A Course in Storage Technologies from EMC Corporation for use in Computer Science and /or Information Technology Curricula

Ed Van Sickle

Abstract— **EMC Corporation, the world leader in data storage, created the EMC Academic Alliance Program to educate students on storage and close the education gap that exists. EMC developed a Storage Technology Foundations course to teach students about the design of storage technologies and the "big picture" of an information infrastructure. The course is "open" and focused on storage technologies, not products. College and universities use the course to teach students about a very important topic in IT: Storage. EMC partners with colleges and universities by providing the course, knowledge transfer sessions to faculty and program support. There is no cost to join and no cost to obtain the courses. EMC requires partners to sign an agreement for course use. Several colleges are using the course as an upper level elective offering and the course is taught by faculty. The alliance program has reduced faculty time to develop a storage course and time to learn the topic. Faculty is responsible for credentialing students and they supplement the course with additional materials. Students are being recruited for jobs by EMC and others, including internships. The Alliance program provides academic institutions with a way to differentiate. This paper will explain the program and the Storage Technology Foundations course.**

***Index Terms*—Computer Science Education, Industry - Education Relationships, Information Science Education, Storage Courses and Curriculum, Storage Technologies.**

I. INTRODUCTION

Corporate CIO's and IT Managers understand that the single most critical asset of any organization is their data. They understand that more data is being created, in various formats (audio, video, mpeg files) and they have regulatory requirements which require data to be available on line, for longer periods of time. IT managers face the task of creating an information infrastructure which can store, protect, manage, optimize and leverage this information. IT departments are implementing SAN (Storage Area Networks), NAS (Network Attached Storage), and Data Replication technologies to solve the problems of storage capacity, data availability, and data mobility. Spending data reveals that storage is an important topic on the minds of CIO's and increased spending on storage makes it one of the fastest growing segments of IT.

According to the *CIO Magazine* Tech Poll for 2007 [1], over 50% of respondents ranked storage and servers as the top items for spending increases over the next 12 months. Comparatively, many expected spending to be flat for networking and telecommunications equipment, infrastructure software and eBusiness. According to a recent Gartner CIO survey, storage technologies were ranked fifth of the top ten technology priorities. CIO's will need to exploit new approaches and technologies to support the business needs of their organizations [2]. In an analysis of U.S. Department of Labor job data and a survey conducted by EMC, it is estimated that over 1 million new storage jobs will be created worldwide by 2012 [3].

Despite all the news and spending data, for most IT professionals, it seems that storage infrastructures are still relatively unknown. This is interesting when considering that storage is not new and has been in existence for some time. Why then, is so much emphasis being placed on information infrastructure technologies now?

A major reason is that system downtime is expensive and not having access to data can cost a company millions of dollars (Figure 1). Information must be continuously available to support the business.

An IT infrastructure that can support this primary requirement is highly desired. Today's IT departments are implementing Information Management Infrastructures to meet those needs.

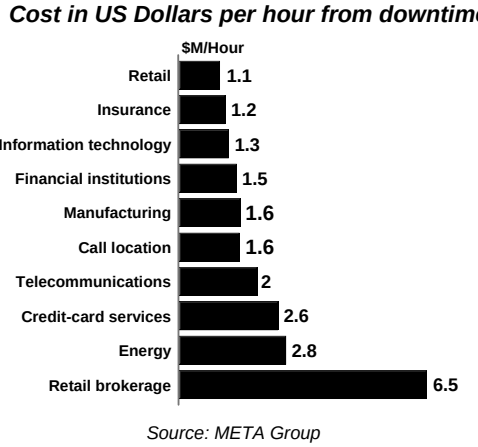

Fig 1 - Cost of Downtime

Typically, at the core of these Information Infrastructure solutions are intelligent storage disk arrays. Intelligent arrays provide organizations with the ability to store, protect, manage, optimize and leverage their data.

Depending on the business needs and solution requirements, different storage technologies can be used to solve the needs for managing data. These technologies include SAN (Storage Area Networks), NAS (Network Attached Storage), DAS (Direct Attached Storage) and CAS (Content Addressable Storage). These technologies and the advancements in hard disk drives and data delivery through high speed networks (fibre channel and IP networks) have enabled organizations to keep data continuously available. Additionally, for each type of storage technology, there is array-based data replication software functionality to copy and move data for backups, business continuance and disaster recovery, migration, and testing efforts.

So, if these storage technologies have become so prevalent, who provides the education on them?

Most vendors aren't. They usually provide training on their product offerings. Typically, the focus of vendor courses is on training, not education. Most Computer Science and Information Technology degree programs at colleges and universities aren't either. Their focus for technical course offerings are on computer architecture, operating systems, databases, networking and software application development – but not storage technologies. If we look at today's IT infrastructure, it's important to note that OS, RDBMS, Networks, Applications **and Storage** are integrated together to form the Five Pillars of IT (Figure 2) [4]. For example, we can see all these technologies in use when an end user using a software application makes a read or write request for data over the network. The operating system processes the request, the database has organized the data into tables and the disk array sends or receives the data, protects it and secures it.

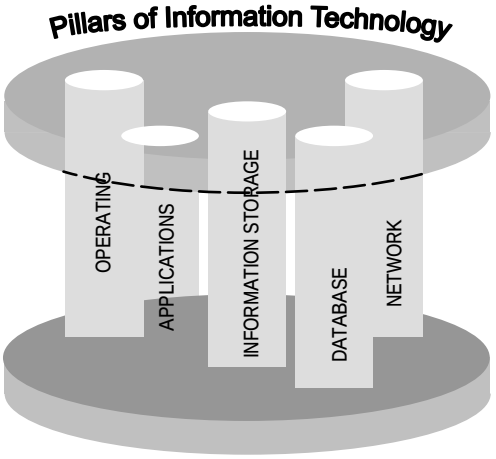

Fig 2- Five Pillars of IT

Most colleges and universities educate students on four of the five Pillars of IT. EMC conducted an extensive survey of IT professionals to acquire data about the storage skills gap. The results of the study drove EMC to develop solutions to meet the needs of industry and academia.

II. EMC STORAGE STUDY AND FINDINGS

Storage and information management infrastructures are being sold rapidly – it's a $60 billion (USD) market in 2007 and growing, which is creating high demand for people with these skills and knowledge (labor). But the supply of available people is low. Because the 5th Pillar is not being addressed in industry and academia, a skills shortage exists in the market. IT departments have the need and hiring demand as they create and implement these information infrastructures with storage arrays as the backbone layer.

EMC conducted a study (global survey of over 1,200 IT Professionals) and identified the key challenges for managing increasingly complex information infrastructures [5]. Figure 3 identifies the key challenges faced by IT / Storage Managers [5].

Challenges Identified by IT and Storage Managers and Professionals
• Managing storage growth
• Designing, deploying, and managing backup and recovery
• Designing, deploying, and managing disaster recovery solutions
• Making informed strategic/big-picture decisions
• Designing and deploying multi-site environments
• Designing and deploying emerging storage technologies (such as storage virtualization, IP SAN, GRID, etc.)
• Lack of skilled storage professionals
• Managing data availability/data retention compliance

Fig 3- Key Challenges for Managing Storage

An activity such as Back Up has been in practice for decades, yet professionals believe they are not doing it to the desired level. Why? EMC believes that this is due to the complexity of the environment and partially due to the fact that individuals managing the environment have not been provided a comprehensive education on storage technologies. Figure 4 further substantiates the complexity issue and identifies key pain points, by priority [5].

IT and Storage Managers	
80%	Managing storage growth
61%	Designing, deploying, and managing backup and recovery
58%	Designing, deploying, and managing disaster recovery solutions
48%	Making informed strategic/big-picture decisions
40%	Designing and deploying multi-site environments
36%	Designing and deploying emerging storage technologies (such as storage virtualization, IP SAN, GRID, etc.)
30%	Lack of skilled storage professionals
27%	Compliance regulations
Storage Professionals	
73%	Managing storage growth
62%	Designing, deploying, and managing backup and recovery
61%	Designing, deploying, and managing disaster recovery solutions
42%	Making informed strategic/big-picture decisions
42%	Designing and deploying multi-site environments
36%	Designing and deploying emerging storage technologies (such as storage virtualization, IP SAN, GRID, etc.)
31%	Lack of skilled storage professionals
23%	Compliance regulations

Fig 4- Key Pain Points for Managing Storage in order of priority by Managers and Professionals

With an "educated" professional, EMC contends that many of these issues would not be pain points, but tasks performed to desired levels.

Figure 5 reveals hiring plans and requirements for storage personnel over the next 12 months [5]. The lack of skilled professionals becomes a serious bottleneck. A major factor for this skills shortage includes the lack of storage technology education in academia and industry. This skills gap needs to be closed.

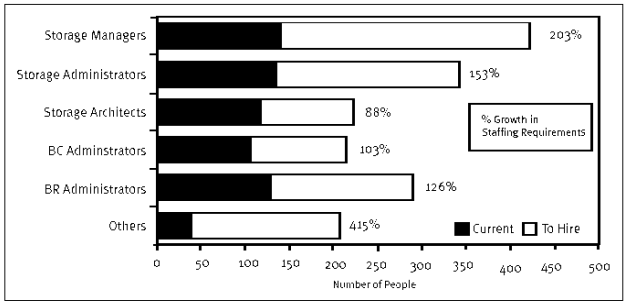

Fig 5- Hiring Requirements of next 12 months

Colleges and universities can close this gap and provide students with a needed education on Storage Technologies through the EMC Academic Alliance Program. The following section describes the program.

III. EMC ACADEMIC ALLIANCE PROGRAM ORIGIN AND CONCEPT

To address the Storage Skills Gap issue, the EMC Education Services team took action with the following steps:

Research with storage customers revealed that the most common pain point was finding individuals who understood "the big picture" of an information infrastructure. IT organizations needed people that understood, at a high level, the Five Pillars of IT.

Research revealed that IT organizations expect to provide an individual with hands-on, product training on the tools that they use. The storage industry has not standardized on a common set of tools and there are many product offerings to perform the same task. Therefore, a focus on equipment based, task-based learning would be a mistake and posed a risk to students.

If a course was created for college students on storage technologies, it would benefit EMC, EMC partners, EMC customers and the storage industry in general (non-EMC customers who have the same needs).

EMC Education had employees with experience in academia and decided to create a set of courses focused on storage technologies, not products. It was decided to develop course materials for use by academia that focused on Storage Theory, Storage Design and Information Management skills, not on using storage equipment.

This "open" curriculum was needed if it were to be accepted by most academic institutions worldwide. Additionally, the course needed to use case study and classroom discussion for the applied learning aspect of the program.

Students preparing for roles in areas like Database or Networking would benefit because they will encounter storage infrastructures in their career and they will need to know about these technologies.

EMC also realized that they could educate far many more people through a partnership program than they could alone. By creating an Academic Alliance program, they could infuse storage education to tens of thousands of college students worldwide.

Executive level support at EMC was provided. Joe Tucci, Chairman, President and Chief Executive Officer of EMC, who serves on the President's Council of Advisors on Science and Technology (PCAST) and as Chairman of the Business Roundtable Task Force on Education and the Workforce, views the EMC Academic Alliance Program as serving the mission of PCAST, as well as a vehicle to improve technology education worldwide. It is an opportunity for EMC to give back to the community. Corporate giving, in-kind contributions, and volunteerism are some of the many ways EMC expresses its commitment to community. Funding is focused on two key areas: championing math and science education and strengthening local communities.

With executive support and funding from the highest levels, the EMC Academic Alliance Program was initiated in late 2005 and launched in July 2006.

IV. EMC ACADEMIC ALLIANCE PROGRAM STRUCTURE AND GOALS

The characteristics of the EMC Academic Alliance Program include a.) a teaching focus that provides CS/IT students with an education on storage b.) a course focused on storage design and management that explains theory and concepts, not products c.) a program that is provided at no cost to the institutes d.) supports the need for academic freedom, and where warranted, the need to supplement the materials d.) EMC support for knowledge transfer, student enrollments, guest lectures, and site visits e.) the exploration of potential research opportunities with EMC CTO Office f.) recruitment and hiring of students by EMC, partners and customers.

The structure of the program includes:

- There is no cost to join the program. EMC courses are also provided at no cost.
- EMC and University/College complete an Agreement.
- EMC remains as the owner of the course.
- Course has to be offered in an undergraduate or graduate degree program, for credit.
- Use of the materials in an adult education or other for-profit program is restricted and must be done through EMC's Learning Partner Channel.
- Course becomes part of the University/College degree program offerings. University/College schedules course

delivery and lists the course in their course catalogue and Web site.

- Universities/Colleges determine how they implement the course, as a special topic elective to Jr. / Sr. students, or as a core offering or permanent elective.
- University/College provides student instruction and the faculty member to teach the course. Credentialing is done by University/College.
- University/College uses the full materials to teach the course and can supplement the materials. Sections of the course can be used in other courses, provided a complete course offering is given. Sections of the course can not be used in other courses unless the complete course is being taught.
- EMC provides program support; training to faculty on material (no cost for attending a session), trademark/logo for use in collateral materials, guest lectures/ site visits, recruiting, research.
- EMC has included the course as part of the EMC Proven Professional Program so that if students are inclined, they can become EMC Certified. This is not a mandatory requirement of the program, but a supplemental offering and option.

The 2007 goals and metrics for the EMC Academic Alliance Program include;

- Sign Agreements with 100 schools worldwide (we will exceed this number and sign 150 by year end)
- Enroll 2,000 students who will complete the Storage Technology Foundations course
- Certify 500 students under the EMC Proven Professional Program

Assessing the program will be done by soliciting feedback from partners and students. EMC will adopt changes that benefit the alliance. The next sections of this paper will describe the Storage Technology Foundations course and its merits.

V. STORAGE TECHNOLOGY FOUNDATIIONS COURSE FROM EMC

The Storage Technology Foundations course provides a comprehensive introduction to Data Storage technology fundamentals. The selection of these topics for the course is based on the wide use, acceptance and adoption of these technologies by industry. Also, EMC's experience as the world leader in storage in providing storage solutions with these technologies (topics) and ability to provide knowledge transfer. Course participants will gain knowledge of the core logical and physical components that make up a Storage Systems Infrastructure. Throughout the course, students will be exposed to the following themes:

i.) The explosion in demand from businesses for highly available and secure access to data.

ii.) The Storage systems and infrastructure architectures and solutions available to support business needs.

iii.) The complexities and challenges in managing storage infrastructures.

Upon successful completion of the course, students should be able to:

a.) Describe storage technology solutions such as Storage Area Networks (SAN), Network Attached Storage (NAS), Content Addressed Storage (CAS).

b.) Understand and articulate the technologies and solutions available to support an IT Infrastructure including Business Continuity, Information Availability, Local and Remote Replication, Backup and Recovery and Disaster Recovery needs of businesses.

c.) Understand the key tasks in successfully managing and monitoring a data storage infrastructure

The course consists of the following sections and modules.

Section 1 - The Complexity of Information Management

Module 1.1 - Meeting Today's Data Storage Needs
- Define the challenges in data storage and the importance of data.

Module 1.2 – Data Center Infrastructure
- List the core elements of a data center including, applications, databases, servers/operating systems, networks (LAN/SAN), storage arrays.
- Describe how these core elements are integrated and show an application example.

Section 2 - Storage Systems Architecture

Module 2.1 – The Host Environment
- List the hardware and software components of the host environment such as CPU, Memory, bus, I/O devices, Registers, L1/L2 cache, file systems, volume management, Host Bus Adapters.

Module 2.2 – Connectivity
- Describe the physical and logical components of a connectivity environment.
- Define the key protocols and concepts including; Bus technology, PCI, IDE/ATA, SCSI, and Fibre Channel.

Module 2.3 – Physical Disks
- Describe the major physical components of a disk drive and functionality including platters, spindle, motor, actuator arm, R/W heads, assembly, drive controller.
- Define the physical structure of disk including sectors, tracks and cylinders.
- Define the access characteristics for disk drives and performance implications for factors such as rotation, positioning and transfer rates.
- Define the partitioning of physical drives using zoned bit recording.

Module 2.4 – RAID Arrays
- Define the concept of RAID for data protection.

- Review and understand the common RAID levels, including RAID 0, RAID 1, RAID 3, RAID 4, RAID 5, RAID 0+1, RAID 1+0, Hot Spares.

AN APPLIED LEARNING ACTIVITY- STUDENT COMPLETES A RAID CASE STUDY

Module 2.5 – Disk Storage Systems

- List the benefits of and components of an intelligent storage system.
- Compare and contrast integrated and modular storage systems.
- Explain how a storage system handles I/O flow.
- Describe the logical elements of an intelligent storage system.
- Define the Cache Structure and data flow through cache.

AN APPLIED LEARNING ACTIVITY– STUDENT COMPLETES A DATA FLOW EXERCISE

Section 3 - Introduction to Networked Storage

Module 3.1 – Direct Attached Storage

- Describe the benefits of a DAS based storage strategy and the physical elements in DAS.
- Define the connectivity options for DAS and distinguish between IDE, ATA and SCSI protocols.
- Describe the I/O flow in a DAS environment.

Module 3.2 - Network Attached Storage

- Define NAS and difference between NAS devices and general purpose file servers.
- Describe the NAS device components and file services protocols used (NFS and CIFS)connectivity options for NAS.
- Describe the I/O flow in a NAS environment.
- List NAS management considerations for performance, storage capacity, back up.
- Provide NAS examples and solutions to technology problems.

Module 3.3 – Storage Area Networks

- Provide an overview of the physical and logical elements of a SAN.
- Define SAN components including, Host Bus Adapter, Fiber Optic cabling, FC Switch, Disk Array, Management SW.
- Define Fibre Channel and connectivity layer including FC ports, World Wide Names, FC Addressing, FC log-in process and Fabric topologies.
- List SAN management considerations such as zoning and LUN masking, security, capacity and performance management.

AN APPLIED LEARNING ACTIVITY– STUDENT COMPLETES SAN CASE STUDY

Module 3.4 – IP SAN

- Define IP SAN topologies including iSCSI, FC/IP, iFCP.
- Describe the differences between FC/IP and iFCP.

Module 3.5 – Content Addressable Storage (CAS)

- Describe the benefits of a CAS based storage strategy.
- Provide an overview of the physical and logical elements of CAS.
- Define the connectivity options for CAS.
- Define the I/O flow in a CAS environment.

Section 4 – Information Availability

Module 4.1 – Business Continuity Overview

- List reasons for planned and unplanned outages and describe the impact of downtime. Differentiate between Business Continuity (BC) and Disaster Recovery (DR).
- Define Information Availability and its importance to the business.
- Define Recovery Time Objective, Recovery Point Objective.

Module 4.2 – Back Up and Recovery

- Define planning requirements for Back Up and Recovery.
- Define Back Up and Recovery Strategies.
- Define how a backup works, plus the business and data decisions for back ups.
- Describe LAN and SAN backup methods and Back Up for databases.

AN APPLIED LEARNING ACTIVITY – STUDENT COMPLETES A BACKUP & RECOVERY CASE STUDY

Module 4.3 – Business Continuity Local

- Describe potential areas of information vulnerability within a data center.
- List the local information availability technologies within the data center.
- Identify the appropriate local information availability technology based on criteria.

AN APPLIED LEARNING ACTIVITY – STUDENT COMPLETES REPLICATION CASE STUDIES 1 & 2

Module 4.4 – Business Continuity Remote

- Describe potential areas of information vulnerability between local and remote data centers.
- List the remote information availability technologies between local and remote data centers.
- Identify the appropriate remote information availability technology based on criteria.

AN APPLIED LEARNING ACTIVITY – STUDENT COMPLETES REMOTE REPLICATION CASE STUDY

Section 5 – Managing and Monitoring

Module 5.1 – Monitoring In the Data Center

- Define the areas to monitor for an information management infrastructure.
- List the issues for monitoring capacity, performance, security, servers, SAN, storage arrays and networks.

Section 6 – Security and Virtualization

Module 6.1 – Securing the Storage Infrastructure

- Define storage security.
- List the critical security attributes for information systems.
- Describe the elements of a shared storage model and security extensions.

- Define storage security domains.
- List and analyze the common threats in each domain.

Module 6.1 – Securing the Storage Infrastructure

- Identify different virtualization technologies.
- Describe block-level and file level virtualization technologies and processes.

VI. INSTRUCTION DESIGN METHOD OF STORAGE TECHNOLOGY FOUNDATIONS COURSE FROM EMC

The Storage Technology Foundations course was developed using a structured instructional design process, including a course design document, course objectives and topics. Each module in the course includes module objectives and individual lesson objectives. At the end of each lesson, the learner participates in "Apply Your Knowledge" activities to reinforce the concepts taught in the lesson.

The course does not require hardware or software labs for the applied learning component. The course does not attempt to teach students to become "hands-on" storage administrators. Instead, the course teaches storage design and architecture skills. This course design requirement is in adherence to input from industry professionals during the research phase of the project. Industry is looking for individuals with "big picture" knowledge of the five pillars of IT. Therefore, the applied learning for this course comes in the form of case study and discussion. Similar in approach to other IT courses such as Systems Analysis and Design, students get exposed to problems by applying the technology learned to solve a case study problem and explaining their technology choices and the overall solution. This method provides for the retention of learning using this case study approach. Current EMC Academic Alliance Program members supplement the course with additional case studies, research projects and home work assignments that they develop to further increase learning and retention.

VII. COMPARISON OF EMC STORAGE TECHNOLOGY FOUNDATION COURSE TO TEXTBOOKS AND OTHER STORAGE COURSES

There are a plethora of storage textbooks that have been written and are currently available (including an entry from the "For Dummies" series). A quick search on Amazon.com using storage networks or storage networking as the search criteria will yield an ample number of choices. In this author's opinion, the text books were developed for working IT professionals, not students. Authors such as Marc Farley, Tom Clark, Meeta Gupta and Daniel Pollack were reviewed and compared to the EMC Storage Technology Foundations course.

The Farley book (Storage Networking Fundamentals, Cisco Press) is a very good fundamentals book. It covers Fibre Channel, SCSI, ATA, and SATA and their use in network storage subsystems. The book also covers volume management, storage virtualization, data snapshots, mirroring, RAID, backup, and multipathing. It does not cover specific storage technologies such as CAS and does not cover intelligent disk arrays using caching algorithms and cache processes to the depth that the EMC course does.

The Gupta and Clark books are heavily focused on SAN technology. A large portion of these books are devoted to Fibre Channel, SCSI and iSCSI network protocols. The EMC course is comprehensive and covers other storage technologies such as NAS, CAS and replication.

There are relatively few other storage course offerings to compare against the EMC course. An offering from the Storage Networking Industry Association (SNIA) was reviewed. SNIA has developed a certification program (SNCP) for IT professionals who are interested in becoming storage certified. SNIA do not provide courses, just exam standards. SNIA has outlined the topics for which certification exams were written and are used to measure skills and knowledge. Training courses are developed by training firms using the standards developed by SNIA. A three day, lecture only, Storage Networking Concepts Foundation class is offered by Knowledge Transfer. The class is designed for working professionals. Like the textbooks, the course covers Fibre Channel, SCSI, iSCSI, SAN and NAS. This course appears to cover a breadth of topics but has no, or limited, applied learning.

VIII. CONCLUSION

EMC's Storage Technology Foundations course covers all the storage technologies used by today's IT departments. The course covers "the big picture" including SAN, NAS, CAS, Back Up and Recovery and array based replication. The course has been through an instructional design process and meets the requirements of pedagogical learning.

This course will provide college and universities an opportunity to meet the needs of industry by providing students with an education in the fifth pillar of IT; Storage. IT is changing and the demand to store, protect, manage, optimize and leverage the massive amounts of data being created is driving corporations to adopt information management infrastructures that utilize storage disk arrays as the backbone. Colleges and universities are urged to join the EMC Academic Alliance Program and use the Storage Technology Foundations course to teach students and prepare them for the emerging challenges of data and information management. The goal of this partnership program is to prepare tomorrow's Storage leaders today.

REFERENCES

[1] *CIO Magazine*, December 2006. http://peoplepolls.com/results/CIO/120706.asp?user=CIO and http://www.cio.com/info/releases/122906_techpoll.pdf

[2] "Gartner EXP Survey of More than 1,400 CIOs Shows CIOs Must Create Leverage to Remain Relevant to the Business" http://www.gartner.com/it/page.jsp?id=501189

[3] Hecker, Daniel, "Occupational employment projections to 2012," *Monthly Labor Review*, February 2004, pp. 80-105.

[4] Clancy, Tom (EMC), "EMC, IIT-B to Bridge Storage Knowledge Gap," *CyberMedia News*, March 14, 2007.

[5] Shrivastava, Alok (EMC), Managing Storage: Trends, Challenges and Options 2007-2008. EMC White Paper. May 2007. http://education.emc.com/main/common/documents/H2159_ManagingStorage_EP.pdf