop BeeGFS连接器API为BeeGFS提供了Hadoop文件系统（HDFS）的支持功能，使得Hadoop应用程序开发人员可以使用通用Hadoop API来访问BeeGFS。

BeeGFS的最低系统要求

目前，本地BeeGFS client和服务器组件适用于x86，x86\_64和PowerPC/Cell架构上的Linux操作系统。一般来说，所有BeeGFS组件都可以在单个机器上运行，甚至只有单个CPU内核，单个硬盘和RAM小于1GB规格。但实际运用中这可能不是你想做的，所以下面是一些关于硬件配置的建议：

****

**存储服务器(Storage Server)：**

* 存储服务器应具有至少4GB的RAM以具有足够的可用于缓存和客户端连接缓冲区的内存。根据磁盘的实际类型和文件系统客户端的数量，每个附加磁盘的1GB RAM通常是一个不错的选择。
* 如果您使用传统的SATA硬盘（即不是固态硬盘或快速SAS驱动器），则通常需要使用每个服务器至少有8个硬盘的RAID。
* 即使存储服务器任务通常不是非常CPU密集型，Linux缓冲区高速缓存可以通过快速RAID阵列产生高CPU负载，因此在这里使用快速CPU可以实现高吞吐量，特别是在高速网络上通常是件好事。
* 注意磁盘数量（整个系统和每个服务器），网络互连带宽和文件系统客户端数量之间的平衡。（例如，根据您的用例，如果服务器仅配备10Gb互连，则每个服务器使用超过12个硬盘驱动器可能无意义。）
* **元数据服务器(Metadata Server)**
* 如果您正在主要使用大型文件，那么元数据服务器的工作不会太多，您可能只想在与存储服务器和同一硬盘分区上运行的相同的计算机上运行它。
* 在具有非常不同类型工作负载的群集上，您可能仍然希望在存储服务器计算机上运行元数据服务器守护程序（拥有大量元数据服务器，而不需要为专用计算机花费额外的资金），但是您应该使用专用磁盘您的存储服务器用于元数据。  
  由于元数据服务器正在进行大量的小型磁盘随机访问（例如，它们正在读取和写入大量小文件），因此如果两者位于同一磁盘上，则可能会对存储和元数据访问产生显着的性能影响。我们还经常看到RAID控制器在为元数据和存储服务器管理不同的RAID卷类型时会遇到（性能）问题，因此您可能希望在单独的RAID控制器上拥有元数据。
* 如果您有大量客户端处理许多（小）文件，则元数据服务器上的CPU使用率可能会很高。因此，确保使用快速的CPU和快速驱动器（通常为SSD）来保证元数据的低访问延迟。
* 与总存储容量相比，元数据所需的空间量取决于实际使用情况（即要存储的文件总数）。对于临时文件系统，元数据所需的空间通常约为总存储容量的0.3％至0.5％，但有时这也是太多。有关详细信息，请参阅此处。
* **客户端(Client)：**
* 无补丁客户端内核模块需要Linux内核2.6.18或更高版本。客户端不需要特定的Linux发行版，即使是最近的香草内核也可以自由使用。由于BeeGFS的客户端组件设计非常轻便，因此没有特殊的RAM或CPU要求。
* **管理守护进程(Management)：**
* 管理守护程序仅使用最少的CPU周期和内存。访问管理守护程序与文件系统性能无关。只需在任何存储/元数据服务器或群集主服务器上运行此守护程序。
* **Admon守护进程(Admon)：**
* admon守护程序经常存储和查询其内部数据库中的大量值，因此它将运行一些存储事务并使用适量的RAM和CPU周期。在任何其他存储/元数据服务器计算机上运行它通常很好，但是您的集群主机可能是此守护程序更好的位置，以确保对文件系统性能没有影响。

BeeGFS支持的网络类型

BeeGFS支持TCP/IP网络、InfiniBand协议（基于OFED ibverbs框架）、Omni-Path和RDMA over Converged Ethernet（RoCE）协议。存储服务器和客户端**Client**可以同时支持混合网络请求（例如，存储服务器配备Infiniband和以太网互连，一些客户端**Client**通过本地InfiniBand连接，而其他客户端**Client**通过TCP /IP的 Ethernet进行连接）。如果主连接路径发生故障，具有多个连接路径（如InfiniBand和以太网或多个以太网端口）的客户端也可以接管实现网络故障转移。

通过SMB/CIFS导出BeeGFS

在BeeGFS中可以导出成SMB或CIFS标准NAS文件系统。将BeeGFS客户端主机配置为Samba服务器。通过BeeGFS配置文件/etc/samba/smb.conf为BeeGFS配置权限和挂载点。重新启动samba服务器服务。现在，您应该可以从windows Samba客户端访问BeeGFS文件系统。

可以通过NFS导出BeeGFS吗？

在BeeGFS中可以导出成NFS标准NAS文件系统。从2014.01发行版开始，BeeGFS支持Linux内核NFS服务器导出。通过NFS导出BeeGFS ，通常它比Samba(SMB)具有更好的性能。  
为了通过NFS提供文件，请将BeeGFS客户端主机配置为NFS服务器，并通过NFS重新导出BeeGFS客户端挂载点。