EOS: An Extensible Object Store

Alexandros Biliris and Euthimios Panagos
AT&T Bell Laboratories
600 Mountain Ave., Murray Hill, NJ 07974
{biliris, thimios}@research.att.com

EOS [1] is a storage manager that has been prototyped at AT&T Bell Laboratories as a vehicle for research into distributed storage architectures for database systems and specially those that integrate programming languages and databases. EOS's overall goal is to provide fast and transparent access to persistent objects independent of their size and their physical location in a distributed computing environment based on a client-server architecture.

EOS objects are uninterpreted byte strings which can range in size from a few bytes to gigabytes. Large objects, spanning multiple pages, can be accessed and updated transparently as if they were small objects, or via byte range operations. The byte range operations are important for very large objects – such as digital video and audio – because there may be memory size constraints that would make it impractical to build, retrieve or update the whole object in one big step. EOS files collect related object together and are stored in EOS databases. EOS databases are stored in one or more storage areas (UNIX files or raw disk partitions). Clustering hints for the physical placement of objects in pages, files, databases and areas are also provided. Any EOS object can be named and subsequently retrieved by its name.

EOS offers extensible hashing supporting variable size keys and user-defined hash and comparison functions. In addition, other index structures can be built by using page objects – objects that expand over the entire available space of a page.

EOS employs the multigranularity two version two phase locking protocol, that allows many readers and one writer to access the same item simultaneously. The option to switch to simple 2PL is also available. EOS uses a write-ahead redo-only logging scheme that offers short logs, fast recovery from system failures, and non-blocking checkpoints. Also, configuration files are provided that can be edited by users to customize and tune EOS performance.

Finally, the EOS architecture has been designed to be extensible. Users may define hook functions to be executed when certain primitive events occur. This allows controlled access to a number of entry points in the system without compromising modularity.

Status: EOS has been implemented in C++ (a C interface is also provided). It runs on SUNs using SunOS 4.1.x, Solaris, and SGI platforms. It is the storage engine of Ode, and it has been distributed free of charge to Universities. To obtain the EOS system, please send e-mail to **eos@research.att.com**.

What is Next? Our experience in using EOS in real applications helped us identify existing system components that needed to be either re-architected or extended. Also, new functionality mainly in the area of distributed computing needed to be designed. Some features of the new system include the following.

- Inter-transaction caching and a number of operation modes for accessing data cached on the server or the client workstations (copy on access, copy on write, shared memory, and virtual-memory mode for databases smaller than virtual memory). These modes allow applications to tailor the storage system and enable users to build multiple specialized servers, e.g., multimedia servers.
- Utilization of client disks to avoid communication with servers for logging and committing transaction updates.
- Users may access and manipulate objects fetched in memory directly on the segment on which they reside without any indirection.
- Database corruption from pointer errors are prevented by storing control structures separately from data. The standard virtual memory hardware facilities are used to protect the control structures and to automatically detect updates.

Status: A first implementation of the new storage system was completed in November 1993; a release is expected by Spring 94. It will be running on the same architectures as EOS as well as on multiprocessor machines such as the NCR 3600.

References

[1] A. Biliris and E. Panagos. EOS User's Guide, Release 2.0. AT&T Bell Laboratories, May 1993.