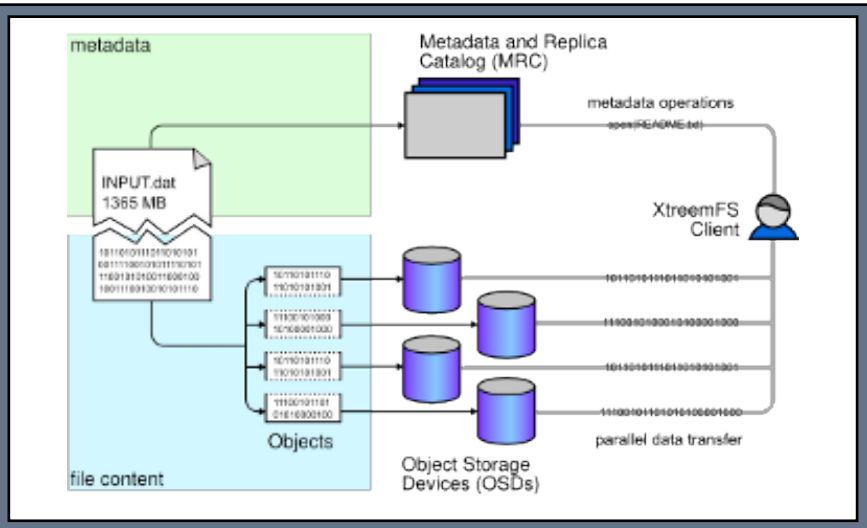

XtreemFS architecture

A fully-featured Cloud file system

XtreemFS is a secure and fault-tolerant file system for the cloud. It has been designed to store peta-scale data volumes across large numbers of distributed servers, while it behaves like a local file system with POSIX® semantics from a user's point of view.

XtreemFS has been specifically tailored to the use in cloud computing environments. The modular design makes it possible to start with a small-scale installation and scale it out when necessary. Its integrated security infrastructure allows to share storage resources between multiple users in a secure and isolated manner. Replication ensures availability and safety of all data at any time.

Examples of **use cases** for XtreemFS are:

Efficient Storage and Distribution of VM Images. XtreemFS provides the scalability to accommodate vast amounts of large virtual machine images for a cloud. Immutable VM images can be replicated to multiple sites in a particularly efficient manner, which allows different clients to access different replicas in parallel, thus balancing load and reducing startup times of distributed cloud applications. Also images with copy-on-write (CoW) format, which have a static part are supported.

Storage of Cloud User Data. There are many cloud storage solutions - but only few provide POSIX file system semantics! Ideally, users can run their applications on a cloud without major modifications. XtreemFS provides the same guarantees as a local file system when files are accessed, even if accesses are directed to different replicas. Furthermore, XtreemFS can use a shared pool of storage resources for the data of many different users while protecting the interests of individual users in terms of privacy and isolation.

XTREEMFS

XtreemFS

A Cloud file system

Contrail File System

contrail
open computing infrastructures for elastic services

ZIB

http://xtreemfs.eu

File replication

One of the core features of XtreemFS is its ability to replicate files over WANs while guaranteeing their consistency. Replication algorithms are designed for dealing with large latencies and complex failure scenarios like network splits. Storage servers are aware of the state and recency of the files they store. XtreemFS offers different kinds of replication:

Read-only File Replication. Read-only file replication provides an efficient mechanism to distribute large volumes of write-once data in a CDN-like manner.

Data transfers are performed asynchronously between all replicas using peer-to-peer mechanisms. This allows replicas to be accessed while they are created, and allows to prioritize the transfer of important parts of the file to the new replica.

Read-Write File Replication. Read-write replication ensures strong replica consistency without write-once restrictions. Internally, it resorts to a primary backup scheme with automatic primary fail-over. Read-write replication ensures that files can be read and written as long as a majority of all replicas remain available.

Metadata Replication. Since metadata and file content are managed separately in XtreemFS, metadata is replicated independently. Metadata replication ensures safety and high availability of metadata in the event of downtimes of metadata servers.

Replica Placement and Selection. The placement and selection of file replicas in XtreemFS is policy-driven. Users can define policies that restrict the creation of replicas to certain servers. Similarly, policies define the preferred order in which clients attempt to access existing replicas.

http://contrail-project.eu

Cloud ready features

POSIX® compatibility. XtreemFS provides the same interface and operation semantics as a common local Linux file system. Applications can thus use XtreemFS without having to be adapted to a specific storage subsystem.

Elasticity. Servers can be easily and dynamically added to an XtreemFS installation in order to increase storage and I/O capacity of the file system. This can happen at any time without having maintenance downtimes. Newly added servers are immediately integrated in the system.

Data Safety. XtreemFS provides robustness in the event of storage device failures by means of replication. Maintaining multiple replicas of files and metadata ensures data safety even if underlying storage devices take physical damage.

High Availability. In peta-scale storage installations, hardware failures and downtimes are the norm rather than the exception. XtreemFS transparently resorts to available replicas of files and metadata if individual servers become unavailable. XtreemFS supports off-site replication over wide area networks to ensure availability even in the event of downtimes of entire data centers.

Security. XtreemFS comes with an integrated security infrastructure that protects data from unauthorized access. SSL connections ensure that data transfers between clients and servers are encrypted. X.509 certificates enable a secure authentication of individual users. POSIX® permissions and ACLs provide the basis for a fine-grained access control to different data volumes.

Extensibility. Most behavior in XtreemFS can be controlled by means of policies. Examples are authentication and authorization of users, placement of files and replicas, and selection of replicas. In addition to using predefined policies, XtreemFS offers a plug-in mechanism to support custom user-defined policies.

Latest release

XtreemFS release 1.4

XtreemFS is available for Linux (openSUSE, SLE, Fedora, CentOS, RHEL, Mandriva, Debian, Ubuntu, Gentoo), Mac OS X and Windows.

XtreemFS offers a range of **additional features:**

Striping. XtreemFS can spread chunks of a single file across multiple storage servers, thus increasing throughput when accessing large files.

Client-side Metadata Caching. XtreemFS clients can maintain a local metadata cache to ensure low-latency access to metadata.

Snapshots. XtreemFS can record consistent snapshots of volumes.

Checksums. Storage servers are capable of calculating and verifying checksums whenever data is read or written, so as to detect corruptions of file content.

Hadoop Support. XtreemFS can be accessed by Apache Hadoop applications through an HDFS adapter.

Monitoring. XtreemFS installations can be easily monitored with third-party monitoring tools like Gangalia and Nagios through an SNMP-based monitoring service.

XtreemFS is partially developed in the Contrail project. Contrail is a project managed by the Contrail consortium. It develops a stack of federated Cloud computing tools that can work together.

Contrail is partially funded by the FP7 Programme of the European Commission under Grant Agreement FP7-ICT-257438.