BeeGFS Parallel File System

Overview

BeeGFS is an open source parallel file system that is suitable for large data storage and environments that require high scalability. Originally known as the Fraunhofer Gesellschaft File System (FhGFS), BeeGFS was designed as a storage solution for high performance computing by the Fraunhofer Institute for Industrial Mathematics in Germany.

As open source software, BeeGFS can be acquired for free. The BeeGFS client is available under the GPLv2 license and the server components are published under the BeeGFS EULA. A support contract for BeeGFS can also be purchased from ThinkParQ which provides support for the software as well as enables additional features for enterprise implementations such as high availability, quota enforcement, and access control lists.

The BeeGFS parallel file system is comprised of four main parts: the management service, the metadata service, the storage service, and the client service.

The management service is a lightweight service that keeps track of the other services and their states. The management service does not store any user data, and typically does not require a dedicated machine.

The metadata service is responsible for storing metadata about the data that is being stored such as directory and ownership information. Data in BeeGFS is stored in chunks using striping that is determined by the metadata service. The metadata service is responsible for providing the stripe pattern to the client, but otherwise is not involved in data access to improve efficiency. The stored metadata is intended to be small and to scale linearly with stored files.

Highlights

- Scale-Out Parallel File System
- Open source software available for free or for purchase
- Buddy Monitoring synchronous replication
- Beegfs-ctl command line and optional admon gui for administration and monitoring
- BeeOND on demand temporary file storage
- POSIX client, NFS and CIFS export
- Metadata separated from data files
- RAID6 or RAIDz2
- TCP/IP, Infiniband, Omni-Path, and RoCE networking

The storage service is the service that is responsible for storing striped user files, referred to as data chunk files, across one or many storage targets. BeeGFS utilizes a scale out design that allows for both capacity and performance to increase as needed. The storage service works with any local Linux POSIX file system. Storage targets typically consist of hardware RAID6 or zfs RAIDz2 volumes.

The final service of BeeGFS, the client service, is responsible for mounting the file system to access stored data. When the client kernel is built, it automatically matches the current running Linux kernel. This eliminates the need for manual updates when the Linux kernel is updated. BeeGFS can also be accessed via NFS, CIFS, or Hadoop. It has been stated that future versions of BeeGFS will include a native client for Windows as well.

BeeGFS also provides advanced features such as Buddy Monitoring synchronous replication to mitigate the risk of hardware failures, storage pools to optimize efficiency of storage resources, and BeeOND (BeeGFS On Demand) to enable the use the internal flash drives of compute nodes as temporary file systems. BeeGFS also features an optional graphical interface for administration and management called "admon".

BeeGFS has additionally added index functionality with BeeGFS Hive Index. Hive Index stores metadata and allows users to run searches or queries on the filesystem with greater efficiency and without impacting the performance of the actual filesystem.

Usage

BeeGFS is intended for large data workloads that need to scale easily without sacrificing performance for added capacity. High performance computing, artificial intelligence, life sciences, oil and gas, and finance are all areas where BeeGFS may be an effective solution.

BeeGFS software is not dependent on specific hardware and can support Intel, AMD, ARM, and OpenPOWER. BeeGFS can store data in any Linux POSIX file system including ext4, xfs, and zfs.

- Characteristics
  - Performance – Supports SSDs for increased performance. Performance scales out with capacity.
  - Availability – High availability of data and metadata achieved through Buddy Mirroring using synchronous replication between storage targets in "buddy" pairs. Storage service is typically configured with RAID6.
  - Replication for BC/DR – Synchronous replication of storage target pairs in different hardware failure domains is supported through Buddy Mirroring.
- Applications
  - BeeGFS is targeted at large file based applications such as high performance computing. The system is capable of scaling out to add both performance and capacity as needed.
- System environments
  - BeeGFS is usually only seen in Linux environments.
- Deployment and Administration

- Four main components: Management server, metadata server, storage server, and client server.
- Administration achieved through beegfs-ctl command line or optional admon graphical interface.

Evaluator Group EvaluScale: BeeGFS - Scale-Out File System Usage

Evaluator Group product review methodology "EvaluScale" assesses each product within a specific technology area. The definitions of the criteria and explanations of how products are reviewed can be found in the **Evaluation Guides**.

	Criteria	Description	Requirement	EG View of BeeGFS	Explanation for BeeGFS
1	Capacity	Current capacity of system to meet demand Number of file systems & Maximum number of files	Must have enough capacity to meet current demand and have ability to scale-up – adding more capacity up to a practical limit. The number of files systems, the maximum size and number of files may be critical in many environments.	Exceeds requirements	No defined limits. Capacity can scale out as needed for large data environments.
2	Price – including data reduction	Cost of system. This includes data reduction effect – compression/deduplication or single instancing	Must be competitive with other leading solutions in this space meaning prices have no more than 20% variance from an average of the other solutions. This includes the effect of data reduction according to the Evaluator Group Data Reduction Estimator tool.	Exceeds requirements	BeeGFS can be used for free under open source licensing. A support contract is available for purchase through ThinkParQ which may be more suitable to enterprise deployments.
3	Performance	Bandwidth SPECsfs file operations per second. Automatic load and capacity balancing TCPIP accelerator	The performance requirement can vary based on the number of nodes and scale for a parallel file system. Balancing I/O for performance and capacity should be a requirement for each product.	Meets requirements	Massive parallelism yields very high bandwidth. Performance can scale out to match growing capacity.
4	Connectivity	Number of ports for access Type of port – 10GigE, 40GigE	The number and type should meet current and planned infrastructure requirements.	Meets requirements	Supports TCP/IP, Infiniband, Omni-Path, and RoCE. Can scale as needed for ports.

5	Scaling – performance and capacity	Ability to increase to meet future demands Maximum number of nodes supported	Scale-out means scaling both performance and capacity to meet demands up to a practical limit – more capacity without sacrificing usually with addition of nodes to increase capacity and performance.	Exceeds requirements	Large scale-out capability enables increase in both performance and capacity.
6	Protocols and sharing support	NFS, CIFS/SMB, HTTP, FTP, file sharing type between protocols	Protocols should include a custom POSIX client and both NFS and CIFS/SMB – with a native SMB3 implementation. File sharing between protocols is a basic requirement.	Meets requirements	POSIX Client, NFS and SAMBA export (limited). Does support GPUDirect.
7	Security	LDAP, AD, NIS. File level locks, TLS, IPSEC Data at rest encryption and key management	Both LDAP and AD are basic requirements. File level locking is a security and integrity issue and may be implemented with a Distributed Lock Manager. Encryption and external key management may be a requirement in some environments. This also includes encrypted communications.	Area for development	Lacking in security features. No encryption at rest.
8	Data protection	Snapshots – file based or filesystem Asynchronous replication NDMP support	Read/write snapshots are a requirement with a number that roughly equals the number of filesystems supported. An added bonus is the ability to snapshot individual files. The high-end enterprise requirement for remote replication is for both synchronous and asynchronous technology while mid-tier and entry usage require asynchronous. NDMP for data protection is a basic requirement. Snapshot or tiering to cloud/object storage may be a benefit to all segments but would only be a requirement in the high-end enterprise.	Meets requirements	RAID6 or RAIDz2. Synchronous replication via Buddy Mirroring. No support for asynchronous replication or snapshots.

9	Economic considerations	Warranty Evergreen updating Environmentals – power & space Simplicity for administration	The overall environmental footprint being roughly on par with other leading systems in this area is the requirement measure. An extended warranty period for devices and an evergreen program for the controllers in the case of an all- flash system is now a requirement given the competitive nature. Simplicity for administration is a basic requirement.	Area for development	No evergreen program. Administration has potential for complexity without access to official support.
10	Advanced features	Automatic migration to clouds / file shares with stubbing and recall S3 object API Multi-tenancy isolation Regulatory compliance features	File tiering to another mount point or clouds/S3 is a competitive issue and a growing requirement but not currently required. Support as a target for objects using S3 is not a requirement as object storage provides a better solution. Multi-tenancy isolation is a requirement in large NAS system environments and for consolidation in enterprises but not usually in the mid-tier or entry. Regulatory compliance features is specific to certain environments and not a general requirement.	Area for development	BeeOND and Hive Index are unique features. Some features not available in free version. No cloud tiering.

Evaluator Group Opinion: Differentiating elements for BeeGFS

BeeGFS was designed as a solution for high performance computing environments and has positioned itself as a viable alternative to other scale out file system solutions such as Lustre or IBM Spectrum Scale. The parallel file system excels with I/O intensive workloads and is able to scale out both capacity and performance as needed.

BeeGFS, which can be used for free under open source licensing, can be a cost-effective way to implement a high-performance scale out file system. For some large enterprises, however, purchasing the software through a support contract with ThinkParQ may be a more practical solution as it adds some more robust features as well as official support to reduce potential complexity associated with deployment, administration, and updates.

BeeGFS offers support for useful features such as Buddy Mirroring, BeeOND, Hive Index and the admon gui, however, it is missing advanced features and security features that may be found in other products. Additionally, the BeeGFS client is currently limited to Linux systems until a native Windows client is developed.

Evaluator Group believes that BeeGFS can be a useful solution for high performance computing workloads that can be acquired at a low cost while allowing hardware flexibility.

Information that is more detailed is available at **http://evaluatorgroup.com**