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

Ed Van Sickle
vansickle_ed@emc.com
EMC Corporation
Franklin, MA 02038 USA

Abstract

Corporate CIOs 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 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. Increased spending on storage makes it the fastest growing segment of IT. EMC Corporation developed a Storage Technology course to teach students about the design of storage technologies. The course is "open" and focused on storage technologies, not products. Several colleges are using the course as an upper level elective offering. The course is taught by faculty. EMC provides knowledge transfer to faculty. There is no cost to join and no cost for the course. College and universities use the course to teach students about a very important topic in IT: Storage. 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. This paper will explain the program and the Storage Technology course.

Keywords: Computer Science Education, Industry - Education Relationships, Information Science Education, Storage Courses and Curriculum, Storage Technologies

I. INTRODUCTION

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. Furthermore, CIOs need to exploit new approaches to transform the business [2]. In an analysis of U.S. Department of Labor job data and a survey conducted by EMC, it is esti-mated that over 1 million new storage jobs will be created worldwide by 2012 [3].
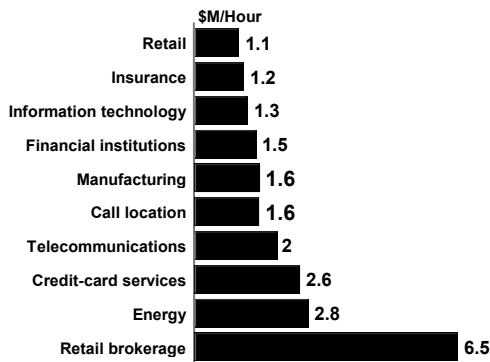
Despite all the news and spending data, for most IT professionals, it seems that storage infrastructures are still relatively unknown. This is interesting when considering that storage is not new and has been in existence for some time. Why then, is so much emphasis being placed on information infrastructure technologies now?

CIOs and IT Managers understand that their organizations most critical asset is data. Without, they can not process customer orders and conduct business as usual. They understand that more data is being created and they are trying to figure out ways to

contend with the tidal wave of information coming at them [4]. Regulatory requirements require them to keep data available on line for longer periods of time (can't archive as quickly as once done). The window of time to perform backups doesn't really exist anymore and many applications require 24x7 uptime. 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 these requirements is highly desired. Today's IT departments are implementing Information Management Infrastructures to meet those needs.

Cost in US Dollars per hour from downtime

	$M/Hour
Retail	1.1
Insurance	1.2
Information technology	1.3
Financial institutions	1.5
Manufacturing	1.6
Call location	1.6
Telecommunications	2
Credit-card services	2.6
Energy	2.8
Retail brokerage	6.5

Source: META Group

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, increased data delivery speed through fibre channel and IP networks and fiber optic cables that have enabled organizations to gain significant ad-

vantages in keeping data 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 [5] [6]. 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) [7]. 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.
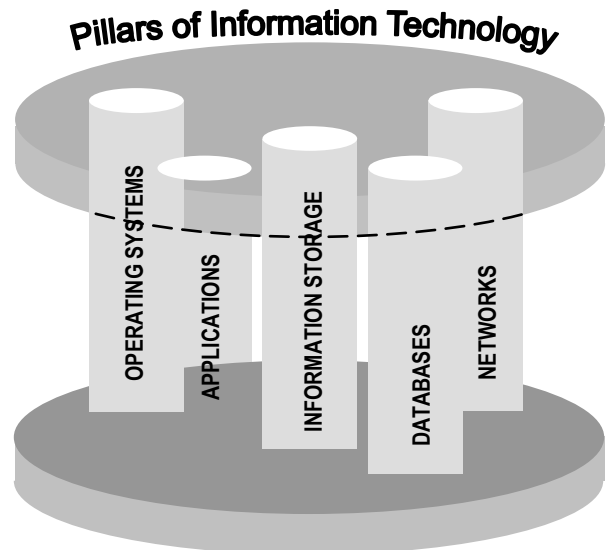
Pillars of Information Technology

Fig 2- Five Pillars of IT

Most colleges and universities educate students on four of the five Pillars of IT. Yet,

businesses and industry are looking for IT professionals, including students, who know all five pillars. Because the 5th Pillar is not being addressed, 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. How can colleges and universities close this gap and provide students with a needed education on the 5th Pillar of IT, Storage Technologies? A strong solution is the EMC Academic Alliance Program [8] [9].

II. EMC ACADEMIC ALLIANCE PROGRAM ORIGIN AND CONCEPT

Storage and information management infrastructures are being sold rapidly – it's a $60 billion (USD) market in 2007 and growing. This is creating high demand for people with these skills and knowledge (labor). But the supply of available people is low. EMC, facing this situation, created an Associates Program where recent college graduates are hired and trained for technical positions within EMC. A hiring requirement was that each new hire had a degree in either Computer Science or Information Technology. Since 2004, the EMC Associates Program has hired over 1,500 new employees, with over 90% being new CS or IT graduates. During the hiring process, it was noted that very few students had any knowledge or experience with storage technologies. It was also noted that each new hire had received an education in four out of the 5 Pillars of IT. In EMC's case, a "boot camp" was created to teach new hires on storage technologies and EMC products.

EMC's customers and partners wanted EMC's help with their hiring needs too. They wanted to know how they could hire personnel that knew about storage. A "boot camp" was not cost effective for them.

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

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 through a partnership program, many more people would become educated on storage than if EMC provided the education alone.

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.

III. EMC ACADEMIC ALLIANCE PROGRAM GOALS AND STRUCTURE

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 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 and no cost for the course.

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 list 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.

This is a global program with U.S., Russia, China and India as the prime locations for participating schools.

IV. EMC ACADEMIC ALLIANCE PROGRAM UNIVERSITY/COLLEGES IN THE UNITED STATES

As of September 2007, EMC has signed agreements with the following institutions in the United States. At this time, EMC is expanding the program and is willing to discuss the alliance program opportunity with other institutions.

Pennsylvania State University, University Park, PA. College of Information Science and Technology

Ball State University, Muncie, IN. Center for Information and Communication Sciences

University of Massachusetts at Dartmouth, Dartmouth, MA. Charlton College of Business, MIS Program

North Carolina A&T State University, Greensboro, NC. School of Technology, Computers, Electronics and Information Technology

North Carolina State University, Raleigh, NC. School of Engineering, Computer Science

Northeastern University, Boston, MA. School of Engineering Technology, Computer Engineering Technology

Indiana University of Pennsylvania, Indiana, PA. Eberly College of Business and Information Technology, MIS

Salisbury University, Salisbury, MD. Perdue School of Business, Information Systems

Springfield College, Springfield, MA. Department of Math, Physics and Computer Science

Quinnipiac University, Hamden, CT. College of Liberal Arts, Department of Computer Science

Southern New Hampshire University, Manchester, NH. Information Systems

Coleman College, San Diego, CA. Computer Networks

Kennesaw State University, Kennesaw, GA. Computer Science & Information Systems

Howard University, Washington, DC. School of Engineering & Computer Science, Department of Systems and Computer Science

St. Edward's University, Austin TX. School of Natural Sciences, Department of Computer Science.

LeTourneau University, Longview TX. Computer Science Department.

Quinsigamond Community College & Worcester Consortium (13 colleges in Worcester, MA area)

University of New Orleans, New Orleans, LA. College of Sciences, Department of Computer Science.

Northwest Missouri State University, Maryville, MO. College of Business, Computer Science / Information Systems.

Worcester State College, Worcester, MA. Computer Science Department.

Framingham State College, Framingham, MA. Computer Science Department.

The next sections of this paper describe the Storage Technology course and a comparison to other courses and textbooks.

V. STORAGE TECHNOLOGY COURSE FROM EMC

The Storage Technology course provides a comprehensive introduction to Data Storage technology fundamentals. 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 increased demand from businesses for highly available and secure access to data.

ii.) The Storage systems, 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), and 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, Disaster Recovery, Security and Virtualization.

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

Data creation, the amount of data and types of data being created

Challenges in data storage and data management

List the solutions available for data storage

Module 1.2 – Data Storage Solutions

Different media and available solutions to address data storage

Describe the role of each solution relative to data storage needs.

Define a Direct Attached Storage (DAS) environment.

Define a Storage Area Network (SAN) environment.

Define a Network Attached Storage (NAS) environment.

Module 1.3 – Data Center Infrastructure

Requirements for storage systems to optimally support the business.

Describe the challenges and activities in managing storage systems in a data center.

Section 2 - **Storage Systems Architecture**

Module 2.1 – The Host Environment

List the hardware and software components of the host environment.

Define the key protocols and concepts used by each component.

Module 2.2 – Connectivity

Describe the physical components of a connectivity environment.

Define the logical components of a connectivity environment.

Module 2.3 – Physical Disks

Describe the major physical components of a disk drive and functionality.

Define the logical constructs of a physical disk.

Define the access characteristics for disk drives and performance implications.

Define the logical partitioning of physical drives.

Module 2.4 – RAID Arrays

Define the concept of RAID and components

Review and understand the common RAID levels.

RAID 0, RAID 1, RAID 3, RAID 4, RAID 5, RAID 0+1, RAID 1+0

CASE STUDY - RAID

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.

Define how a storage system handles data flow.

Describe the logical elements of an intelligent storage system.

Cache Structure and data flow through cache, cache algorithms

DATA FLOW EXERCISE

Section 3 - **Introduction to Networked Storage**

Module 3.1 – Storage Networking Overview

Describe the evolution of networked storage.

Module 3.2 – Direct Attached Storage

Describe the benefits of a DAS based storage strategy.

Define the connectivity options for DAS and distinguish between IDE, ATA and SCSI protocols.

Describe the I/O flow in a DAS environment.

Module 3.3 - Network Attached Storage

Provide an overview of the physical and logical elements of a NAS.

Describe the connectivity options for NAS.

List common NAS topologies.

Compare and contrast connectivity devices.

Describe the I/O flow in a NAS environment.

List NAS management considerations given a particular environment.

Module 3.4 – Storage Area Networks

Provide an overview of the physical and logical elements of a SAN.

Describe the connectivity options for SAN.

List common SAN topologies.

Compare and contrast connectivity devices.

Overview the Fibre Channel log-in process.

Describe the I/O flow in a SAN environment.

List SAN management considerations given a particular environment.

SAN CASE STUDY

Module 3.5 – IP SAN

Describe the benefits of an IP based storage strategy.

Provide an overview of the physical and logical elements of an IP SAN.

Describe the connectivity options for IP SAN.

List common IP SAN topologies.

Module 3.6 – 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

Describe the impact of downtime

Differentiate between Business Continuity (BC) and Disaster Recovery (DR)

Define Information Availability and its importance to the business

Define RTO, RPO, and RGO

Module 4.2 – Back Up and Recovery

Planning for Back Up and Recovery, Back Up and Recovery Strategies

How a backup works, business and data decisions

Database backup methods, Back Up Topologies for LAN and SAN based backups,

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

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.

REMOTE REPLICATION CASE STUDY

Section 5 – Managing and Monitoring

Module 5.1 – Monitoring In the Data Center

Define areas to monitor.

Use an appropriate tool for data center management activity.

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 6.2 – Securing the Storage Infrastructure

Identify different virtualization technologies

Describe block-level virtualization technologies and processes

Describe file-level virtualization technologies and processes.

VI. INSTRUCTION DESIGN METHOD OF STORAGE TECHNOLOGY COURSE FROM EMC

The Storage Technology 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. Similar in approach to other IT courses such as Systems Analysis and Design, students get exposed to problems by applying the learned topic to solve a case study problem. Students also explain the technology choices and the overall solution for the case study. Retention of learning through case study is a prove 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 several storage textbooks that 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. All of these text books were developed for IT professionals working in the field. Authors such as Marc Farley, Tom Clark, Meeta Gupta and Daniel Pollack were reviewed and compared to the EMC Storage Technology 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 intelligent disk arrays using caching algorithms and cache processes that the EMC course covers.

The Gupta and Clark books are heavily focused on SAN technology. A large portion of the books are devoted to Fibre Channel, SCSI and iSCSI network protocols. The EMC course is more 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 does 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

The EMC Storage Technology 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 educa-

tion in the fifth pillar of IT; Storage. IT is evolving and the demand to store, protect, manage, optimize the massive amounts of data being created is driving corporations to implement information infrastructures that depend on storage disk arrays as the backbone. Colleges and universities are urged to join the EMC Academic Alliance Program and use the Storage Technology course to teach students for the emerging challenges of data and information management. The goal of this partnership program is to better prepare tomorrow's Storage leaders today.

IX. REFERENCES

[1] *CIO Magazine*, December 2006. http://peoplepolls.com/results/CIO/1 20706.asp?user=CIO and http://www.cio.com/info/releases/12 2906_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?i d=501189

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

[4] Gantz*, J.F. (2007).* The expanding digital universe: A forecast of world wide information growth through 2010. *IDC White Paper.*

[5] Morgenstern, D. (September, 2003). Missing from the Resume: SAN Higher Education*.* http://findarticles.com/p/articles/mi_ zdewk/is_200309/ai_ziff59296 (accessed September 11, 2007).

[6] Morgenstern, D. (October, 2003). Storage Education Still on Hold*.* http://findarticles.com/p/articles/mi_ zdewk/is_200310/ai_ziff108565 (accessed September 11, 2007).

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

[8] Courte, J., & Bishop-Clark, C. (2005). Strategies for making connections with industry: Creating connections: Bringing industry and education together. *Proceedings of the 6th Conference on Information technology Education SIGITE '05.*

[9] Prigge, G.W. (2005). University-industry partnerships: What do they mean to universities? *Industry & Higher Education, 19*(3).