CM-5 Scalable Disk Array

JAN 1 3 1993

Introduction

The CM-5 Scalable Disk Array is an extremely high performance, highly expandable RAID disk storage system that is completely integrated into the CM-5 parallel architecture. Capacity ranges from 9 Gigabytes to 3.2 Terabytes, with sustained transfer rates of 12 Mbytes/s to 4.2 Gbytes/s. Data is stored on individual Disk Storage Nodes that connect to the same Data Network as the CM-5 Processing Nodes. Both the capacity and the transfer rate of file structures scale independently of the number of processing nodes, resulting in a system that is uniquely configurable both in terms of processing power and storage capability.

With the introduction of the CM-5 disk array, Thinking Machines Corporation continues its leadership in supercomputer disk technology. The Data-Vault RAID storage system, introduced by Thinking Machines in 1988, established an industry first by uniting large numbers of small disk units to achieve high performance and reliability in demanding applications. Now Thinking Machines has taken this technology a step further by combining processing nodes and disk storage nodes into a peer-to-peer relationship that is tunable and scalable across an extremely broad range of application requirements.

The CM-5 disk array grows in performance and capacity with the number of Disk Storage Nodes connected to the Data Network. Systems vary in size from a few nodes to several hundred, designed from the ground up to offer very high performance, easy system maintenance, and high value.

Hardware Description

The basic Disk Storage Node is a RISC microprocessor-based controller with a CM-5 Network Interface, a large disk buffer, four advanced SCSI controllers, and eight hard disk drives. Additional custom hardware is provided to augment the transfer of data directly between the disks, the buffer, and the CM-5 Data Network. This Disk Storage Node provides up to 9.2 Gbytes of storage, a peak transfer rate of over 19 MBytes/s, and sustainable transfer rate of up to 12 MBytes/s. Hundreds of Data Storage Nodes can be accommodated within the Data Network of the CM-5.

Software Description

The CM-5 "stripes" the data across all the Disk Storage Nodes in the system, so that each node contributes to the overall transfer rate available to the user. The software offers numerous options for establishing multiple file systems across the Scalable Disk Array. High availability is assured by the inclusion of a flexible redundancy scheme. Following a hardware failure, redundant data is used to regenerate the lost data during a read.

Data stored on the Disk Storage Nodes is always available in serial order. The CM-5 provides a seamless mechanism for moving data between the Disk Storage Nodes, partitions of processing nodes, other CM-5 I/O devices, other serial computers, and I/O devices connected to them. Special-purpose hardware on the Disk Storage Node controller assists the operating system in handling data-ordering issues. This combination of hardware and software support relieves the applications programmer from the burden of dealing with data ordering.

Application programs access the file system through the standard UNIX interfaces with extensions to support large parallel transfers. The NFS protocol is implemented to allow external access to the disk array.
(over)

CM5

CM-5 Scalable Disk Array (continued)

Features

- Preserves serial ordering of data sets
- File system can be read or written from different sized partitions
- Parallel and serial reads and writes can be interspersed
- UNIX compatible, NFS mountable
- Supports multiple file systems
- Automatic on-line disk recovery

Specifications

- Disk Storage Node
  - RISC microprocessor-based Scalable Disk Array controller
  - Direct connection to the Data Network through a Network Interface (NI)
  - 8 3.5", 1.2 Gbyte disk drives
  - 1 spare drive (optional)
  - 1 parity drive (optional)
  - Up to 9.2 Gbytes storage capacity
  - Up to 12 Mbytes/s transfer rate sustained

- Disk Array Module
  - 3 Disk Storage Nodes
  - 1 Disk Array Module=1 equivalent 32 PN volume
  - 1.2 Gbyte drives
  - 22 data drives
  - 1 spare and 1 parity drive
  - 25 Gbyte capacity per module
  - 33 Mbytes/s transfer rate sustained

- Scalable Disk Array
  - Consists of 1 to 384 Disk Storage Nodes
  - Scalable capacity: 9 Gigabytes to 3.2 Terabytes
  - Scalable transfer rate: 12 Mbytes/s to 4,224 Mbytes/s sustained

- CM-5 Scalable File System (SFS)
  - 64-bit UNIX File System
  - UNIX-compatible interface from applications written in CMF, C*, or CMMD
  - Functionality: RAID 3
  - Mountable via NFS
  - Supports a flexible file system composed of a collection of blocks on a collection of disks
  - Each file system runs on 1 SPARC-based I/O Control Processor

Scalable Disk Array Configurations

Cabinet Space Required	CM-5 Address Space	SDA EPV's*	Storage Nodes	Total Number of Drives	Capacity (Gbytes)	Real Application Bandwidth (Mbytes/s)	Peak Hardware Bandwidth (Mbytes/s)
1/8	32	1	3	24	25.3	33	48
1/4	64	2	6	48	50.6	66	97
1/2	128	4	12	96	101.0	132	194
1	256	8	24	192	202.0	264	387
2	512	16	48	384	405.0	528	774
4	1,024	32	96	768	810.0	1,056	1,549
8	2,048	64	192	1,536	1,619.0	2,112	3,098
16	4,096	128	384	3,072	3,238.0	4,224	6,195

* EPV=Equivalent 32 PN volume