

IBM Enterprise Storage Server (ESS)

ESS provides:

- ✓ *Ultra High Availability*
- ✓ *Massive Scalability 420GB to 11.2TB*
- ✓ *High Performance with I/O Rates Exceeding 45,700 Per Second*
- ✓ *SAN Ready*
- ✓ *Advanced Copy Features*
- ✓ *Heterogeneous Support for NT, UNIX, and IBM eSeries Data Storage*
- ✓ *Business Continuance Options*
- ✓ *Investment Protection*
- ✓ *Flexible IBM Lease Terms*
- ✓ *Price Competitive*
- ✓ *Non Disruptive Upgrade Options*
- ✓ *RAID-5 or Non RAID*
- ✓ *Automatic I/O Load Balancing*
- ✓ *Synergy with IBM Tape Devices*
- ✓ *SCSI, Fibre Channel, ESCON, FICON*
- ✓ *FICON channel performance can run over two times faster than ESCON*
- ✓ *Full Hardware Redundancy*
- ✓ *Significant new S/390 ESS EX Performance Package Enhancements*
- ✓ *Web Management Tools*

The Storage Server Standard For The New Millennium

E-business is driving a data explosion, generating exponential growth in the need for immediately accessible, always available and highly functional storage capacity. In the increasingly e-centric world, information demand is like electricity: you plug in, you get it. Appliance-like intelligent storage served by its own network is required to provide "information as a utility."

The IBM Enterprise Storage Server (ESS) is the ultimate Storage Networking utility, providing the information "fuel" that runs the e-business "engine". Extensive heterogeneous server connectivity makes the ESS a natural fit for server consolidation requirements. The ESS supports rapid universal access to vast quantities of data through many advanced functions and features, making it a workhorse for support of business intelligence and other business-critical applications. Because of its enterprise-wide support and management scope, the ESS is tailor made to help provide consistent, efficient and effective enterprise resource planning. Using the ESS to address any or all of your strategic and tactical business initiatives will give your organization the business advantage needed to survive and thrive in the e-world.

The ESS is the third generation of the IBM Seascope (TM) architecture for disk systems. It is a solution that provides the outboard intelligence required by Storage Networking solutions, offloading key functions from host servers which frees up valuable processing power for applications. As a comprehensive SAN-based storage solution, the ESS provides considerable management flexibility to meet the fast-paced requirements of today's businesses.



Universal Data Access and Future Enhancements

ESS allows you to consolidate your data from different platforms on a single high performance, high availability storage server.

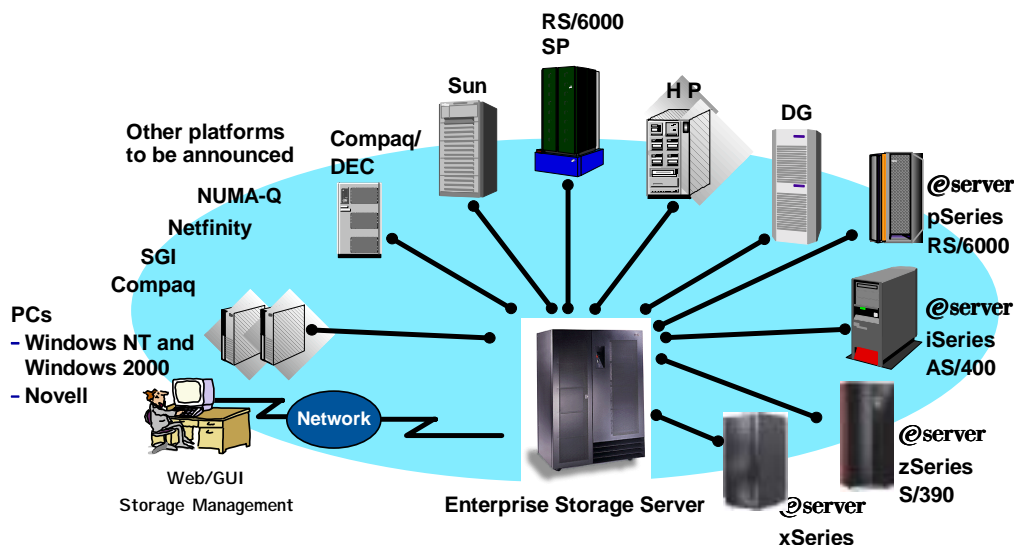
For example, 50 NT, and 5 UNIX servers could use Fibre Channel connected to the ESS.

The integrated Web based StorWatch Specialist is used to configure the ESS subsystem.

The optional StorWatch ESS Expert product provides asset, capacity and performance management functions which are especially useful for open systems. The Expert provides functions such as number of I/O, bytes transferred, read and write response times, etc.

With a capacity of up to 11TB and up to 32 direct host connections (more with daisy-chaining and switches), an ESS can meet both your high capacity requirements and your performance expectations.

Most upgrades such as additional capacity are non-disruptive.



The ESS supports many diverse platforms including the IBM's pSeries (RS/6000), many leading UNIX variants (Sun Solaris, HP-UX), IBM xSeries (NetFinity®) and other Intel-based PC servers (Compaq, etc.) running Windows NT/2000, Novell Netware, iSeries (AS/400®) running OS/400® and zSeries (System/390). In addition, the ESS supports System/390® servers running OS/390®, VM/ESA®, VSE/ESA® and TPF®. This rich support of server platforms is not limited in any way; any combination of these heterogeneous platforms may be used with the ESS. Storage capacity is partitioned among the attached servers using the flexible, web based StorWatch Enterprise Storage Specialist management tool. The ESSNet feature even provides a browser for the StorWatch interface!

Different platform types use different connection protocols, and the ESS is equipped to handle many different types concurrently including connectivity for UNIX, Intel-based PC and iSeries servers, Fiber Channel, and ESCON or FICON for System/390 servers. The ESS has up to 16 Fibre Channel or 100 MB/sec full duplex FICON ports (up to 100KM without data rate "droop") offered in pairs. Direct Fibre Channel attachment as well as SAN is available for connections to Windows NT/2000, Novell NetWare, and UNIX servers from IBM, Sun Microsystems and Hewlett-Packard.

The ESS also supports Storage Capacity Upgrade on Demand which allows you to get the capacity your e-business needs when it needs it without disrupting the flow of work. Storage Capacity Upgrade on Demand ESS, A Step Ahead, allows you to have extra disk storage capacity available when you need it. It gives you the option of receiving an extra amount of disk storage which is shipped and installed in the ESS, and paid for when used. This should be particularly attractive for customers with dynamic or unpredictable growth, or those who just want the assurance that the storage will be there when needed. Technical details about the ESS can be found at the IBM ESS web site: <http://www.storage.ibm.com/hardsoft/products/ess/ess.htm>

Business Value of the ESS

In today's information-dependent world of e-business, the ability to gather and store information from every imaginable source and deliver it to users across the extended enterprise is the key to business success. The ESS provides:

- Extensive Connectivity - The ESS provides simultaneous attachment to a wide variety of host servers with storage capacity partitioned among the attached servers using flexible web-based StorWatch management tools.
- Copy Services Functions - The ESS supports several hardware copy functions for two purposes: Mirroring for disaster recovery solutions, and copy functions that provide a near instant copy of data.
- Comprehensive Availability for Applications and e-Business 24x7x365 - Everything is designed for redundancy: power, cooling, adapters, buses and microprocessor complexes. Redundancy of components means that ESS has been designed with virtually no single points of failure as to hardware components.
- Breakthrough Performance - The ESS hardware can be configured to provide the capacity and performance required for applications supporting over 41,000 I/Os per second. New OS/390 options allow multiple simultaneous I/O operations to the same logical volume, which reduces IOSQ time.
- Low Total Cost of Ownership - Snap-in-building blocks keeps IBM development costs low -- with the savings passed on to you. IBM Global Financing can develop a lease to fit every requirement. Three-year warranty is included.
- FICON Performance - FICON System/390 channels, can reduce tape backup elapsed times by up to half, provide 4:1 reduction in ESCON channels, and can run over two times faster than ESCON channels
- Extreme Scalability - The ESS adapts to growing storage requirements created by e-business applications scaling from 420GB to over 11TB with non-disruptive upgrades. Also, storage consolidation from multiple platforms reduces costs.
- E-Business - When you're an e-business, the store is always open, so you need data storage that's available 24xforever. Your customers drive your workload, so you need high performance data storage that can meet unexpected demands. Customers can flock to your door without warning, so you need storage that is scaleable. Suddenly, everyone in the world is your user, so you need data storage on which you can bet your business!
- Storage Networking - The ESS offers customers the leading storage subsystem, supporting the key servers in the industry, and enabling customers to realize the promise of nonproprietary SANs and Network Attached Storage (NAS).
- Remote Hardware Support - The IBM hardware support center call home connection results in virtually no user intervention required on hardware incidents.

SCSI, FICON, Fibre Channel, and ESCON provide robust host server connectivity.

PPRC, XRC, Concurrent copy and FlashCopy provide data copy functions.

Maintenance and upgrades normally can be performed concurrently with full operation, including the concurrent activation of microcode.

ESS "future-proofs" users by providing a new "snap-in" design. With ESS, you can easily add performance, capacity and connectivity upgrades, providing investment protection and 24x7 productivity, saving money and time.

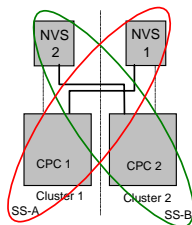
ESS has one of the smallest footprints in the industry for 11TB configurations.

Architecture Overview

ESS has full redundancy of hardware components.

Designed to have the highest quality ever for an IBM disk storage subsystem.

ESS requires two power cords for redundancy.

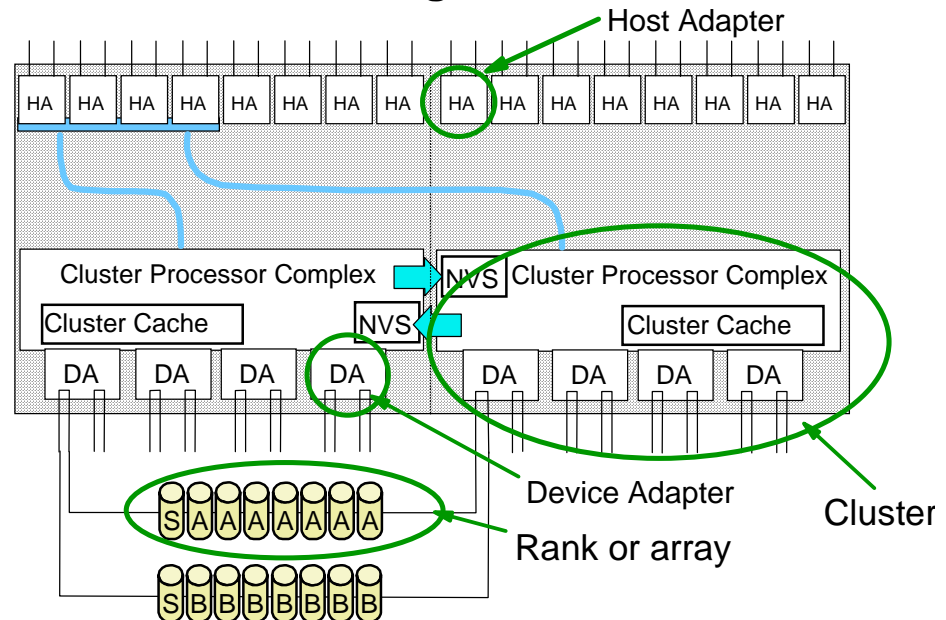


Some subsystems have only one copy of data in cache -- if the cache fails data is lost!

If a single hardware cluster fails there should be no loss of data access since the other cluster has a Disk Adapter that can access all the data. Write data is retrieved from NVS.

The schematic below illustrates the major components of the ESS. Starting at the top, there are 16 Host Adapters (HAs) which are the external interfaces. Each adapter supports two ports, either SCSI or ESCON. Fibre Channel or FICON HA's support one channel per HA. Each HA connects to *both* Cluster Processor Complexes so that either cluster can handle I/Os from any host adapter.

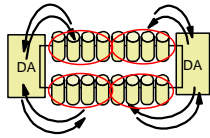
Overview of ESS Logical Structure



There are two Cluster Processor Complexes in the ESS, each working independently. Each contains four 64 bit RISC Symmetric Multi Processors (SMP), with up to 16GB of cache (total 32GB), non volatile cache (NVS), and four device adapters. All host data is written to and from the cache using adaptive record, partial track, sequential or full track staging algorithms. In addition all write data is automatically placed into two separate caches for integrity. The NVS cache has its own 7-day battery and is managed using a least recently used (LRU) algorithm. On data writes, one copy of data is placed in one cluster and the second copy of the write data is placed in the NVS of the other cluster. In the event of a either a planned or unplanned cluster outage, write data for the failed cluster is available on the other cluster and the remaining cluster takes over the functions in progress from the failed cluster. The design philosophy is that once data is in the ESS, it is intended to be fully protected.

Within each cluster the Device Adapters (DAs) are used to connect disks to the Cluster Processor Complexes. DAs are always installed in pairs, one in each cluster. Disks arrays, or ranks (IBM 9, 18 or 36 GB disks) are attached to the two DAs via 160 MB Serial Storage Architecture (SSA) loops. The ranks can be configured as RAID 5 or non-RAID arrays (Just a Bunch of Disks -- JBOD).

The ESS uses the latest IBM 160 MB Serial Storage technology in its device adapters which connect disks to the cluster processors. Disks are connected via a serial loop architecture with two simultaneous read and write data links. Each link operates at 40 MB per second for a total bandwidth of 160 MB around the loop. Each Device Adapter supports two independent SSA loops, giving a total bandwidth of 320 MB per second. There are four pairs of Device Adapters in an ESS, providing a total disk bandwidth capability of 1,280 MB per second.



Up to 8 simultaneous read / write I/O operations can be performed at once per loop. Up to 8 loops and 64 concurrent I/O operations can be performed.

- Use 18 GB drives for most applications.
- Use 9 GB for the most demanding performance requirements.
- Use 36 GB drives for large capacity and less demanding performance requirements.
- Every disk can transfer data at the same time

Up to 128 disks can be installed in the ESS base frame. For larger capacity requirements the ESS Expansion enclosure is available that holds an additional 256 disks.

Due to advanced designs there is negligible RAID 5 write penalty.

Data is striped across RAID ranks which achieves the highest possible device performance.

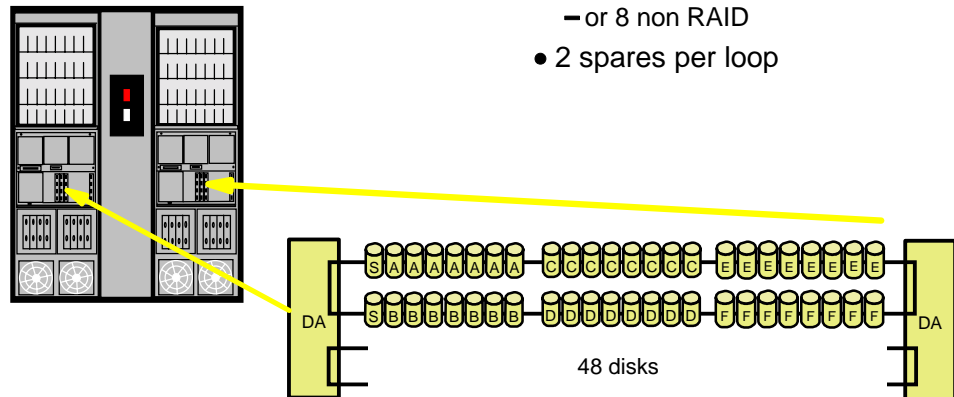
Components..Device Adapters

SSA 160 Device Adapters

- 4 DA pairs per Subsystem
- 4 x 40 MB/sec loop data rate
- 2 Loops per Device Adapter pair

Up to 48 disks per loop

- No mix of '8-packs' and drawers
- Each group of 8 is
 - RAID-5 array, 6+P+S or 7+P
 - or 8 non RAID
- 2 spares per loop



A RAID-5 collection of disks is called a rank. Illustrated above are 6 RAID ranks labeled A-F. A and B both have spare disks that are used across the loop in case of a disk failure. In the event of a disk failure the failed disk is replaced and becomes the new spare. The rank is the basic unit of capacity in an ESS and is owned by one Logical Subsystem, in either fixed block (open systems) or Count Key Data (S/390) format. A rank is formatted as a set of Logical Volumes (LVs). The number of LVs in a rank depends on the capacity of the disks in the rank (for example 9 GB, 18 GB or 36 GB) and the capacity of the logical disks being emulated (for example 3390-3 for S/390). As an example, a group of 8 disks (6+P+S) has capacity of about 53 GB (assuming 9 GB disks). This can be formatted into 18 3390-3 disks or one 53 GB LUN for use by a UNIX system. RAID operations (such as parity generation, etc.) are managed by the DA.



A non-RAID option is also available, however, in the event of a disk failure all data on the failed disk would be lost. This option might be used if mirroring is done by the server operating system or application software.

You have a choice of 32 standard configurations to meet your business needs. Capacity on the standard configuration scales from 420 GB to over 11 TB. Portions of the storage may be configured as RAID-5 or non-RAID. Storage capacity is customized as logical volumes of variable sizes. LUNs are used for UNIX, Windows NT and AS/400 storage. S/390 storage may be identified as 3880 or 3390 track format volumes. The chart below shows the original E models and the newest F model availability's and capacities:

ESS has internal batteries that supply power for a few minutes where power is lost or to bridge during fluctuations.

For extended power outages the ESS will shutdown gracefully after destaging of all write data.

Many configurations are possible. The choice depends on your performance and capacity requirements.

18 GB Drives	E-model availability	F-model availability	36 GB Drives	E-model availability	F-model availability
420 GB	available	6/16 - N/A	840 GB	4/28 - N/A	4/28 - N/A
630 GB	available	6/16 - 6/30	1260 GB	4/28 - 6/30	4/28 - 6/30
840 GB	available	6/16 - 6/30	1680 GB	avail-6/30	avail-6/30
1050 GB	6/16 - 7/28	6/16 - 7/28	2110 GB	4/28 - 7/28	4/28 - 7/28
1260 GB	available	6/16 - 6/30	2520 GB	available	3/31 - 6/30
1470 GB	6/16 - 7/28	6/16 - 7/28	2940 GB	4/28 - 7/28	4/28 - 7/28
1680 GB	available	6/16 - 6/30	3660 GB	available	avail-6/30
2170 GB	4/28 - 7/28	4/28 - 7/28	4340 GB	4/28 - 7/28	4/28 - 7/28
2660 GB	available	3/31 - 6/30	5320 GB	4/28 - 6/30	4/28 - 6/30
3150 GB	4/28 - 7/28	4/28 - 7/28	6300 GB	4/28 - 7/28	4/28 - 7/28
3640 GB	available	3/31 - 6/30	7280 GB	available	avail-6/30
4130 GB	4/28 - 7/28	4/28 - 7/28	8260 GB	4/28 - 7/28	4/28 - 7/28
4620 GB	available	3/31 - 6/30	9240 GB	4/28 - 6/30	4/28 - 6/30
5110 GB	4/28 - 7/28	4/28 - 7/28	10220 GB	4/28 - 7/28	4/28 - 7/28
5600 GB	available	3/31 - 6/30	11200 GB	available	avail-6/30

New ↑ ↑ Upgrades

There are various configuration options depending on the operating systems you plan to use on the ESS. Some of the choices are shown below:

Configuration Options

Many new hardware functions are now available for open systems hosts that were not possible before.

For S/390, there are new performance features that can significantly reduce execution run times.

- Open System considerations**
- Open System support
 - pSeries, AIX
 - PC Servers
 - Windows NT and 2000, Novell NetWare
 - iSeries, OS/400
 - Sun, Solaris
 - HP 9000, HP-UX
 - Data General, DG/UX
 - Compaq Alpha, OpenVMS & Tru64 UNIX
 - Linux Servers
 - SCSI parallel device support
 - 15 Targets, 64 LUNs
 - SCSI-FCP device support
 - 1 Target per adapter on host system
 - 4096 LUNs per target with LUN masking
 - Open System LUN's
 - 0.5 GB - 239 GB per logical volume
 - iSeries LUN's
 - 4.19GB, 8.59GB, 17.54GB, 35GB, 36GB
 - Subsystem Device Driver
 - Configure via ESS Specialist

- S/390 considerations**
- 16 control unit images
 - 3990-6, 3990-3, 2105 emulation
 - Up to 256 Devices / control unit image
 - 3390-2/3/9 with 3390 track
 - 3390-2/3 with 3380 track
 - Maximum 32 ESCON ports
 - Each 64 logical path
 - Maximum 2048 logical paths
 - 128 Logical Path / control unit image
 - Maximum 16 FICON ports
 - Each has up to 256 logical paths
 - Maximum 2048 logical paths
 - 128 Logical Path / control unit image
 - OS/390, VM, VSE, TPF, z/VM, z/OS, Linux
 - Configure via ESS Specialist

ESS Open System Overview

ESS supports AS/400, RS/6000, Netfinity, NUMA-Q, Intel, HP, Sun and other host servers. See for a complete list:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

ESS offers open systems users 24x7 availability with redundant design and non disruptive maintenance.

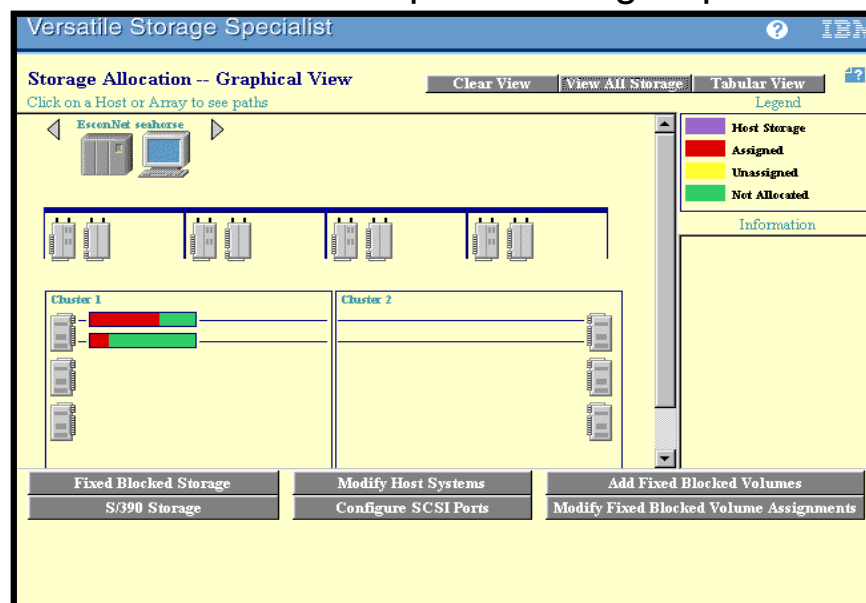
The StorWatch ESS Specialist enables centralized physical storage management which reduces cost and management complexity.

IBM 7133 and VSS storage systems capacity can be used with the ESS providing investment protection for assets already paid for.

TSM (formerly AD5M) can be used to backup ESS storage.

The ESS provides universal data access including: RS/6000 and SP2 running AIX, many leading UNIX variants, IBM Netfinity and other Intel-based PC servers running Windows NT and Novell Netware, and AS/400 running OS/400. This rich support of server platforms is not limited in any way; any combination of these heterogeneous platforms may be used with the ESS (including S/390). The storage capacity is partitioned among the attached servers using flexible Web-based StorWatch Enterprise Storage Specialist management tools. ESS's in one location are attached to a local area network and to an external processor (supplied with the ESS) that provides a means of accessing the StorWatch tools and can provide a gateway to your WAN for remote access and management from anywhere in the world. StorWatch ESS Specialist is used to configure, install, and allocate storage to the various host servers that will use the ESS. One sample Web window is illustrated below:

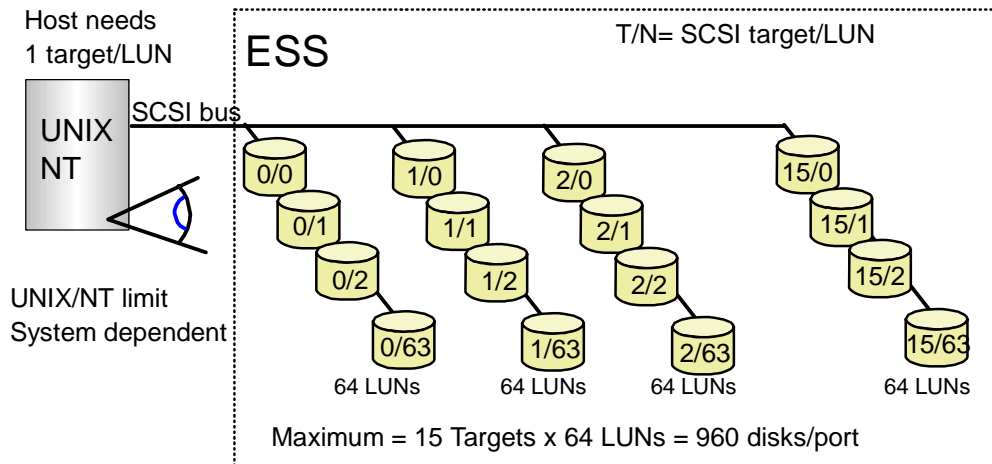
IBM StorWatch Enterprise Storage Specialist



For SCSI attached hosts, a storage facility consists of one or more SCSI I/O interfaces (buses), each with one or more SCSI targets. Each target has an associated set of logical units. Each logical unit has a unique logical unit number (LUN) on the target. A given target is accessible through a single SCSI bus. A given SCSI bus may access one or more targets. A set of logical devices are associated with a logical subsystem. A logical subsystem is managed by a single device adapter and a single Cluster Processor Complex at any given time. Since the ESS has two Cluster Processor Complexes, if one fails or is taken offline, the other cluster takes over all the work on the cluster that is no longer available. When one cluster is taken offline, write data that has not been de-staged is retrieved from the other cluster's NVS. Normal operations continue.

If you imagine a SCSI host's view of an ESS, it looks like a group of SCSI disks attached to a SCSI bus. The actual number that any UNIX or NT system can support is considerably less than the maximum shown below.

A UNIX or NT View of an ESS



Each Fixed Block (FB) Logical Subsystem supports up to 256 logical volumes.

Logical volume size can vary from 0.5GB to 245GB (the size of a RAID rank with 36GB disks)

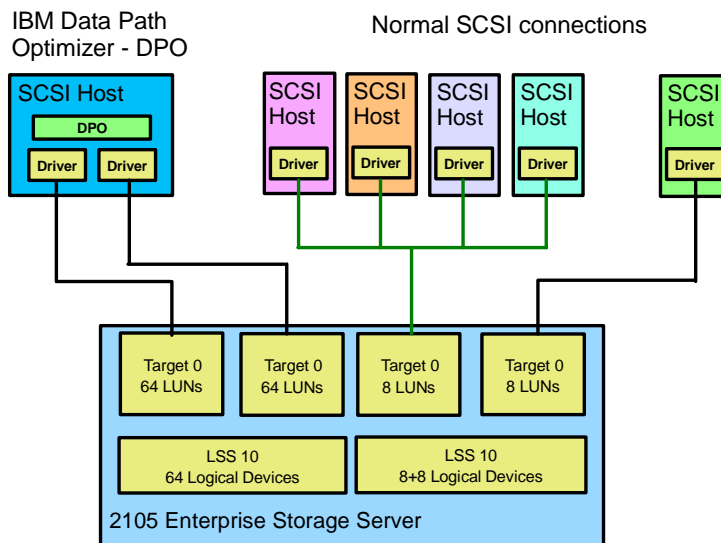
A total of 4096 logical devices can be defined.

IBM Subsystem Device Driver (SDD) provides:

- More than one path from the host to the ESS.
- A single LUN can appear as 2 - 16 LUNs
- Host path failover
- Load distribution across paths
- Supports SCSI AIX and Windows NT/2000.
- Supports NT, Windows 2000 and AIX Fibre Channel
- Sun and HP native Fibre Channel

One target / LUN is used for each host attached to the ESS SCSI port. Typically you will only have one host per SCSI port, leaving you with 15 target IDs and a number of LUNs per target. For availability purposes, you can configure a logical device in the ESS as a shared device using two different SCSI ports in the ESS. This will allow you to interconnect your host to two or more separate SCSI ports. If your host OS is AIX or Windows NT, you can use the IBM Subsystem Device Driver (SDD) to distribute I/O activity across the SCSI adapters in the host. SDD will recover I/Os that fail due to a path problem using an alternate path. A sample configuration is illustrated below:

SCSI Host Connectivity



Open Systems Performance Examples

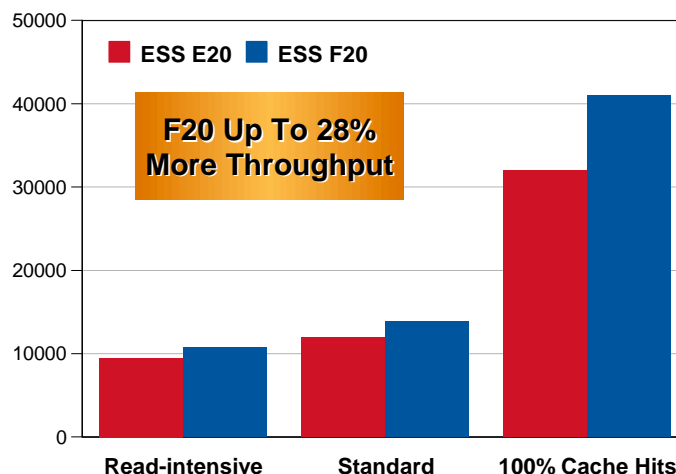
IBM performance tests have demonstrated that the ESS has significantly exceeded IBM's design goals. The example below used four RS/6000 processors, 32 Ultra SCSI ports, 128 18.2GB drives with 4K records achieving over 40,000 I/Os per second. No other competitor comes close.

Open Systems Maximum Throughput

This outstanding performance provides:

- Better handling of e-business unpredictable workloads
- Less manual tuning because the ESS stripes data requests over many disks
- Ability to handle peak loads with ease
- Ability to run both open systems and S/390 data on the same storage server
- The ESS can scale much higher than the 4 SCSI interfaces used for this comparison. (up to 32 concurrent transfers).
- As you increase the number of host interfaces and capacity the ESS throughput scales proportionately

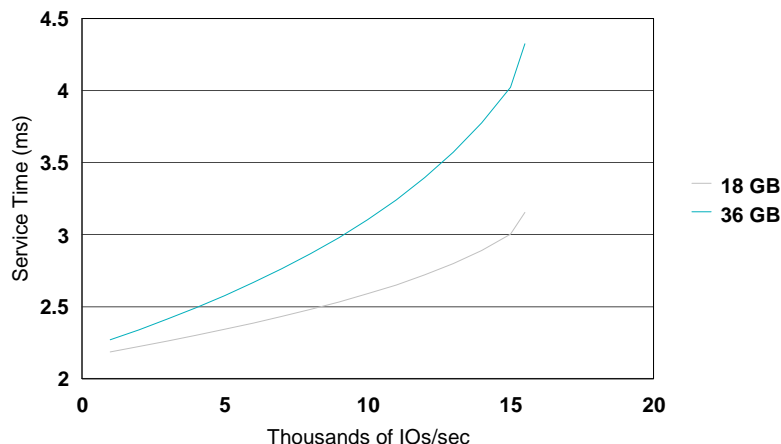
Maximum Throughput (Accesses per sec)



Four RS/6000 F50 processors; 32 Ultra SCSI ports; one ESS; 128 18.2GB disks; 4KB records.

The IBM 7133 family is regarded in the storage industry as the fastest non-cached storage solution (as the chart below shows). The ESS design is almost as fast as the 7133, with IBM leaving one storage competitor way behind in read intensive operations.

Performance of New Drive Family Cache Standard 3.4 TB



Open Systems Copy Functions

Some of the advantages with an ESS in an open systems environment are ESS's excellent performance, reliability, scalability, and serviceability. In addition to that, the ESS provides new functions for the open systems environment.

ESS hardware provides a number of distinct ways to share and copy open systems data.

PPRC is the ideal way to provide for business continuance in the event of a major outage at one location.

With PPRC the ESS ensures that write data is in two separate ESS's before letting the application continue.

PPRC is an industry standard..

PPRC is established on a volume level and can be used by midrange and S/390 systems.

PPRC uses direct connections between ESS systems with no host involvement.

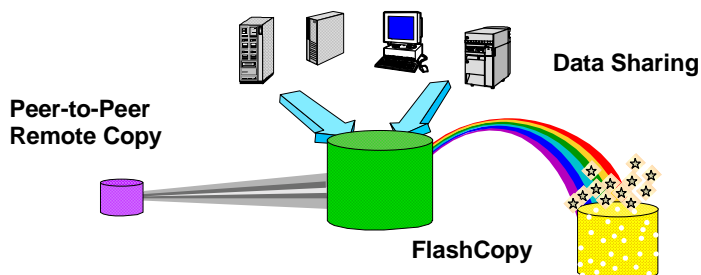
FlashCopy designed to make copies of volumes available in less than 2 seconds!

PPRC and FlashCopy are separate options on ESS systems.

More details on PPRC and FlashCopy are in the S/390 section.

ESS Functions for Open Systems

- Data Sharing
 - Access to the same data from like systems
 - Host software must manage data integrity
- Peer-to-Peer Remote Copy
 - Disaster recovery solution
- FlashCopy
 - Immediate copy available for backup, data mining and other use



ESS's capability to share logical volumes between like systems allows concurrent access to data bases by different hosts (both IBM DB2 Universal Database and Oracle Parallel Edition support this feature). The data base products provides locking mechanisms to guarantee data integrity.

With ESS's Peer to Peer Remote Copy (PPRC) function, midrange systems (NT, NetWare, UNIX, iSeries) can now realize a disaster recovery solution to protect data should one location be destroyed. PPRC provides a synchronous copy of selected data up to a distance of 103 km. PPRC is administered from the web browser interface of IBM's StorWatch ESS Specialist. No special software is required to use PPRC. ESCON cables connect the ESS systems. Write operations are completed on both copies (local and remote) before they are reported complete to the operating system. More information on PPRC can be found later on in this document.

FlashCopy technology -- available for the mainframe since July 2000 -- is now available for UNIX, iSeries or Windows NT host systems. FlashCopy provides a near-instantaneous copy of data and makes it available to applications as needed, without affecting the overall performance of the storage subsystem or the application being used by the host server. With traditional copy techniques, applications are often tied up for the entire duration of the time required to copy data -- which can be many minutes or even hours. Such an extended period of time is simply not an option for today's dynamic e-business applications.

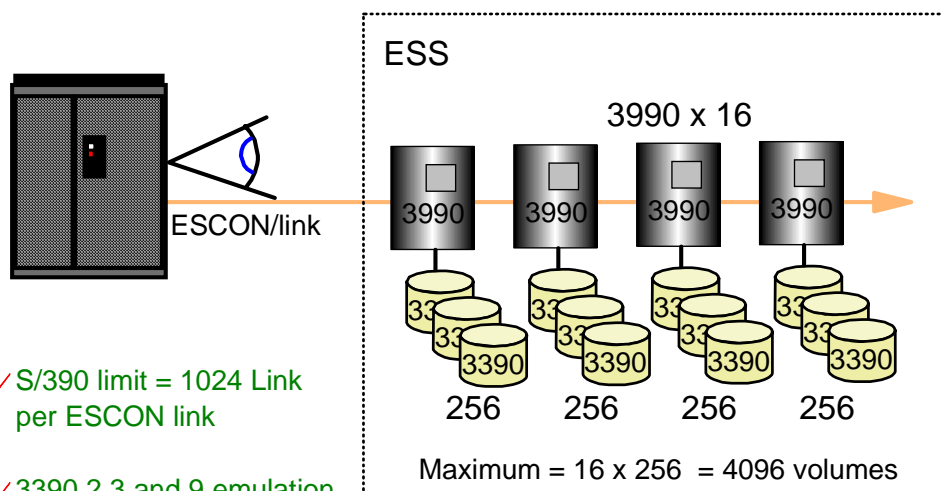
[zSeries \(System/390\) Overview](#)

ESS supports OS/390, z/OS, VM/ESA, z/VM, TPF and VSE/ESA, and S390 Linux operating systems.

The ESS supports S/390 servers and inherits the S/390 functions previously provided by the 3990-6 and RAMAC families. However, the ESS far surpasses these legacy technologies in all areas such as performance, hosts supported, availability, functions, scalability and non-disruptive operation. The ESS is designed specifically to support both S/390 and open systems workloads at the same time. Typically S/390 workloads are cache friendly and take advantage of large caches, whereas the open systems workloads often are very cache unfriendly. For the S/390 workloads there are sophisticated cache management algorithms and a large cache. For workloads that do not take advantage of large cache sizes, the ESS has high performance disk arrays with 160 MB disk loops. So whatever the workload, even mixed workloads, the ESS delivers high performance. The S/390 hardware view of the ESS is shown below:

S/390 View of Logical Subsystem

Up to sixteen 3990 control units can be defined in the ESS.



- ✓ S/390 limit = 1024 Link per ESCON link
- ✓ 3390 2,3 and 9 emulation
- ✓ 3380 track format
- ✓ Variable size 3390 and 3380 volumes - (Custom Volumes)

Custom Volumes may also be configured. Custom Volumes allow a volume to have any number of cylinders you choose. One would typically define a small number of cylinders to reduce contention for high activity data sets.

In most cases, however, the PAV function would be a better choice.

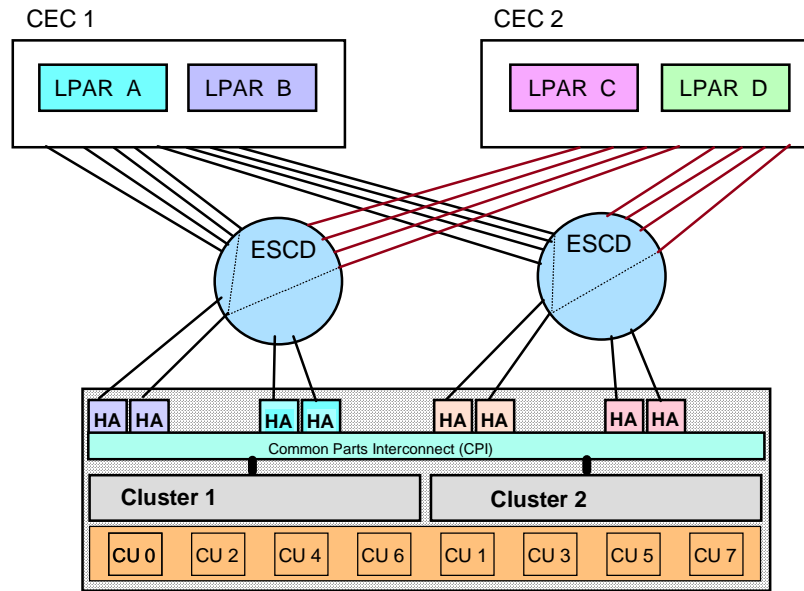
The number of 3390 or 3380 devices available depends on the type of disk installed (9 GB, 18 GB, 36 GB), how many are installed and the type of emulation you select.

The combination of StorWatch Specialist and HCD / IOCP are used to configure the ESS.

IBM has defined standard physical configurations that meet most customers requirements and are intended to make it easy to install an ESS. The standard configurations include options for ultra-high performance, large capacity, ESCON or FICON only, fix block only (open), and mixed environments (SCSI, ESCON). For complex environments, the ESS can be configured to match your exact requirements.

In the example below, two S/390 processors connect via ESCON links to an ESS with eight 3390 control units defined..

Sample ESCON Connectivity



The configuration on the right is fault tolerant in that HAs, clusters, disks, buses, disk adapters, ESCON directors, or paths can fail and the applications will not be affected. The same configuration can be set up using FICON but with less hardware and more performance.

FICON typically reduces ESCON channels 4 to 1.

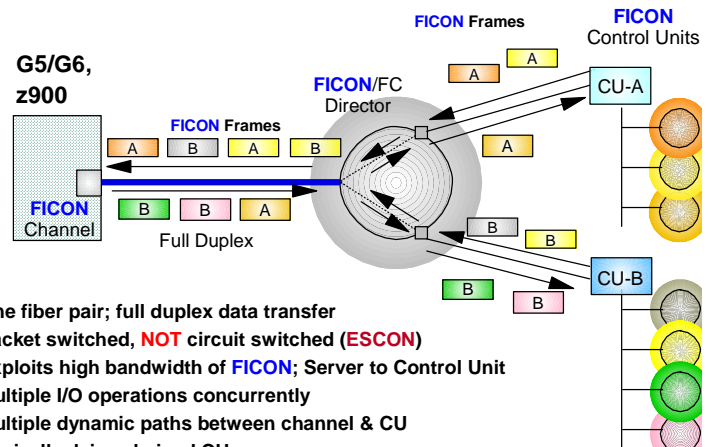
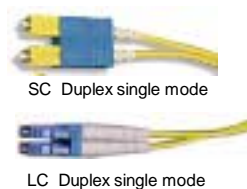
FICON has no performance degradation up to 100KM.

FICON has eliminated many ESCON architectural constraints - for example FICON channels can address up to sixteen times more devices on a channel than ESCON.

For tape operations the FICON enabled combination of the IBM A60 tape controller and the IBM 3590 tape drives with the FICON enabled ESS can provide significant reductions in disk and tape elapsed times. One example of this is shown on the right with a 144% thrupt improvement. Also backup elapsed times can be reduced up to 50% or more.

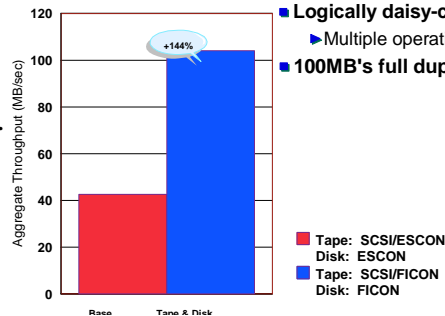
With high speed microprocessors, 160 MB disk interfaces, large cache and large capacity the ESS provides extraordinarily high performance and with FICON thrupt can be as much as **2.5 times better** than ESCON.

FICON Overview



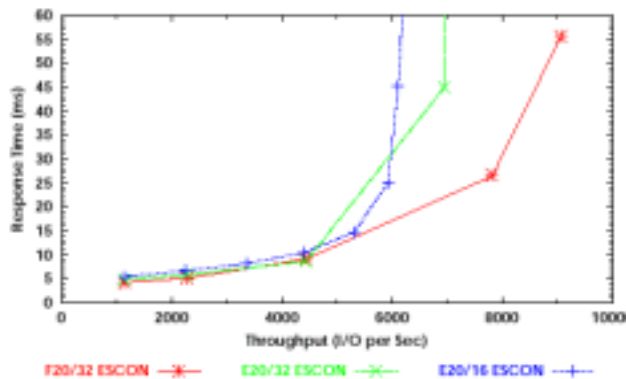
- One fiber pair; full duplex data transfer
- Packet switched, **NOT** circuit switched (ESCON)
- Exploits high bandwidth of FICON; Server to Control Unit
- Multiple I/O operations concurrently
- Multiple dynamic paths between channel & CU
- Logically daisy-chained CUs
 - Multiple operations at a time from / to *each* CU
- 100MB's full duplex 60-65mb per second per channel

DFSMSdss Full Volume Dump, Eight Tasks



Even with cache hostile workloads the ESS can drive a very high rate of I/O's with good response times.

ESS 1.1 Cache Hostile Performance



Read hit ratio = 45%, R/W ratio = 2:1

The ESS can handle significantly more work than is shown here as more channels and capacity is added.

ESS Priority I/O Queuing:

- I/Os can be queued in ESS
- Priority can be set by WLM
- I/O priority for systems in a SYSPLEX is honored

IBM's storage server design team knew they would have the fastest storage server in the industry for S/390 -- but they didn't stop there. They teamed with the IBM OS/390 software development lab and invented three new revolutionary improvements which can increase performance 2 to 3 times over what the raw hardware can deliver! The performance gains you will see are entirely dependent on your particular environment, however, as you understand these new innovations, it may become obvious how these innovations will apply to your environment.

Parallel Access Volumes (PAV) allow a single S/390 server to simultaneously process multiple I/O operations to the same logical volume (the SCSI architecture already provides this function for open systems hosts). **Multiple Allegiance** is a similar enhancement that expands the simultaneous logical volume access capability across multiple attaching OS/390 servers. These two functions enable the ESS to process more I/Os in parallel, dramatically improving performance and enabling greater use of large volumes, which helps to simplify storage management and reduce costs. **Priority I/O Queuing** allows the ESS to use information provided by the OS/390 Workload Manager (WLM) to manage the processing sequence of I/O operations so that they match application priorities. The ESS also has implemented enhanced commands that significantly reduce overheads, or time spent by the system but not directly related to the transfer of information, which further improves performance.

These innovations, which are called, the *ESS EX Performance Package*, can dramatically improve performance and significantly reduce the elapsed times of all the applications in your systems.

S/390 architecture only allows one I/O operation at a time to disk devices. The new ESS changes this so that more than one I/O can be active at a time to one disk device.

Since it was developed, System/390 architecture allows only one channel program to be active to a disk volume at a time to ensure that data being accessed by one channel program cannot be altered by the activities of some other channel program. OS/390 does not try to issue another I/O to a disk volume when a previous I/O is already active for that volume. Unit control blocks (UCBs) and OS/390 queues keep track of I/Os so they are processed one after another. The ESS and OS/390 software change all this.

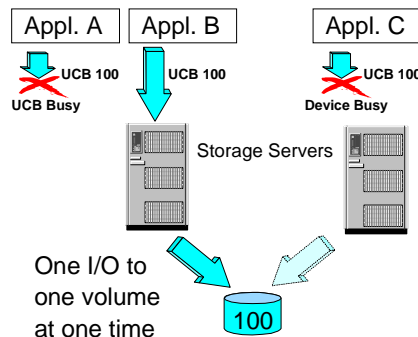
OS/390 ESS PAV Support:

- Transparency support has PTFs to use the ESS as a 3990 only
- Exploitation static support has PTFs that recognize the ESS with PAV support and can use alias and base addresses for I/O
- Exploitation dynamic support uses OS/390 2.7 plus PTF's. The Workload Manager will automatically tune the OS/390 I/O system using alias address by the use of policies

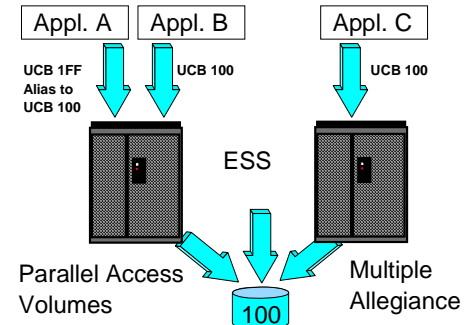
PAV's:

- Sharply reduces IOSQ time
- Provide multiple UCBs per logical volume
- Aliases are visible only to IOS in OS/390
- PAV support is an ESS option

Traditional OS/390 Behavior



Enterprise Storage Server (ESS) Behavior

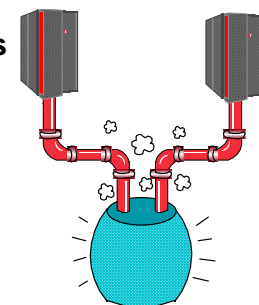


The ESS introduces PAV, a capability to do more than one I/O to a S/390 logical volume. The ESS introduces the concept of alias addresses. Instead of one UCB per logical volume, an OS/390 host can now use several UCBs for the same logical volume. The ESS also accepts I/Os to a shared volume coming from different hosts in parallel. This capability is called Multiple Allegiance. In the table below, the first row shows hosts A and B running separately to a DB2 volume. The second row illustrates how host B can significantly slow down host A I/O. With the ESS and Multiple Allegiance, there is no slowdown, as shown on the third row!

Multiple Allegiance

Concurrent access from multiple Path Groups (system images) to a volume

- Incompatible I/Os are queued in the ESS
- Compatible I/O (no extent conflict) can run in parallel
- ESS guarantees data integrity
- No special host software required, however:
- Host software changes can improve global parallelism (limit extents)



Example

Improved system throughput

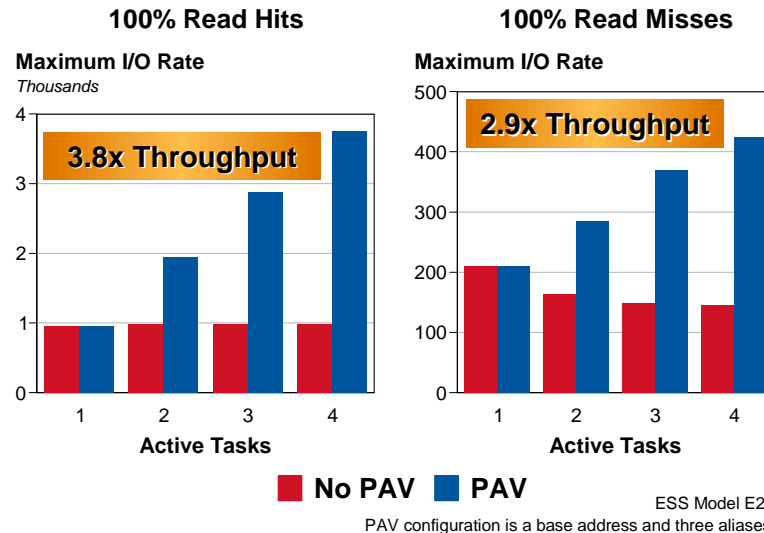
- Different Workloads have less impact on each other

	Host 1 (4K read hits)	Host 2 (32 record 4K read chains)
Max ops/sec - Isolated	767 SIOs/SEC	55.1 SIOs/sec
Max ops/sec - 100% Extent Conflicts	59.3 SIOs/SEC	54.5 SIOs/sec
Max ops/sec - full Multiple Allegiance	756 SIOs/SEC	54.5 SIOs/sec

These state of the art advancements in the ESS result in an extremely fast storage subsystem. Performance is dependent on the workload and particular hardware used, however, the illustration below is typical of what one might see when using the ESS and PAV:

PAV Effect for Database Workloads on One Volume

This chart demonstrates the outstanding performance improvements using PAV's. Imagine this effect for more than one volume!



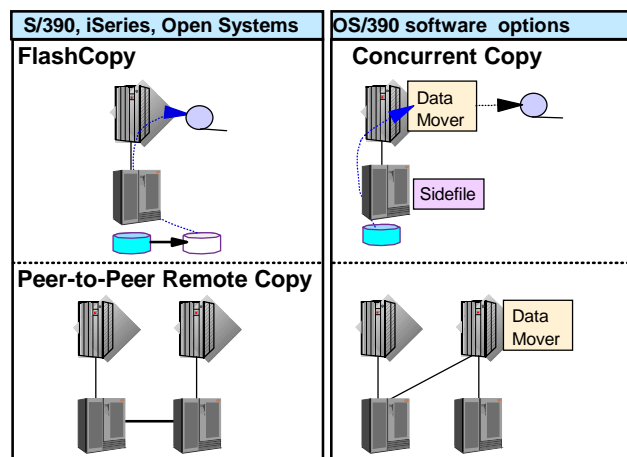
ESS Copy Services

The ESS supports several hardware copy functions for two purposes: mirroring for disaster recovery solutions, and copy functions that provide an instant copy of data.

Copy functions are supported and can be initiated by OS/390. UNIX systems use the StorWatch Specialist software to initiate and control copy services.

- FlashCopy provides an instant copy of a volume for open systems, S/390, iSeries
- Concurrent Copy for S/390, of a volume or data set
- Peer to Peer Copy (PPRC) is synchronous mirroring: open systems, S/390, iSeries
- Extended Remote Copy (XRC) provides asynchronous mirroring for OS/390

PPRC and XRC copy architecture is an industry standard.



FlashCopy is an optional feature of the ESS. S/390, iSeries and open systems can use it.

FlashCopy can be used to:

- Create test data
- Create temporary checkpoint copies
- Asynchronous backup - copy data and make the backup from the copy while your application access the source
- Perform Data Mining and Data Warehousing on the copy of the data.

FlashCopy can be started via z/OS DFSMSdss or the StorWatch Specialist.

z/VM has the ability to FlashCopy mini disks in addition to entire real or virtual volumes.

VSE/ESA can also dataset level for certain types of datasets.

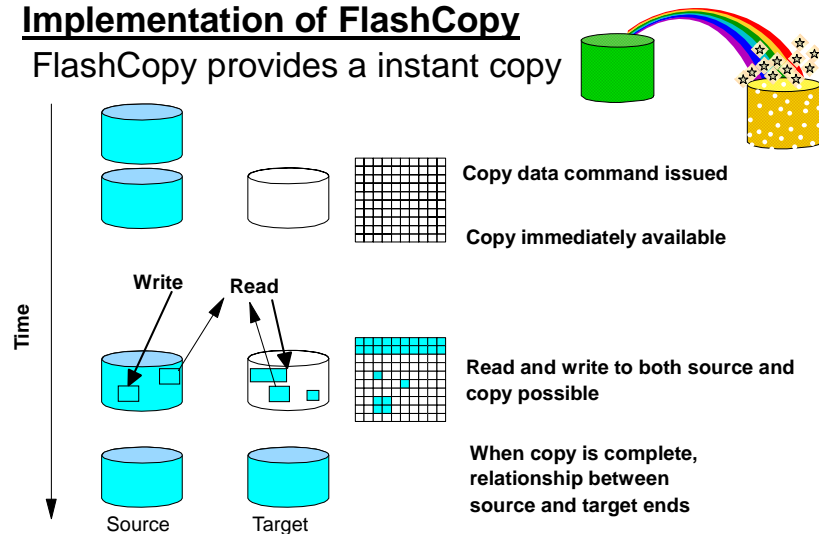
PPRC is used for business recovery in the event of a data center disaster.

S/390 and open systems can use the function at the same time.

With SnapShot™ on the IBM RAMAC Virtual Array (RVA) or Concurrent Copy you get a copy of the original volume or data set in a few seconds. The FlashCopy function provides a similar capability on the ESS for volume copies. When using VM entire volumes or minidisks and be copied and with VSE/ESA certain types of datasets can also use FlashCopy. As with SnapShot, you get an almost instant copy when you start the command. Once the relationship between the source and target volumes is established, a background task copies the tracks from the source to the target. As the background task is making the copy, both the source and target volumes can be used for reading or writing independently.

Implementation of FlashCopy

FlashCopy provides a instant copy



PPRC is an ESS hardware remote copy or mirroring solution that can help preserve data integrity, even in a rolling disaster.

Peer-to-peer Remote Copy (PPRC)

■ Concept

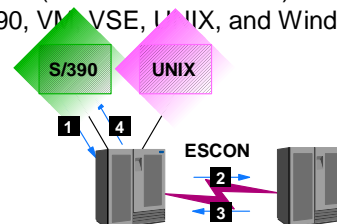
- Synchronous copy, mirroring (RAID-1) to another ESS
- Disaster recovery solution
- Established on a volume level
- Direct connections between ESS systems using ESCON links
- Full GDPS support

■ Interfaces

- TSO/E for OS/390 systems (or ICKDSF for S/390)
- ESS Specialist for OS/390, VM, VSE, UNIX, and Windows NT

Benefits

- Data Integrity protection
- Remote Mirroring
- De facto standard
- Offloads server
- Common facility for all attached hosts



PPRC is a defined architecture that has many unique data integrity and performance features.

PPRC for OS/390 can be controlled by either ICKDSF, TSO/E commands, or StorWatch Specialist.

Open systems use the StorWatch Specialist Copy Services functions.

Would your critical applications survive a real disaster?

GDPS provides automated services using the ESS to provide near continuous operation, even in a disaster.

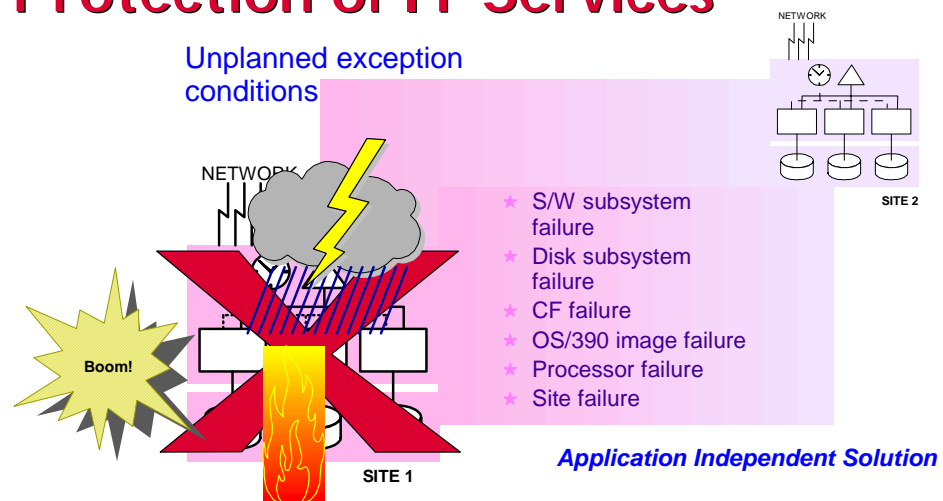
GDPS is only for OS/390 systems.

Offers the combination of CMOS processors, ESS, high performance, and IBM Global Services to provide a comprehensive disaster recovery solution.

PPRC is a synchronous mirroring solution, which means that an I/O does not complete until it is acknowledged from the remote site. PPRC is a de facto standard in that other vendors have licensed the architecture from IBM to use in their storage subsystems. PPRC is set up on a volume or LUN basis. Two or more ESS's are connected by ESCON links (fibre optic links using the ESCON protocol). Updates to a PPRC volume on the primary site go first in the cache and non volatile storage in the primary storage control (see (1) in figure above). The updates are then sent over the ESCON links to the remote the ESS (2). When the data is in cache and NVS on the secondary site, receipt of the data is acknowledged (3) and the primary the ESS signals the application that the I/O is complete (4). This process places data in two places before the application considers the operation complete. In the ESS, the efficiency of the ESCON protocol has been significantly improved over prior implementations which provides extended distances of up to 103 km.

PPRC is very powerful as a stand alone solution, but it is not enough for the most demanding customers, such as banks, hospitals, e-business, and 24x7x365 solutions. If one site goes down, there is no automatic way for the other site to take over or even know a site is no longer available. This is illustrated below:

Protection of IT Services



IBM has an answer to this dilemma for System/390. It is called Geographically Dispersed Parallel SYSPLEX (GDPS). GDPS is a multisite management facility that is a combination of system code and automation that utilizes the capabilities of Parallel SYSPLEX technology, storage subsystem mirroring (PPRC) and databases to manage processors, storage and network resources. It is designed to minimize and potentially eliminate the impact of a disaster or planned site outage. It provides the ability to perform a controlled site switch for both planned and unplanned site outages, maintaining full data integrity across multiple volumes and storage subsystems and the ability to perform a normal DBMS restart (not DBMS recovery) at the opposite site.

IBM has proven the GDPS concept with multiple customers. For example, with GDPS, a recent simulated site disaster at a customer site caused no data loss and the recovery window was reduced from 12 hours to 22 minutes.

Geographically Dispersed Parallel SYSPLEX consists of a base or Parallel SYSPLEX cluster spread across two sites separated by up to 40 kilometers (km) — approximately 25 miles — with one or more OS/390 systems at each site. The multisite Parallel SYSPLEX cluster must be configured with redundant hardware (for example, a Coupling Facility and a SYSPLEX Timer® in each site) and the cross-site connections must be redundant.

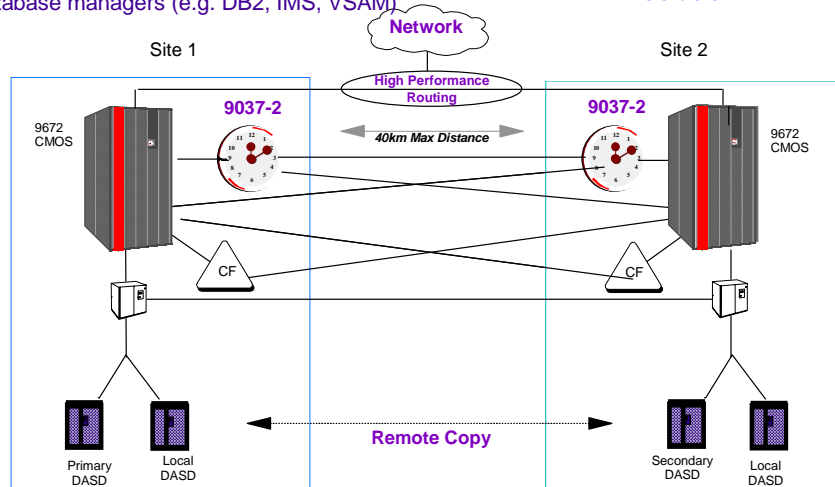
IBM GDPS provides:

- near continuous application and data availability
- Planned application availability across both reconfigurations and unplanned reconfigurations
- Provides data consistency and integrity. Simplifies operations from a single point of control.
- Application independence

A Key ingredient of GDPS is the ESS PPRC function, with the high performance and data integrity the ESS provides.

GDPS Topology

- ★ Multi-site SYSPLEX
 - ★ Remote DASD Mirroring
 - ★ Automation
 - ▶ Tivoli NetView for OS/390
 - ▶ System Automation for OS/390
 - ★ Transaction managers (e.g. CICS, IMS)
 - ★ Database managers (e.g. DB2, IMS, VSAM)
- Planned and unplanned exception conditions
- Application Independent Solution*



GDPS consists of production systems, standby systems and controlling systems. The production systems execute the mission critical workload. The standby systems normally run expendable work which will be displaced to provide processing resources when a production system or a site is unavailable. GDPS uses a combination of storage subsystem, Parallel SYSPLEX technology, and environmental triggers to capture, at the first indication of a potential disaster, a data consistent secondary site copy of the data, using the new, recently patented PPRC freeze function.

GDPS consists of on-site IBM Global Services to plan for the installation, installation of GDPS automation software, and customizing the environment.

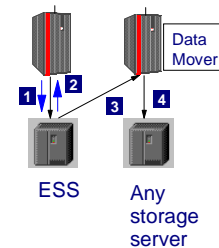
Extended Remote Copy was previously available on IBM 3990 control units and is for OS/390 systems only.

PPRC and XRC can also be used to move data from one ESS to another, for instance if leases are up. P/DAS is an OS/390 function that can be used in conjunction with PPRC and XRC. It is designed to allow primary and secondary volumes to be swapped without an application outage.

For those situations that don't require synchronous mirroring or require long distances, Extended Remote Copy (XRC) provides the solution. XRC is a de facto standard that is an asynchronous remote mirroring solution. It requires the OS/390 System Data Mover (SDM) to move the data. Applications doing write I/Os to primary (source) volumes (see 1 below) receive "write complete" as soon as the data has been secured in cache and NVS of the primary ESS (2). The SDM reads the update to XRC source volumes from cache and sends it to the secondary volume on a remote storage subsystem. The ESS implementation of XRC is enhanced with full suspend / resume, which enables the secondary copy to be re-synchronized following either a planned or unplanned telecommunications outage without performing a time consuming full-volume copy operation.

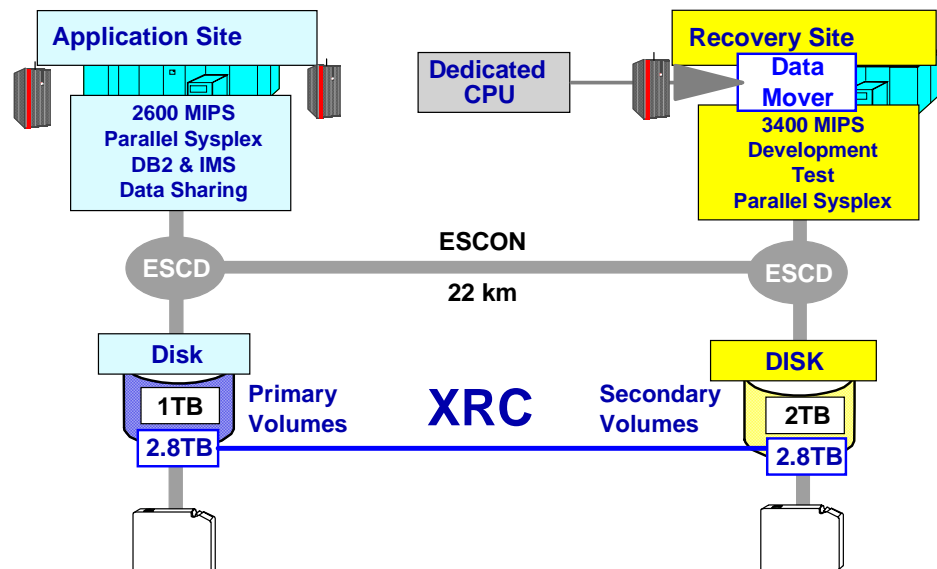
ESS XRC Implementation

- Remote mirroring of volumes
- Asynchronous copy process
- Supports large distances
- OS/390's System Data Mover (SDM) moves data from primary subsystem to a remote storage subsystem
- ESS's XRC implementation is compatible with the previous implementation on IBM 3990-6
- ESS supports all advanced XRC V2 functions
 - Multiple reader support (max 64 /LSS)
 - Dynamic balancing of application write bandwidth vs SDM read performance
 - Floating utility device
 - Or use 1-cylinder utility device
- Full Suspend / Resume support
- Use of new performance enhanced CCWs
- XRC for Open Systems volumes possible with XPE
 - Store your Open Systems data on 3390-3 type volumes
 - Copy these volumes with XRC from an OS/390



XRC can be used to clone an entire ESS or subset of volumes. This capability could be used to move a data center from one location to another as shown below.

Database Cloning via XRC



XRC is transparent to applications. XRC is controlled via TSO/E commands.

Enterprise Storage Server Summary

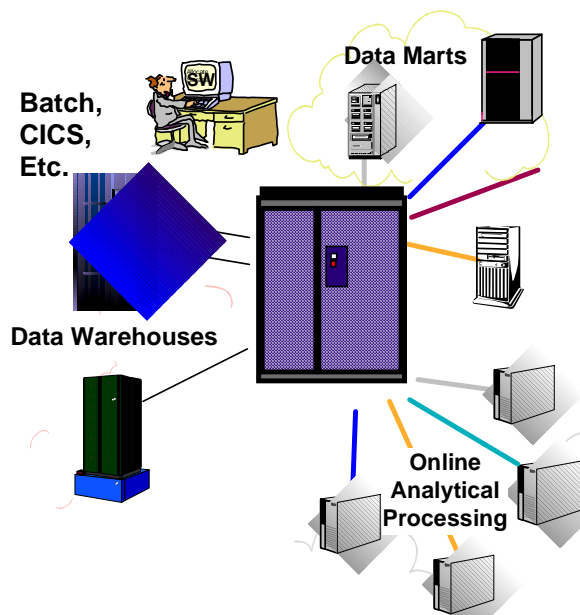
The Enterprise Storage Server is the storage of choice for those wanting the ultimate in enterprise disk storage consolidation and data sharing on multiple or heterogeneous server platforms with combined storage requirements of 420 GB or more. Whether it is direct attached or part of your storage networking strategy the ESS is designed to meet your needs now and in the future.

The Enterprise Storage Server is the longer term strategic enterprise disk storage solution for those that want to manage all of their storage from a single location. Enterprise Storage Server is the right choice for those requiring high-performance RAID-5, read and write cache, or the flexibility of common storage for multiple servers with ESCON, FICON, SCSI, and Fibre Channel attachments.

The Enterprise Storage Server supports rapid universal access to vast quantities of data through many advanced functions and features, making it a workhorse for support of business intelligence and other business-critical applications. Because of its enterprise-wide support and management scope, ESS is tailor made to help provide consistent, efficient and effective enterprise resource planning. Using ESS to address any or all of your strategic and tactical business initiatives will give your organization the business advantage needed to survive and thrive in the e-world.

*Your Data
is the
Business*

The Enterprise Storage Server is THE server for the next Millennium.



It has **PERFORMANCE WITHOUT BORDERS**



09/2001 Prepared by Kim Frank, Technical Architect 908-931-4588
KCFRANK@US.IBM.COM

Technical editing by Chris Saul, Storage Systems Division ES DAS Market Support

References in this publication to IBM products, programs, or services do not imply that IBM intends to make them available in all countries in which IBM operates.

This publication contains insights into IBM's future product development plans. These development plans are subject to change or withdrawal at any time without prior notice.

IBM reserves the right to change specifications or other product information without prior notice. All information contained in this publication is believed to be accurate as of the date of initial publication. IBM has no obligation to update any such information.

Copyright 2000, International Business Machines Corporation. All Rights Reserved.

Published performance information was modeled or obtained in a controlled environment based on the use of specific data. The results that may be obtained in other circumstances may vary significantly. IBM makes no representation and provides no assurances the same or similar performance results may be obtained. Users of performance information should verify the applicable data in their specific environments.

System/390, S/390, OS/390, CICS, IBM Parallel Sysplex DB2, MQseries, IMS, FICON, ESCON, VTAM, RS/6000, AS/400, OS/400, VM/ESA, TPF, VSE/ESA, PAV, AIX, ESS, PPRC, XRC, FlashCopy, RAMAC, DB2, StorWatch, SSA, P/DAS, iSeries, zSeries, xSeries, pSeries, z/OS, z/VM, and NetFinity are trademarks or registered trademarks of International Business Machines Corporation in the US and other countries.

Lotus, and Domino are trademarks or registered trademarks of International Business Machines Corporation in the US and other countries.

Windows and NT are trademarks or registered trademarks of Microsoft Corp. in the US and other countries.

SnapShot is a trademark and the property of Storage Technology Corporation for a duplication product.

UNIX is a trademark of The Open Group.

IBM Storage www.ibm.com/storage and www.storage.ibm.com/hardsoft/products/ess/supserver

GDPS <http://www.as.ibm.com/asww/offerings/mww62b1.htm>

IBM software www.ibm.com/software

IBM Servers <http://www-1.ibm.com/servers/home.html>