Spring 2008

Designing and Implementing a Backup and Recovery System for Kentucky's Cooperative Extension Service

Wesley G. Justice
Regis University

Regis University
College for Professional Studies Graduate Programs
Final Project/Thesis

Disclaimer

Running head: KENTUCKY CES BACKUP & RECOVERY

Designing and Implementing a Backup and Recovery System for Kentucky's Cooperative Extension Service

Wesley G. Justice

Regis University

School for Professional Studies

Master of Science in Computer Information Technology

Regis University

School for Professional Studies Graduate Programs

MSCIT Program

Graduate Programs Final Project/Thesis

Certification of Authorship of Professional Project Work

Print Student's Name _____

Telephone _____ Email _____

Date of Submission _____ Degree Program **MSCIT**

Title of Submission _____

Advisor/Faculty Name _____

Certification of Authorship:

I hereby certify that I am the author of this document and that any assistance I received in its preparation is fully acknowledged and disclosed in the document. I have also cited all sources from which I obtained data, ideas or words that are copied directly or paraphrased in the document. Sources are properly credited according to accepted standards for professional publications. I also certify that this paper was prepared by me for the purpose of partial fulfillment of requirements for the Master of Science in Computer Information Technology Degree Program.

_____ _____

Student Signature *Date*

Regis University

School for Professional Studies Graduate Programs

MSCIT Program

Graduate Programs Final Project/Thesis

Authorization to Publish Student Work

I, _____, the undersigned student, in the Master of Science in Computer Information Technology Degree Program hereby authorize Regis University to publish through a Regis University owned and maintained web server, the document described below ("Work"). I acknowledge and understand that the Work will be freely available to all users of the World Wide Web under the condition that it can only be used for legitimate, non-commercial academic research and study. I understand that this restriction on use will be contained in a header note on the Regis University web site but will not be otherwise policed or enforced. I understand and acknowledge that under the Family Educational Rights and Privacy Act I have no obligation to release the Work to any party for any purpose. I am authorizing the release of the Work as a voluntary act without any coercion or restraint. On behalf of myself, my heirs, personal representatives and beneficiaries, I do hereby release Regis University, its officers, employees and agents from any claims, causes, causes of action, law suits, claims for injury, defamation, or other damage to me or my family arising out of or resulting from good faith compliance with the provisions of this authorization. This authorization shall be valid and in force until rescinded in writing.

Print Title of Document(s) to be published: _____

_____

_____

_____

_____ _____

Student Signature *Date*

Check if applicable:

_____ The Work contains private or proprietary information of the following parties and their attached permission is required as well: _____

Name of Organization and/or Authorized Personnel

Regis University

School for Professional Studies Graduate Programs

MSCIT Program

Graduate Programs Final Project/Thesis

Releasor Authorization to Publish Student Work WWW

I, _____ the undersigned, _____

Print Name of Company/Organization Representative *Representative's Title*

on behalf of _____ ("Releasor") do hereby authorize

Company/Organization Name

Regis University to publish through a Regis University owned and maintained web server, the document described below ("Work") and acknowledges that the Work contains personal or proprietary information of the Releasor. Releasor further acknowledges and understands that the Work will be freely available to all users of the World Wide Web under the condition that it can only be used for legitimate, non-commercial academic research and study but that this restriction on use will be contained in a header note on the Regis University web site but will not otherwise be policed or enforced. This authorization shall be valid and in force until rescinded in writing.

Print Student Name: _____

Title(s) of document(s) to be published: _____

_____

BY: _____ **DATE:** _____

Company/Organization Releasor Signature

Note: It is the student's responsibility to obtain the necessary release(s) prior to submitting the Final Project for publication. Please print your name and list all applicable documents.

Regis University

School for Professional Studies Graduate Programs

MSCIT Program

Graduate Programs Final Project/Thesis

Advisor/Professional Project Faculty Approval Form

Student's Name: _____ Program _____

PLEASE PRINT

Professional Project Title: _____

PLEASE PRINT

_____

Advisor Name _____

PLEASE PRINT

Project Faculty Name _____

PLEASE PRINT

Advisor/Faculty Declaration:

I have advised this student through the Professional Project Process and approve of the final document as acceptable to be submitted as fulfillment of partial completion of requirements for the MSCIT Degree Program.

Project Advisor Approval:

_____ _____

Original Signature *Date*

Degree Chair Approval if:

The student has received project approval from Faculty and has followed due process in the completion of the project and subsequent documentation.

_____ _____

Original Degree Chair/Designee Signature *Date*

Regis University

School for Professional Studies Graduate Programs

MSCIT Program

Graduate Programs Final Project/Thesis

Project Paper Revision / Change History Tracking

Version	Submitted To	Date	Changes
1	Erik Moore	08/30/2007	Initial submission
2	Erik Moore	10/04/2007	Added several sections; various corrections / adjustments per feedback; changed formatting to meet APA standards
3	Erik Moore	10/12/2007	Various minor adjustments; final draft

Acknowledgement

This effort is dedicated to my wife, Jackie, my daughter, Kennedy, and my parents, Greg and Adrianne Justice. All of them provided endless support and inspiration throughout the entire process.

Abstract

This project proposes the development and proof-of-concept implementation of a comprehensive backup and recovery plan for Kentucky's Cooperative Extension Service. Currently, no standardized backup system is in place. Each CES office location contains between five and forty Windows-based workstations and at least one server, and backup methods vary from office to office. Current backup processes are inadequate in several key areas. To ensure the availability and integrity of mission-critical data, the goal of this project is the analysis, design, and implementation of a standardized backup and recovery plan. The project will consider multiple hardware and software solutions (both commercial and open source), along with best practices for implementation and maintenance. A select number of offices will be chosen for implementation, and the project will be considered complete when a successful proof-of-concept has been established in these locations. A consistent, reliable backup solution, with both onsite and offsite components, will provide a much-needed safeguard to enterprise information and protect against costly data loss.

Table of Contents

List of Tables

List of Figures

Designing and Implementing a Backup and Recovery System for Kentucky's Cooperative Extension Service

Chapter 1: Introduction

Problem Statement

The Kentucky *Cooperative Extension Service* (CES) is an outreach and engagement organization based in the University of Kentucky's College of Agriculture, with regional offices located in each of Kentucky's 120 counties. As with many organizations, digital information plays a large and ever-increasing role in CES's business processes. Prior to the inception of the project described herein, CES lacked a standardized, comprehensive backup and recovery process, and risked losing critical data.

Data is always at risk, being constantly susceptible to hardware and software failures, theft, or unforeseen disasters. Likewise, human error presents a very significant risk: it accounts for an amazing 32% of data loss incidents, and is one of the primary reasons that an effective backup system is necessary (Ray, 2004). Data is one of an organization's most precious assets, the loss of which can bring devastating effects ("The three pillars of data," 2007).

A consistent, reliable backup solution, with both onsite and offsite components, will provide a much-needed safeguard to enterprise information and protect against costly data loss.

Review of Existing Situation (Prior to Project)

A typical CES office contains between five and forty Windows-based workstations, and at least one file server. The previous backup and recovery process was inadequate on numerous levels. Because there was no standardized plan in place, backup procedures varied widely from office to office. However, each office shared at least some of these common characteristics:

- No offsite backup component: One of the largest flaws in the existing system was the lack of an offsite backup component in virtually all CES offices. No backup and recovery plan – no matter how good at the local level – is complete without this critical element. Under these circumstances, any office that experienced a theft, natural disaster, or similar occurrence would face permanent data loss.
- Excessive user intervention required: Many offices employed a "manual" backup system,

wherein the user was expected to manually replicate their own data. Such circumstances can lead to a high probability of technical error and policy violation. In these cases, an automated system provided a far greater solution.

- Lack of hardware redundancy: The server hardware in most CES locations was non-redundant. Thus, even in situations where the server employed some form of automated backup procedure, recovery time was high when hardware failed. CES support personnel are centrally located at the University of Kentucky campus in Lexington. In the event of hardware failure – for example, a hard disk crash – users had to wait on the technician to travel to the CES office, physically replace the drive, reinstall and reconfigure the OS, and restore the backup. Fault-tolerance technology such as RAID could have turned the same hard disk crash into a virtual non-issue, from the perspective of the user.
- No access to previous file versions: In CES offices where a "manual" backup was used (as earlier described), or when backup software made a simple "mirror backup," users had little or no

access to prior file versions. When a user made an incorrect or unintended alteration to a file or files, and a few days passed before the mistake was realized, the last known "good" version of the file might have already been removed from the backup.

- Support issues: A help desk located on campus provides front-line IT support for all CES locations. Since backup procedures were different from office to office, help desk personnel were at a disadvantage, and problem resolution times were higher than necessary. Standardized backup and recovery procedures provided a solution, allowing for quicker troubleshooting and issue resolution.
- Lack of data integrity verification: Where inadequate software or "manual" backup systems were in place, backup media was not verified for integrity. Any backup system missing this verification presented a false sense of security, as data backups might have been unknowingly corrupt.

Goal to be Achieved

The goal of this project was the analysis, design, and proof-of-concept implementation of a data backup and recovery plan for Cooperative Extension Service offices. Multiple hardware and software solutions (both commercial and open source) were considered, along with best practices for implementation and maintenance. A select number of offices were chosen for implementation. The selected office locations will serve as "proof-of-concept" for future installations in other offices.

Barriers & Issues

- Budget: Financial constraints were a primary concern. Each CES office operates within a unique financial situation. The "Cooperative" part of "Cooperative Extension Service" indicates that federal, state, and local county governments cooperate to fund each office. However, the great majority of this funding is obtained at the local level. Each individual county government determines the level of funding to provide the local CES office. The end result is 120 offices with vastly different financial situations. Some are quite well funded, while others get by with a

shoestring budget. Thus, while CES offices are under the *administrative* control of the University of Kentucky, they are essentially under the *financial* control of their respective local governments. It was important to consider these budget issues when designing the backup and recovery solution. It was necessary for the final product to meet the financial requirements of all offices, including those not participating in the initial implementation.

- Time: The expected completion time for the project was five months, with an estimated completion date of August 10, 2007. Though no specific external factors specified this particular date, both management and the project manager recognized that the existing backup situation was very deficient, thus needed to be replaced as soon as possible.
- Support: As mentioned earlier, a centrally located help desk provides front-line IT support to CES offices. The backup solution was required to be designed such that help desk personnel are able to perform basic support and maintenance tasks, with minimal training.

- Technology Constraints: The project was authorized to consider new server hardware, but was required to function with existing user workstations. Thus, it was necessary for any potential client-side software to be compatible with the existing Windows/Intel-based machines.
- Business Requirements: The project was required to comply with any additional business requirements and/or constraints that were determined during the analysis phase.

Project Scope

The project focused exclusively on the stated goal of providing a comprehensive backup and recovery solution for the CES offices selected for implementation. This included the analysis, selection, and implementation of appropriate hardware and software, along with the identification of best practices for implementation, support, and maintenance. No other IT systems were included in the project's scope.

It is also important to note that the implementation phase involved *only* those offices that were selected during the analysis phase for the proof-of-concept implementation. However, the overall system design considered the

collective needs of all offices, in preparation for future installations.

Definition of Terms

Technical terms, or terms relating specifically to the University of Kentucky or the Kentucky Cooperative Extension Service, used in the project report include:

- CES: Acronym for Cooperative Extension Service.
- D2D: Acronym for disk-to-disk; a type of backup wherein data is backed up from one fixed disk to another (usually from a client workstation to a backup server).
- D2D2T: Acronym for disk-to-disk-to-tape; same as D2D (above), except that backed up data is subsequently archived to tape.
- Differential backup: A type of backup that occurs after a full backup; backs up all changes since the last full backup. Differential backups do not consider data copied during the last differential backup (if any). To restore from a differential backup, only the most recent full backup and the most recent differential backup are needed.

- Full backup: A type of backup that copies all specified data; does not rely on any previous backup, and is complete in and of itself. To restore from a full backup, only the most recent full backup is needed.
- Incremental backup: A type of backup that occurs after a full backup; backs up all changes since the last incremental backup (or since the last full backup, if no prior incremental backup has occurred). Incremental backups have the most complex restore procedure, as restoration requires the most recent full backup and *all* subsequent incremental backups.
- LTO-2: Acronym for Linear Tape Open 2; second generation of the LTO tape data storage technology. Also referred to as "Ultrium 2."
- Metadata: Data about data. Concerning backup technology, a given backup system's metadata would normally contain information regarding the backed up data; it essentially serves as an "index" to allow administrators to better handle relatively large, distributed data stores.
- Mirror backup: A type of backup wherein the

destination literally mirrors the source. New and modified data in the source is automatically added to the destination; data deleted from the source is also deleted from the destination.

- Open Source: Software licensed in such a way that allows the source code to be freely used, modified, or distributed.
- Proof-of-concept: An implementation of a given concept or idea, often on a relatively small scale, to demonstrate practicality and/or feasibility.
- RAID: Acronym for Redundant Array of Independent Disks; a term for a series of data storage technologies that split or replicate data among an array of hard drives. Used to increase performance and/or reliability.
- SDLC: Acronym for a systems development lifecycle; a framework that helps to ensure a project stays within scope, satisfies identified requirements, and meets its stated goals.
- SMB: Acronym for Small / Medium Business.
- Snapshot backup: A type of backup that provides a snapshot of a given disk (or disks) at a

specified point in time. Often, snapshot backups are merely full and incremental backups "under the service," but software logic allows them to appear as multiple, full backups.

- VPN: Acronym for Virtual Private Network. A private network (such as an internal local area network) that is "tunneled," via encryption technology, over another network (such as the public Internet).

Summary

This project involved the analysis, design, and proof-of-concept implementation of a data backup and recovery system for Kentucky's Cooperative Extension Service, to overcome the numerous problems associated with the backup system as it existed before the project.

Chapter 2: Review of Literature / Research

Overview of All Literature and Research on the Project

After an exhaustive search for information on backup-related projects having specifically taken place in other states' Cooperative Extension systems, the project manager found that the majority were facing a situation similar to that of the Kentucky CES (prior to the completion of the project). Some Extension Services did not have publicly available information regarding the topic. Of the ones that did, many exhibited trends that matched the situation in Kentucky: user-driven backups, no centralized administration, lack of redundancy, etc.

For example, the University of Arkansas, Division of Agriculture, recommends that CES personnel perform their own individual backups using the Windows Backup Utility, and provides a limited set of instructions for doing so (University of Arkansas, 2006). South Dakota State University also recommends that CES users handle their own backup needs, and lists a set of best practices. When discussing archival backup media, one instruction states, "If you require a full year's worth of data in your backup arsenal, use twenty-one sets of media; you'll have four

dailies, five weeklies, plus twelve monthlies" (South Dakota State University, 2007). While theoretically sound, instructions such as these are ambiguous at best and add an unnecessary workload for non-IT oriented users.

On the other end of the spectrum, the University of Nebraska-Lincoln appeared to have a reliable backup process in place for CES offices. The system, named *NSave*, is a university-wide resource utilizing Tivoli Storage Manger technology to back up workstations and servers to a secure, centralized location. Published information indicates that the system is effective, well documented, and well supported (University of Nebraska-Lincoln, 2007). However, because NSave was developed for the entire campus at the University level (not just CES offices, though CES offices appear to be welcome to participate), it is not an entirely appropriate model for a CES-only project such as the one being addressed here. In fact, the University of Kentucky does have a TSM-based backup resource available for on-campus workstations and servers. However, current policy restricts access to systems located on the UK wide area network; CES offices are not.

An additional resource that provided valuable insight into an external CES program's backup-related circumstances was a recent audit of Texas Cooperative Extension business

operations, performed by the Texas A&M University System Internal Audit Department. The audit's findings were published in a publicly available document, and included a section on current backup and information security procedures. Many of these findings were quite familiar when compared to the discoveries of this project own analysis, such as:

- "Research data is stored on employees' computers without systematic formal backup procedures. This elevates the risk of data loss in the case of a hard drive crash or data theft."
- "IT personnel are generally spread so thin that backup is performed irregularly."
- "Backup tapes are kept onsite with no offsite copies for insurance in the event of an unforeseen disaster."

(TAMU Internal Audit Department, 2004)

However, it is interesting to note that in response to these findings, the Texas CES still recommended a user-driven backup approach. Management specifically responded that "all units have been instructed that all relevant or sensitive data, including research data, that is stored on personal computers must be backed up on a systematic and regular basis; they have also been instructed to keep a

copy of the back up at a secure, off-site location" (TAMU Internal Audit Department, 2004). This project, while sharing very similar initial circumstances, will pursue a decidedly different solution.

Literature and Research that is Specific/Relevant to the Project

In contrast to the relatively small amount of backup-related research specifically pertaining to CES, there is a vast amount of literature published on backup technologies in general. The project manager consulted a variety of resources, including industry trade publications, technical magazines, books, and web-based material. When narrowing down these resources to those that were relevant to this project – i.e., concerning enterprise backup solutions for a wide user base – a few common themes arose. These included:

- Recent emergence of "snapshot" backups as an alternative to traditional full, incremental, and differential backup types: Snapshots record complete or partial system states at regular intervals, and essentially simulate an ongoing set of full backups (Kay, 2006).
- Continued importance of secure, offsite backups

for disaster recovery: While certainly not a new concept, recent literature continues to stress offsite backups as absolutely essential. Furthermore, offsite disaster recovery storage must meet the same data-security standards as the primary data store (Chernicoff, 2005).

- Importance of choosing a backup solution that fits the situation at hand: When considering the near limitless field of available technologies, care must be taken to choose a solution that integrates into the current technical environment, maintains regulatory compliance, and fits applicable requirements. For successful development of a backup and recovery strategy, it is key to ensure that the business requirements have been properly captured and properly valued; the analysis of these business requirements yields the technical requirements (Dow, 2004).
- Increasing popularity of disk-to-disk-to-tape (D2D2T) as a viable backup solution: A relatively recent innovation, D2D2T combines the speed of disk-based backups with the capacity and archival benefits of tape. The concept behind D2D2T is to back up from production disk to

> backup disk as quickly as possible; once this "D2D" has finished, files can be backed up or migrated to tape at a more leisurely pace (Gerber, 2004).

Summary of what is Known and Unknown about the Project Topic

As indicated above, there is a substantial amount of literature available on backup technology and practice, thus much is known about the project topic in general. However, also as previously indicated, very little information has been published regarding backup solutions in use in Cooperative Extension offices. This project attempts to explore the topic from that specific angle.

Contribution Project will Make to the Field

Based on discovered research, this project will be the first to publish a publicly available, in-depth report regarding the analysis, design, and implementation of a backup and recovery solution specifically for CES offices. Because every county in every state in the US has a CES office, the project's findings will be a valuable resource for those seeking to implement similar systems for CES offices in other states, or for other organizations with a

technical and logistical structure similar to that of CES.

Summary

An overview of available research revealed that when considering backup strategies, many states' CES programs are in a situation similar to that of Kentucky. While an abundance of information relating to general backup technology is available from a variety of sources, virtually no information was published on efforts by other universities to implement an enterprise-grade backup system specifically for CES offices. This project intends to contribute to the field by filling that void.

Chapter 3: Project Methodology

Research Methods Used

The project utilized various methods of research in order to gather information pertinent to backup practices and technologies. Such methods included online research, offline research, interviews, and project stakeholder meetings.

Online research served as a starting point for gathering data, and provided the bulk of the project's supporting background information. Numerous online resources were consulted. Because the amount of publicly available, Internet-based information regarding the topic is truly vast, it was necessary to narrow down the selection by vetting resources according to authority, practicality, and usefulness. A large variety of online resources were consulted, including the major search engines (*Google*, *Yahoo*, etc.), specialized backup-centric search engines (*SearchStorage.com*, *BackupCentral.com*, etc.), and magazine article / trade journal publication databases (*LexisNexis Academic*, *ACM Digital Library*, *Thomson's Computer Database*, etc.). The more informal resources, such as information found via search engines and

magazine articles, were used as a practical guide when completing the project's various phases. Meanwhile, scholarly research, found in trade publications and academic journals, was used to build the project's theoretical base, and to support the project's central concepts. Together, these online resources provided a virtually limitless source of up-to-date information.

A variety of offline, print-based resources were also consulted. These included physical trade publications (*NetworkWorld*, *ComputerWorld*), computing magazines (*Storage Magazine*, *Wired*), and books (Preston's *Backup & Recovery*). An additional form of "offline research" involved formal and informal meetings with colleagues, which often served as "brainstorming" sessions.

Systems Development Life Cycle

The project made use of the Systems Development Life Cycle model. The SDLC is a systems development framework that helps to ensure the project stays within scope, satisfies the identified requirements, and meets its stated goals. In particular, the project utilized the *waterfall* model of the SDLC, wherein the output of each project phase became the input for the next. Figure 1 illustrates the project's specific SDLC implementation.

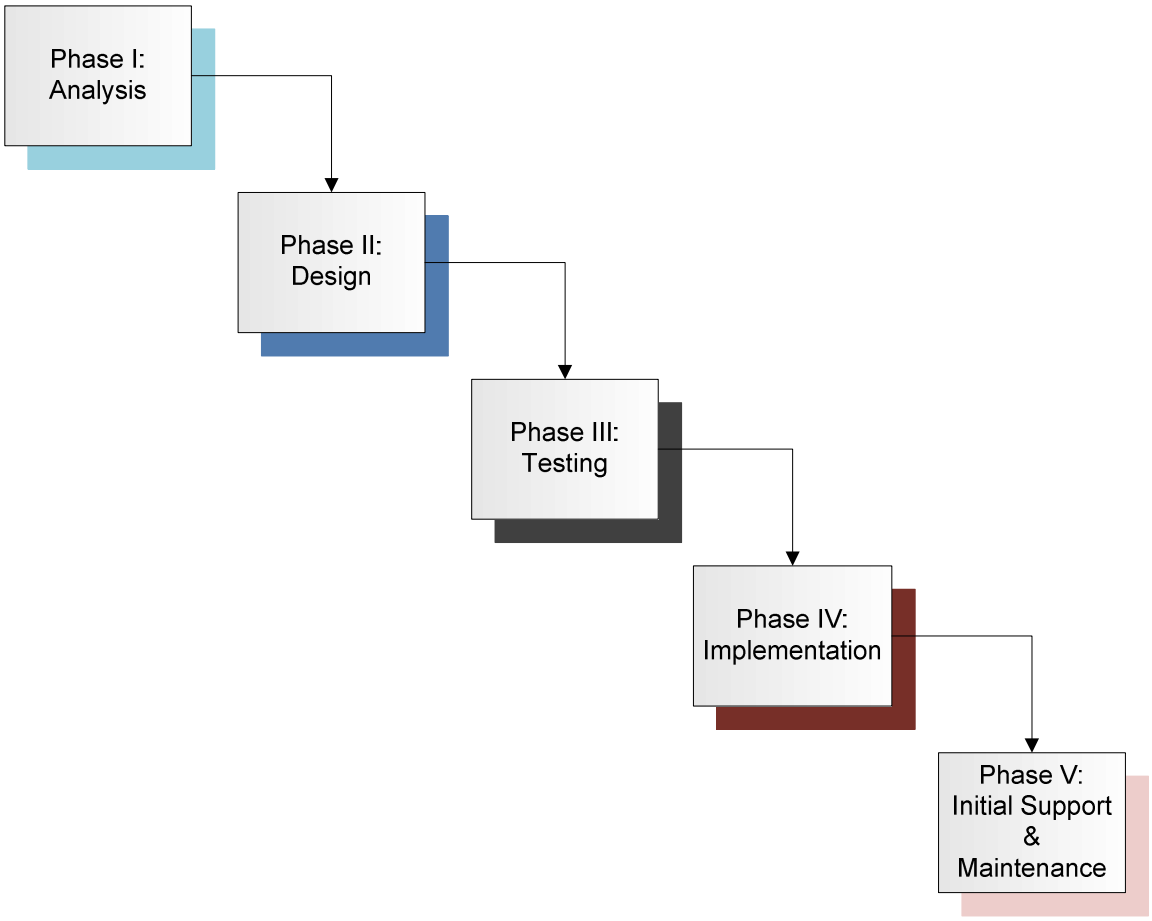


Figure 1: SDLC, waterfall method

The project's five phases, detailed in the following sections, included:

- Analysis
- Design
- Testing
- Implementation
- Initial Support & Maintenance

Phase I: Analysis

The first step in the analysis phase involved selecting three CES county offices, out of 120 total offices, to serve as participants in the proof-of-concept implementation. It was necessary to perform this step *prior* to detailed information gathering, as conducting a thorough analysis on all 120 offices was simply not practical and would exceed the scope of the project. Per meetings with project stakeholders, a number of criteria for identifying implementation locations were identified. These criteria included:

- Diversity in office size: The chosen offices should each represent a different relative size, both in number of employees and complexity of the local technical infrastructure. Ideally, relatively small, medium, and large-sized offices should be included, to provide an adequate representation of the state as a whole.
- Willingness to participate: The local employees should understand that the implementation is part of a proof-of-concept demonstration, and be willing to provide feedback that could be later useful to a large-scale, statewide implementation. (It should be noted, however,

that the gathering and application of such feedback is not covered within the scope of this project.)

- Available budget: Due to the unique financial situation of CES offices (see *Barriers & Issues*, above), it was necessary that each selected office have funds available for the purchase of any required hardware or software.

After the above criteria were identified, the final selection of offices was left to the project manager (pending approval from the offices themselves). The final selected offices were as follows:

- Carroll County Cooperative Extension Service. *Carrollton, KY*. One of Kentucky's smallest CES offices. Four local employees, including three county extension agents and one staff assistant. Rural area.
- Kenton County Cooperative Extension Service. *Covington, KY*. Mid-sized office. Twenty local employees, including county agents, agent assistants, and staff assistants. Moderately populated location just outside the Cincinnati metro area.
- Jefferson County Cooperative Extension Service.

Louisville, KY. Kentucky's largest CES office in terms of both staff and business volume. Over forty local employees, including numerous county agents, agent assistants, technicians, and staff assistants. Urban area.

After the above offices were selected, a detailed analysis of the existing situation was performed. Visits were made to the Carroll, Kenton, and Jefferson CES offices. During the visits, information was gathered using two primary methods:

- Interviews: Individual users were interviewed. Interviews were used to give the users an overview of the project's objectives, and – most importantly – to collect information from the users themselves. The interviews were performed by the project manager, and detailed notes were logged. Information gathered included:
  - Details on data and applications
  - Business requirements
  - Performance expectations
  - Budget / financial details
- Existing hardware / software / network examination: After the interviews, the project manager gathered information on existing

hardware. Notes were taken on the servers in each office, a selection of user workstations (representing users from each CES business area), and the local area network structure.

After information had been gathered from all three offices, the subsequent analysis produced the following information.

- Application data: While many of the applications in use have been converted to web applications in recent years (and are thus hosted on the UK campus and outside of the project scope), CES personnel continue to use a variety of locally-hosted business applications, each containing mission-critical data. Applications include:
  - Martech Youth Enrollment: *Youth Enrollment*, from Martech Systems, Inc., is a software application designed to track members and leaders in each county's numerous 4-H clubs. Features include interactive project, activity, and awards tracking, leader certification tracking, literature ordering and tracking, project lists, mailing labels, statistical reports, club reports, and activity reports (Martech Systems, 2007).

The application's data is stored in a single folder that can itself be stored in a variety of locations. Smaller CES offices often designate a single workstation to host the Youth Enrollment data, whereas larger offices almost always store the data on the central file server. The data is shared from the server or workstation, and accessed from clients via a mapped drive.

- UK SoilData: Used by the Agriculture and Horticulture departments of the Cooperative Extension Service, *SoilData* is an internally developed application for entering, analyzing, transferring, and archiving soil test information. Because it is a front end to a local Access database, SoilData's data is stored in a single Microsoft Access file. As was the case with the Youth Enrollment software, the data is often stored on the CES office's file server, but sometimes stored on a particular user's workstation. The server or workstation hosting the Access file accesses it through the SoilData application itself; clients accessing it

over the network must use a different application known as *SoilDataNet*.

- PATIM: *PATIM* (Pesticide Applicator Training Information Management) is an internally developed application used by CES offices to ensure that local private pesticide applicators maintain current training and licensing. It is a legacy, 16-bit application that has been in use at the University for some time. PATIM is a front end to a local FoxPro database, and unfortunately has no network capability. Only one workstation in each CES office is designated to run the PATIM software, and that workstation must host the data itself.
- NEERS: *NEERS* (Nutrition Education Evaluation and Reporting System), developed by the US Department of Agriculture, is used by CES's Expanded Food and Nutrition Education Program (EFNEP). EFNEP assists limited-resource audiences in acquiring the knowledge, skills, attitudes, and changed behavior necessary for nutritionally sound diets, and contributes to their personal

development and the improvement of the total family diet and nutritional well-being (United States Department of Agriculture, 2007). The NEERS software was designed to facilitate tracking and reporting on the program's efforts at the local level. Similar to the Youth Enrollment software, the application's data is stored in a single folder that can be stored in numerous locations. It is often stored on the office's local file server, though is sometimes maintained on a designated workstation.

- Mailroom Toolkit: Satori Software's *Mailroom Toolkit* is a series of COM and .NET-based controls that provide address quality and mailing features to CES offices' local mailing list databases (Satori Software, 2007). It performs single address verification, batch processing for multiple addresses, presorting options for bulk mailing operations, and label generation and printing. Mailroom Toolkit is essentially a plug-in for Microsoft Access – thus, similar

to SoilData, the data is stored in a single Access file, located either on the file server or on a designated workstation.

- User data: Beyond CES-specific application data, user workstations contained numerous instances of other data, including:
    - Office documents: All CES employees currently use Microsoft Office 2003. User workstations contain an abundance of Office documents, including files created with Word, Excel, Access, Publisher, and PowerPoint.
    - E-mail: Outlook 2003 serves as the current "official" CES e-mail client. While a few employees choose to use Outlook Web Access to access the University's Exchange server (and thus have no locally stored e-mail), most have e-mail archives, contacts, distribution lists, calendar data, and notes stored in Outlook PST files.
    - Browser favorites: Internet Explorer or Firefox
    - Media including:
        - Photos

        - Audio
        - Video
    - Financial data: Quicken or QuickBooks
    - Miscellaneous items: Other data stored in various locations (such as the Desktop, various folders under C:\, etc.)
- Existing backup methods: Existing methods for safeguarding data varied widely from user to user. Methods included:
    - No backup system at all: Unfortunately, this "method" was discovered to be far too common.
    - Manual backups to various media (CDR, DVDR, flash drives, external hard drives, network shares, etc.): This method was the second-most commonly used. While better than nothing at all, there were numerous negative aspects of users manually backing up their own data. The process was not automated, thus it was time-consuming and required the user to remember to perform the backup. It did not provide for data integrity verification. It required the user to be at least somewhat technically knowledgeable.

And, in most cases where users performed their own backups, a significant amount of data was overlooked. For example, the vast majority of users that employed this process were unaware of the location of Outlook's PST files.

- Backup4All software: Some years ago, the University purchased a statewide volume license for Backup4All, a simple backup application that is seemingly aimed at the home PC market. The analysis revealed that the software was still in use on some CES workstations, many of them at the Jefferson County office. Even before the start of the project, it was the opinion of the project manager, management, and users that Backup4All was, at best, minimally useful. It had developed a somewhat notorious reputation for constant crashing, failure to perform scheduled tasks, and botched recovery attempts. The software also contained no server-side component, and had to be administered individually at each workstation; thus centralized management was

not possible.

- Workstation hardware: As was expected, user workstation hardware varied widely. Each office contained one Dell workstation per user; overall, approximately 75% were Optiplex models, while the remaining 25% were from Dell's Dimension, Inspiron, and Latitude lines. Table 1 summarizes the findings of the analysis on workstation hardware in each CES office.

CES Office	Make	Model(s)	OS	Age
Carroll	Dell	Optiplex (various): 3 Latitude D820: 1	Windows XP, SP2	0 – 2 years
Kenton	Dell	Optiplex (various): 16 Latitude D620: 2 Latitude D820: 1 Dimension 4400: 1	Windows XP, SP2	0 – 3 years
Jefferson	Dell	Optiplex (various): 32 Latitude D420: 1 Latitude D620: 6 Latitude D820: 2 Dimension 2400: 1 Dimension 4400: 1	Windows XP, SP2	0 – 3 years

Table 1: Workstation hardware summary: Carroll, Jefferson, Kenton CES offices

- Server hardware: Each office contained one server, used for file and print services. As was the case with workstation hardware, server models varied between offices, according to the local office's budget and needs. Findings are summarized in Table 2.

CES Office	Make	Model	OS	Age
Carroll	Dell	PowerEdge SC400	Windows Server 2003, SP2	2 years
Kenton	Dell	PowerEdge SC400	Windows Server 2003, SP2	2 years
Jefferson	Dell	PowerEdge 1800	Windows Server 2003, SP2	3 years

Table 2: Server hardware summary: Carroll, Jefferson, Kenton CES offices

- Network structure: Network architectures in each office were relatively simple. In the Carroll and Kenton offices, all workstations and the server were wired to a single Linksys 10/100 Mbps unmanaged switch. The Jefferson CES office is spread out over two floors. Each floor has its own Linksys Gigabit unmanaged switch; the two

switches are connected via fiber. All offices make use of a single Linksys router that is connected to a local broadband ISP. The Kenton and Jefferson offices use a DSL connection, whereas the Carroll office subscribes to a Cable ISP. CES offices are not part of the UK campus wide area network; data transactions to and from campus make use of the public Internet. In instances where a CES workstation must be connected to the UK WAN, VPN client software is used. Figure 2 illustrates a typical CES office network structure.

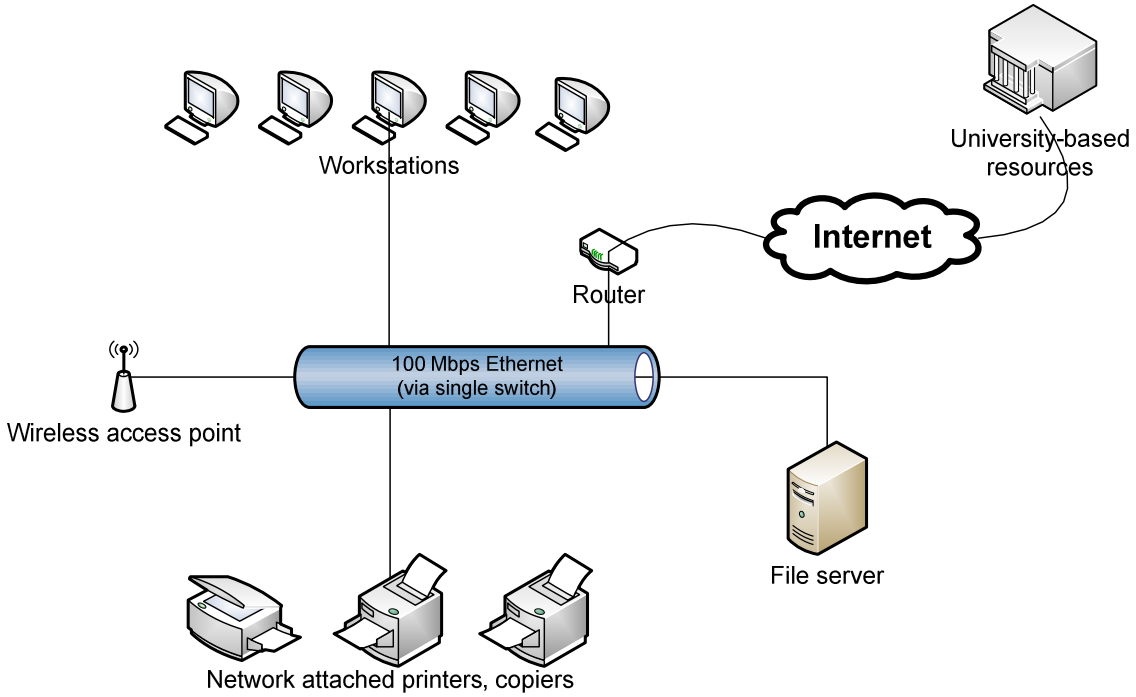


Figure 2: Typical CES office network structure

The final stage of the analysis phase was requirements gathering, considered by the project manager and stakeholders to be one of the project's most crucial processes. Because the project's ultimate goal – a comprehensive, reliable backup system for CES offices – involves a system that should be relatively transparent to end-users, it was important to differentiate between business and technical requirements.

Business requirements originated primarily from user input. Because of the nature of the project, business requirements were relatively few. Different types of projects – for example, development of a software application, website, or similar system – often run into the issue of "feature creep" as new requirements and features are continually added. However, when seeking requirements for this particular project, it became readily apparent that a common theme was "It should just work." As such, both business and technical requirements reflected a desire for the finished project to be efficient and transparent to the users.

After numerous interviews with potential users, the identified business requirements were compiled and placed into a business requirements document for review by all project stakeholders. Key business requirements included:

- Reliability: The newly implemented backup system must be consistently reliable. As data loss can happen at any time, the system must be available at a moment's notice.
- Transparency: The system should be transparent to end users, and should require absolutely no user intervention to perform scheduled backup tasks. It should not interrupt users' workflow. Users should not need to think about the backup system until a data loss situation occurs.
- Speed: To match the ever-increasing pace of business, backup and recovery operations should be relatively fast. For typical recovery scenarios (for example, single file restoration), end users should not need to wait on the physical presence of their designated IT support person; rather, recovery should be accomplished with a quick call to the Computing & IT Helpdesk.
- Disaster-readiness: The system must guard against localized disasters. Data must be regularly duplicated and stored in a secure, offsite location.

Technical requirements were derived from the analysis findings, as well as the business requirements themselves.

Key technical requirements included:

- New hardware (Dell): Due to the "preferred vendor" contract in place between Dell and the University of Kentucky, all newly purchased server and workstation hardware must be acquired from Dell. If non-Dell hardware is to be purchased, the project manager must prepare a written justification stating the reasons why equivalent Dell hardware will not meet the project's needs.
- Compatibility with existing hardware: The system must be interoperable with existing workstation hardware.
- Redundancy: The system must provide a level of hardware redundancy to safeguard against hardware failures.
- Automation: Routine backup tasks should be completely automated, requiring no human intervention.
- Uniformity: To simplify logistical and support issues, as well as to prepare for a future statewide implementation, the system should be as "uniform" as possible across CES offices. While

different budgets and technical needs might necessitate that one implemented system might not exactly match another, variations in purchased hardware and software should not be extreme.

Phase II: Design

The goal of the design phase was the actual design of the backup system, including identification of potential technologies (both hardware and software), and eventual selection of the technologies that provided the best solution to meet the project's goals.

Upon beginning the design phase, an immediate concern of the project manager was "information overload." It quickly became apparent that there is a virtually limitless amount of backup solutions available, and that evaluating all of them would be quite impractical. Thus, when considering hardware and software, it was first necessary to "limit the field" to a finite number of potential solutions. Research demonstrated that organizations of similar size, geographic distribution, and technical structure used some common criteria when deciding on an initial list of software candidates (Hope, 2005). Based on these, a number of criteria were developed for the software "vetting" process:

- System requirements, as identified in analysis phase
- Budget constraints
- Initial research (see *Research Methods Used*)
- Input from management and other colleagues
- Product reputation (solutions that were generally well-regarded within the industry took prevalence over those that were lesser known)

When considering hardware, a significant limiting factor was the University's "preferred vendor" contract with Dell. The contract requires that UK's workstation and server hardware be purchased from Dell.

Server candidates, selected from Dell's PowerEdge Performance Tower series, included the PowerEdge 840, PowerEdge 1900, and PowerEdge 2900 models. Table 3 summarizes each candidate's features and technical specifications.

Model	PowerEdge 840	PowerEdge 1900	PowerEdge 2900
Description	Entry-level 1S tower server	Entry-level 2S tower server	Performance Tower
Form factor	Tower	Tower	Tower
Benefits	Affordable server with advanced hardware and systems management features	Delivers performance, scalability, and manageability at a value price	Delivers high performance, scalability and availability for departmental applications
CPU(s)	Single dual-core Intel Xeon CPU, Intel Pentium D CPU or Intel Celeron D CPU	Up to two 64-bit quad-core Intel Xeon CPUs	Up to two 64-bit quad-core Intel Xeon CPUs
Memory	512MB – 8GB ECC DDR2 533/667 SDRAM	256MB – 16GB; fully buffered DIMMs	256MB – 48GB; fully buffered DIMMs
PCI slots	Five total: two PCI Express, two 64-bit PCI-X, one 32-bit PCI	Six total: four PCI Express, two 64-bit PCI-X	Six total: four PCI Express, two 64-bit PCI-X
Integrated controllers	Embedded four-channel SATA, optional SAS	Embedded two-channel SAS/SATA, optional 4-port SAS/SATA, optional SCSI (for tape)	PERC 5/I RAID or SAS 5/I (SAS or SATA support)
RAID controller	PERC 5/I PERC 5/E SAS 5/I R	PERC 5/I PERC 5/E SAS 5/I R	PERC 5/I PERC 5/E PERC 4e/DC
Integrated NIC	Single-embedded Broadcom Gigabit NIC	Single-embedded Broadcom Gigabit NIC	Dual-embedded Broadcom Gigabit NICs
Maximum internal storage	SAS: 1.2TB SATA: 2TB	SAS: 1.8TB SATA: 4.5TB	SAS: 3TB SATA: 7.5TB
External storage	SAS storage systems	SCSI and Fibre Channel storage systems	SAS, SCSI, and Fibre Channel storage systems
Availability features	Highly serviceable tool-less chassis; ECC memory; hot-put SAS and SATA drives; options hardware SATA	ECC memory; Single Device Data Correction (SDDC); optional PERC with battery-	ECC memory, SDDC, Spare Bank; hot-plug SAS/SATA hard drives; optional hot-

	RAID; OpenManage Systems Management Support	backed 256MB DDR cache; tool-less chassis; cluster support; full OpenManage Systems Management Suite Support; validated for Dell/EMC SANs	plug redundant power; hot-plug redundant cooling; tool-less chassis; high availability Dell/EMC Fibre Channel and PowerVault SCSI cluster support

Table 3: Server hardware candidates

(Dell, 2007)

After considering server hardware candidates and consulting with management, the decision was made to choose the PowerEdge 840 server for small-to-mid-sized offices (represented in the project by the Carroll County CES office), and the PowerEdge 1900 server for mid-to-large-sized offices (represented in the project by the Kenton and Jefferson County CES offices). Price and available features were primary factors in the choice. In relation to the project's technical requirements, the processing power, memory, storage space, and additional features of the PowerEdge 840 and 1900 models made them the most reasonable choices when considering price and available budgets. While the PowerEdge 2900 certainly would have been a more than adequate choice, its relative high price

and technical specifications (many of which could have been considered "overkill" for a project of this scale) did not make it a practical contender.

Prior to acquisition, the technical specifications of the servers were customized as follows:

- PowerEdge 840:

Model	Dell PowerEdge 840 Performance Tower
Description	Lower-end server for relatively smaller CES offices (Carroll)
Form factor	Tower
OS	Windows Server 2003, Standard Edition Academic, SP2
CPU	Dual Core Intel Pentium E2160, 1.8GHz
Memory	2GB DDR2,667MHz (2x1GB) Dual Ranked DIMMs
Storage	146GB 10K RPM Serial-Attach SCSI drives (4), 586GB total storage
RAID controller	PERC 5/I
Tape Drive	PowerVault 110T, LTO2-L Tape Backup, 200/400GB, Internal
Integrated NIC	Single-embedded Broadcom Gigabit NIC

Table 4: PowerEdge 840: Key specifications as configured

- PowerEdge 1900:

Model	Dell PowerEdge 1900 Performance Tower
Description	Higher-end server for mid- to large-sized CES offices (Jefferson, Kenton)
Form factor	Tower
OS	Windows Server 2003, Standard Edition Academic, SP2
CPUs	Quad Core Intel Xeon E5310 (2)
Memory	4GB 667MHz (4x1GB), Dual Ranked Fully Buffered DIMMs
Storage	300GB 10K RPM Serial-Attach SCSI drives (4), 1.2TB total storage
RAID controller	PERC 5/I
Tape Drive	PowerVault 110T, LTO2-L Tape Backup, 200/400GB, Internal
Integrated NIC	Single-embedded Broadcom Gigabit NIC

Table 5: PowerEdge 1900: Key specifications as configured

Analysis and selection of software candidates was an entirely different process. There were no vendor constraints, thus virtually all available backup software qualified as an initial candidate. Therefore, as was mentioned earlier, it was necessary to limit the field to a pre-selected group of candidates, and focus evaluation on those. Software candidates identified using the criteria outlined above are summarized in Table 6.

PRODUCT	DEVELOPER	DESCRIPTION	LICENSING
Backup Exec for Windows Servers	Symantec	Formerly from VERITAS (now purchased by Symantec), Backup Exec is the company's flagship backup product.	$928.65 (one server, unlimited clients)
Retrospect Single Server	EMC Insignia	D2D2T and snapshot-focused software aimed at SMBs	$500 (one server, unlimited clients)
Data Protector Express	HP	SMB edition of HP's enterprise-class backup solution	$779 (one server, unlimited clients)
Tivoli Storage Manager Express	IBM	SMB edition of IBM's Tivoli Storage Manager product	Varies, depending on number of clients and processor value units (PVUs)
BackupPC	Open source	Enterprise-class, open source, server-based backup system for D2D backups. (No tape / archival component).	n/a
Duplicity	Open source	Client-based, open source backup application utilizing rsync algorithm	n/a
Rsnapshot	Open source	Client-based, open source backup application. Uses rsync and snapshot technology to create virtual "full" backups.	n/a

Table 6: Software candidates

After careful consideration of the above solutions, the project manager and stakeholders agreed that EMC Insignia's Retrospect provided the best fit for the project's need. Primary reasons for the choice included price (an academic license was available for $500 per server, and covered *all* clients, regardless of the number of employees in the office), scalability (the software appeared well-suited for all CES offices sizes, whether there were four employees or forty), and support (management was impressed with EMC's support offerings, and was happy to see that product updates were issued on a regular basis). Selected features of the software are demonstrated in the screenshots in Appendix A.

Another important stage of the design phase was development of an agreed-upon set of "best practices" for data backup. These best practices were researched and developed by the project manager, and reviewed and approved by management. See Appendix B, *Best Practices*, for a detailed listing.

Finally, it was necessary to develop a maintenance plan that contained guidelines for ongoing support and maintenance of the implemented system. The maintenance plan can be found in Appendix B, *Maintenance Plan*.

Phase III: Testing

Once the analysis phase was complete, the analysis results were thoroughly tested. Though the analysis phase provided a good deal of information, it was very important to see the proposed hardware and software solutions at work in a "real world" environment prior to actual implementation.

To facilitate testing of the designed system, a "test lab" was created. The lab contained a technical architecture similar to that of a typical CES office, in addition to hardware and software that had been selected during the design phase.

Test lab hardware included four Dell Optiplex 745 workstations, and one Dell PowerEdge 1900 server.

Test Preparation

In order to create a true representation of a CES office, a number of CES business applications were loaded onto the test workstations. In addition, these applications were loaded with a set of sample data provided by the Jefferson County CES office. These applications, described previously, are illustrated in Table 7.

APPLICATION	DATA VOLUME / DISTRIBUTION	PRIVACY ISSUES	RECOVERY REQUIREMENTS
Martech Youth Enrollment	Up to 1 GB, stored on either workstation or server	Contains private personal information (SSNs, contact information) on clients	Ideally, downtime should be less than one day
SoilData	Up to 500 MB, stored on either workstation or server	Contains private information	Availability is less crucial than Youth Enrollment, yet prolonged downtime is still unacceptable
PATIM	Up to 100 MB, stored on workstation	Contains private information	Used to serve walk-in clients; downtime must be minimal

Table 7: Test applications & sample data

In addition to these specialized business applications, the workstations were loaded with software typically used by CES employees, including Office 2007 (all components), Internet Explorer, and QuickBooks. A set of sample data was loaded for these general applications, including:

- Assorted Office 2007 documents, placed into the user's "My Documents" folder as well as other locations on the local drive

- Outlook PST files (primary and archive)
- Internet Explorer favorites
- Miscellaneous desktop items
- Miscellaneous media files in various locations (pictures, video, music, etc.)

Due to the impracticality of purchasing a separate server merely for testing, the test server used was the actual Dell PowerEdge 1900 purchased by the Jefferson County CES office. In preparation for testing, the system was loaded with Windows Server 2003, patched and updated, and configured as a file server (including the loading of several types of sample data, similar to that described above). Finally, the Retrospect software was installed on the server, and configured according to identified best practices.

Testing Process

During the first week of testing, no data loss scenarios were performed. The server and workstations were allowed to run as normal, with various updates to the sample data being performed on a daily basis. The workstations were backed up to the server once daily; the server received an initial full "offsite backup" and was subsequently backed up to tape according to the identified

best practices (see Appendix B).

After the initial week of typical operation, several data loss and recovery scenarios were performed:

- Scenario 1 – Server drive failure: The first test scenario simulated a failure of one of the server's internal drives. To simulate the drive failure, the power source to a single, randomly selected drive was disconnected while the server was running.
    - Results: The server's RAID 5 implementation allowed system operations to continue with no downtime. System performance experienced a mild decrease as the designated hot spare drive was automatically rebuilt with the contents of the failed drive. When rebuilding was complete, the hot spare drive took the place of failed drive in the RAID 5 array. In the event of an actual drive failure, IT support personnel would visit the CES office after working hours to install a replacement drive and designate it as the new hot spare for the array.
- Scenario 2 – Workstation drive failure: The next test scenario simulated drive failure in a user

workstation. To simulate the failure, the Optiplex 745 test machine was powered down, and the drive data cable was physically disconnected from the system board.

- Results: As existing CES workstations do not have the redundancy features of the server hardware, moderate downtime was inevitable. The workstation was unavailable as a replacement drive was installed, the appropriate software image was applied and customized, the Retrospect client software was installed, and the user data was restored from the Retrospect server. In a real-world situation, the project manager estimates that such a failure would necessitate from 1-24 hours of downtime, depending on external variables such as the availability of the replacement drive hardware, as well as the availability of IT support personnel to perform the drive replacement and data restoration. While up to 24 hours of downtime is not desirable, the results of this test nonetheless represent a huge improvement over previous

conditions. For example, if the workstation previously used no backup method at all, the data would be permanently lost.

- Scenario 3: Workstation data loss / corruption: The third test scenario simulated loss or corruption of specific data on a user workstation, rather than loss of an entire drive. The test involved several "sub-tests," in which specific application data was intentionally deleted from the test machine. These included:
  - Youth Enrollment: The test system's locally-hosted sample data for the Martech Youth Enrollment application was intentionally removed.
  - SoilData: Sample data for the SoilData application was intentionally removed.
  - PATIM: Sample data for the PATIM application was intentionally removed.
  - User data: Selected files from the test user's Documents folder were removed.
  - Results: All of the above data loss scenarios were successfully corrected by restoring the affected data from the Retrospect server to the test machine. In a

real-world environment, the restoration could be performed by 1st-tier help desk personnel, resulting in minimal downtime.

- Scenario 5 – Disaster: The final test simulated a disaster in a CES office, such as theft, fire, or natural disaster. To accomplish the test, the test server was simply unplugged and set aside, as such an occurrence would result in complete loss of the server hardware. In many disaster scenarios, workstation hardware would also be lost.
    - Results: A relatively long downtime is required as the server (and workstations, if necessary) are replaced and imaged and customized with appropriate software, and data is restored. Assuming that local tape backups were lost in the disaster, the latest offsite backup is used to restore the server data. The project manager estimates that such a disaster would result in downtime lasting from one to several days, or possibly longer, again depending on specific circumstances and external factors such as the availability of replacement

hardware and support personnel.

Phase IV: Implementation

The goal of the implementation phase was to be as seamless and transparent to the end users as the finished product itself. Implementation in each of the three selected CES occurred on a Saturday, thus occurring outside of working hours and preventing any disturbance of workflow.

Implementation at each office began at the workstation level. The Retrospect client software was installed and configured at each workstation. (Very little configuration was required at this level; the majority of Retrospect's client configuration options are handled via the server.) Following workstation software installation and configuration, data was copied from the existing server, and settings (such as share names, file permissions, user accounts, and network configuration information) were carefully recorded. The existing server was physically removed, and the new server was installed and configured using the previously recorded settings. Because these settings remained the same on the new equipment, the transition to a new file server was essentially seamless. Mapped drives and file shortcuts on workstations operated

just as before.

After configuration of file server operations was complete, the Retrospect software, having been preconfigured with scheduled backup tasks according to identified best practices, was launched. Each client was added to the server and placed into a designated "Backup Clients" container, so that client operations could be performed on all workstations as a group, rather than individually. An initial full backup was performed to tape, to serve as the first offsite backup. An LTO-2 cartridge was left in the server's tape drive to prepare for nightly server backups. Finally, initial client disk-to-disk backups were performed.

Implementation in each office went smoothly and as expected, with users noticing no apparent changes in the client/server environment (with the exception of a considerable increase in space available on the file server).

Phase V: Support & Maintenance

The final phase encompassed the first three weeks of support and maintenance for the newly implemented system. (Of course, support and maintenance will continue indefinitely, but this initial support phase was identified

to maintain a clearly defined project scope). During this phase, server software (including Retrospect as well as the operating system itself) was checked and updated on a weekly basis. Beyond routine updates, backup sets were checked to verify integrity, and occurrences of data loss were reported to and handled by the project manager.

During these initial weeks, no server or workstation hardware failures were experienced. Several data loss instances occurred, including three in the Jefferson County office, two in the Kenton County office, and one in the Carroll County office. Of these five, four involved data corruption due to user error, and one involved accidental deletion of a file by a user. In each case, after the project manager was notified by the CES office, data restoration was performed quickly (via remote access to the server), and downtime was minimal. Users were notably pleased with the newly implemented system, as compared to the various methods previously in use.

Specific Procedures

Progress Tracking

The project plan was designed, maintained, and tracked using Microsoft Project software. The project plan is available in Appendix C. In addition, detailed notes were

maintained during each phase by the project manager, to assist with ongoing system maintenance as well as the preparation of this report.

Progress reporting was handled via bi-weekly meetings with selected project stakeholders, hosted by the project manager. The length and formality of these meetings varied, depending on the project phase and amount of information to be reported. E-mail updates were utilized when it was necessary to report important information between bi-weekly meetings.

Management approval was required after each major milestone, prior to continuing the project. These "major" milestones were identified as the completion of each of the project's five phases. Approval was given during informal meetings between management and the project manager, called on an as-needed basis. After completion of the last phase (Initial Support & Maintenance), a final approval was requested and granted, signifying the overall project's completion and success.

Change Management Procedure

Changes to the project plan were to be described in a written summary, and required stakeholder review and management approval.

Formats for Presenting Results / Deliverables

Project deliverables were delivered via e-mail and/or printed documents, as required.

Review of Deliverables

Project deliverables for included the following:

- Phase I: Analysis
    - Feasibility analysis
    - Requirements summary
- Phase II: Design
    - Design summary
    - Best practices document
    - Network diagram
    - Maintenance plan
- Phase III: Testing
    - Test plan
    - Test results summary
- Phase IV: Implementation
    - Implemented system, per design specifications

Outcomes

The final project outcome was considered a success by the project manager and stakeholders. After implementing the system and monitoring its progress during the initial support and maintenance phase, it became clear that the project's original goal – the analysis, design, and implementation of a standardized backup and recovery plan – was successful. A case study demonstrating the project's effectiveness can be found in Appendix E.

Summary

In order to organize and manage the project, the waterfall method of the Systems Development Life Cycle was utilized. Project phases included Analysis, Design, Testing, Implementation, and Initial Support & Maintenance.

Chapter 4: Project History

How the Project Began

For some time after acquiring his position with the University of Kentucky, the project manager was concerned with the state of backup technology in CES offices. Backup solutions, if they existed at all, were outdated, non-standardized, and – in many cases – simply didn't function as needed. Thus, implementing an overhaul to the backup system had been a priority almost from the start. The professional project process provided a great framework with which to put this idea into action, and all of the project stakeholders – from management, to the project manager, to the users themselves – were happy to see the idea become a reality.

How the Project was Managed

The details and daily tasks of the project were managed and undertaken entirely by the project manager. The project manager provided progress reports to his immediate supervisor, as well as the section manager, on a bi-weekly basis. The Agricultural Communications unit director, Dr. Haven Miller, served as the project's

sponsor, and provided logistical support and guidance on an as-needed basis. An abridged organizational chart, displaying only project-related personnel, is shown in Figure 3.

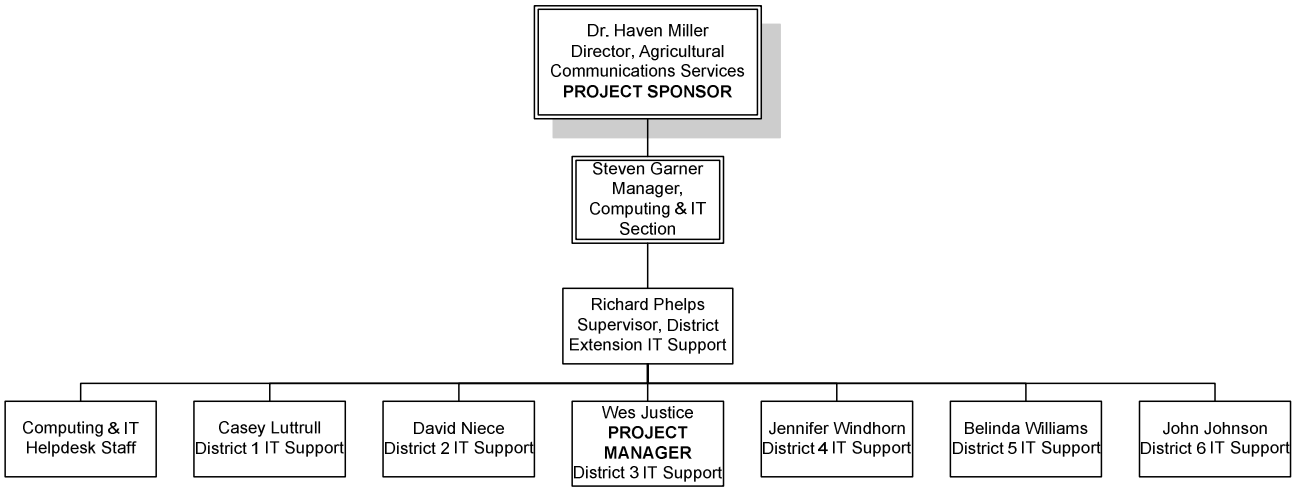


Figure 3: Abridged Agricultural Communications organization chart (project-related personnel only)

Project Stakeholders

The project's stakeholders included the project manager, management (as identified in the diagram above), and the end users in each of the three CES offices selected for the project.

Significant Events / Milestones

Significant events and milestones throughout the course of the project were as follows:

- Project approval (University of Kentucky): Clearing the first major "hurdle," the project received approval from the project manager's superiors.
- Proposal approval (Regis University): The project proposal was completed and approved by Regis faculty.
- Analysis complete: The project's analysis phase was completed, providing crucial information for carrying out the remainder of the project.
- Design complete: The design phase was completed, providing the necessary blueprint for implementation.
- Testing complete: The testing phase was successfully completed, providing evidence that the design was functional and ready for implementation.
- Implementation complete: The planned design was physically implemented in the selected CES office locations.

- Initial support & maintenance complete: The initial weeks of support and maintenance were completed.

Changes to the Project Plan

From an overall perspective, there were relatively few changes to the project plan. One significant change involved server consolidation. Though not originally planned, it was discovered in the design phase that adding a dedicated backup server *in addition* to the office's existing server(s) increased complexity and decreased efficiency. Because of the size of the offices and the relatively light duty of the file servers, it was decided that the new backup server hardware would also take over the file-sharing functions. This provided a number of benefits, including:

- Efficiency of support: One server per office, instead of two or more
- Decreased licensing cost: The Retrospect software costs significantly less when only used to back up one server per location
- Use of displaced server hardware: The removed servers could benefit other CES offices, particularly those with lesser budgets

Thus, while not originally foreseen, this particular change provided a positive impact on the project as a whole.

Did the Project Meet its Stated Goals?

As previously stated, the primary goal of the project was the analysis, design, and proof-of-concept implementation of a standardized backup and recovery plan. This goal was indeed met. The project resulted in the implementation of a comprehensive backup plan that matched the originally identified objectives. The final product was scalable enough to fit the needs of both large and small CES offices, utilized redundant hardware, was essentially transparent to end users, provided an offsite backup component, and required a relatively low amount of support and maintenance. Because of the project's success, the CES offices that participated in the project are able to serve as models for implementations in future offices.

What went Right, What went Wrong?

Many aspects of the project can be said to have gone "right," as all of the project's major goals were accomplished. Taken as a whole, the analysis, design, testing, implementation, and support phases all proceeded

mostly as planned.

That said, while successful, the project could hardly be considered close to nearing perfection. A number of issues arose that caused the project to stray, albeit only slightly, from the project manager's original vision and plan. These included:

- Limited purchasing power due to variable budgets: as emphasized in the section below, budgets were variable and dependent upon each particular office. While adequate in purchasing all necessary equipment and software, larger office budgets would have taken the project's goals to an even greater end. For example, redundancy was identified as a key requirement, in order to eliminate downtime due to hardware failures. Each server was configured with a RAID 5 array, providing redundancy for one of the most prone-to-fail components: the hard disks. However, budgets did not allow the purchasing of more expensive server models featuring redundant power supplies. Thus, in the event of a power supply failure (which is, however, much less likely than a disk failure), moderate downtime will be required while the failed part is replaced.

- Personal circumstances. Due to personal circumstances beyond the control of the project manager (an illness in the family), the project start was put on hold for several months past the original start date. This served as a valuable reminder that even the best-planned projects can sometimes be thrown awry by unforeseen circumstances.

Project Variables & Their Impact

Project variables included:

- Office size: As mentioned, the size of CES offices varies widely, based on the size of the local population served by any particular office. Office size was a very important variable to consider; though the project aimed for a relatively uniform solution, it was critical to determine whether a single hardware/software solution could practically and efficiently serve the needs of all offices. While a single software package was eventually chosen, it was necessary to customize server hardware based on local needs.
- Office budget: Individual office budgets were

also highly variable. This variable had a significant impact on the design phase, as any identified solutions were required to fit a large variety of budgets. Budget concerns were a major factor in the eventual selection of Retrospect, because its licensing structure allowed for an unlimited number of clients at the same relatively inexpensive rate.

Findings / Analysis Results

Considering the results of the entire project of a whole, including the analysis and the findings after examining the implemented system, the project was considered a definite success, providing a vast improvement over the previous backup and recovery methods in use. The Retrospect software, when combined with the chosen hardware and identified best practices, provided an excellent solution. The implemented system proved to be reliable, scalable, and configurable enough to provide a "custom fit" for the needs of the Kentucky Cooperative Extension Service. The project's management looks forward to using the findings as a basis for future implementations in CES office locations throughout the state.

Summary

Through the project's various phases and milestones, many issues were encountered: some expected, some unexpected, some with positive impact, and others with negative. Inevitably, it was necessary for the project plan to change – though relatively little – in order to adapt to the project environment. While some aspects of the project went wrong, many others went right, and the final implemented system provided a very effective solution.

Chapter 5: Lessons Learned

What was Learned from the Project Experience?

The project experience offered numerous lessons – both technical and practical. From the project management perspective, one major lesson was that some things are under the control of the project manager, and some simply are not. No matter how much time and effort is placed into a project plan, things can – and often do! – go wrong. As such, it is important to give substantial consideration to this fact when developing the project plan. For example, a given phase in a project might be estimated to take two weeks. However, any number of unforeseen circumstances might lengthen this time – shipping delays, personal circumstances, workplace political issues, etc. Thus, when planning, it is better to *over*estimate than underestimate the resources – time, budget, and otherwise – required to complete a given part of the project.

From a technical perspective, the project served as an in-depth exploration of the myriad hardware and software technologies available for backup and recovery. By becoming more familiar with these technologies, and gaining hands-on experience with deployment and support, the

project manager was further prepared to expand the system from “proof-of-concept” to a large-scale, statewide implementation.

What would have been Done Differently?

While significant effort was placed into the analysis phase, I believe the need for an *automated* offsite backup procedure was underestimated. A scheduled, automated, Internet-based transfer of critical data to an offsite location (likely the College of Agriculture’s data center) would have provided a positive addition to the project, and yet another safeguard against disaster. Though initially considered, the idea was dismissed, perhaps too quickly, as being outside of the project’s scope. That said, the project *does* provide an offsite backup component, though the tape-based backup requires more human intervention.

Initial Project Expectations Met?

As stated in Chapter 1, the original goal of the project was the analysis, design, and proof-of-concept implementation of a data backup and recovery plan for Cooperative Extension Service offices. The finished project was expected to be efficient and relatively transparent. As detailed in the requirements discussion,

it was expected to "just work," and to prevent data loss on multiple levels.

Based on these initial goals, it is the opinion of the project manager, management, and users that the implemented proof-of-concept system did indeed meet expectations.

Next Evolution of Project

As previously stated, the project provided a proof-of-concept implementation. Because of the project's success, the project manager has been authorized to begin initial planning on a statewide implementation of the designed system in all 120 CES offices.

Conclusions / Recommendations

Due to the project's success, the project manager strongly recommends the continuation of the project on a statewide basis. The lessons learned during this proof-of-concept implementation are expected to be a valuable tool. Specifically, the project manager recommends retaining the uniform quality of the developed system, while using available budget resources to maximize the system's benefit in each local office.

Summary

The project experience offered many valuable lessons, both technical and practical. Though a few aspects would have been handled differently given the chance, the project was nonetheless considered a strong success. Based on the successful outcome, the project manager recommends expanding the project to a statewide scale.

Appendix A: Screenshots

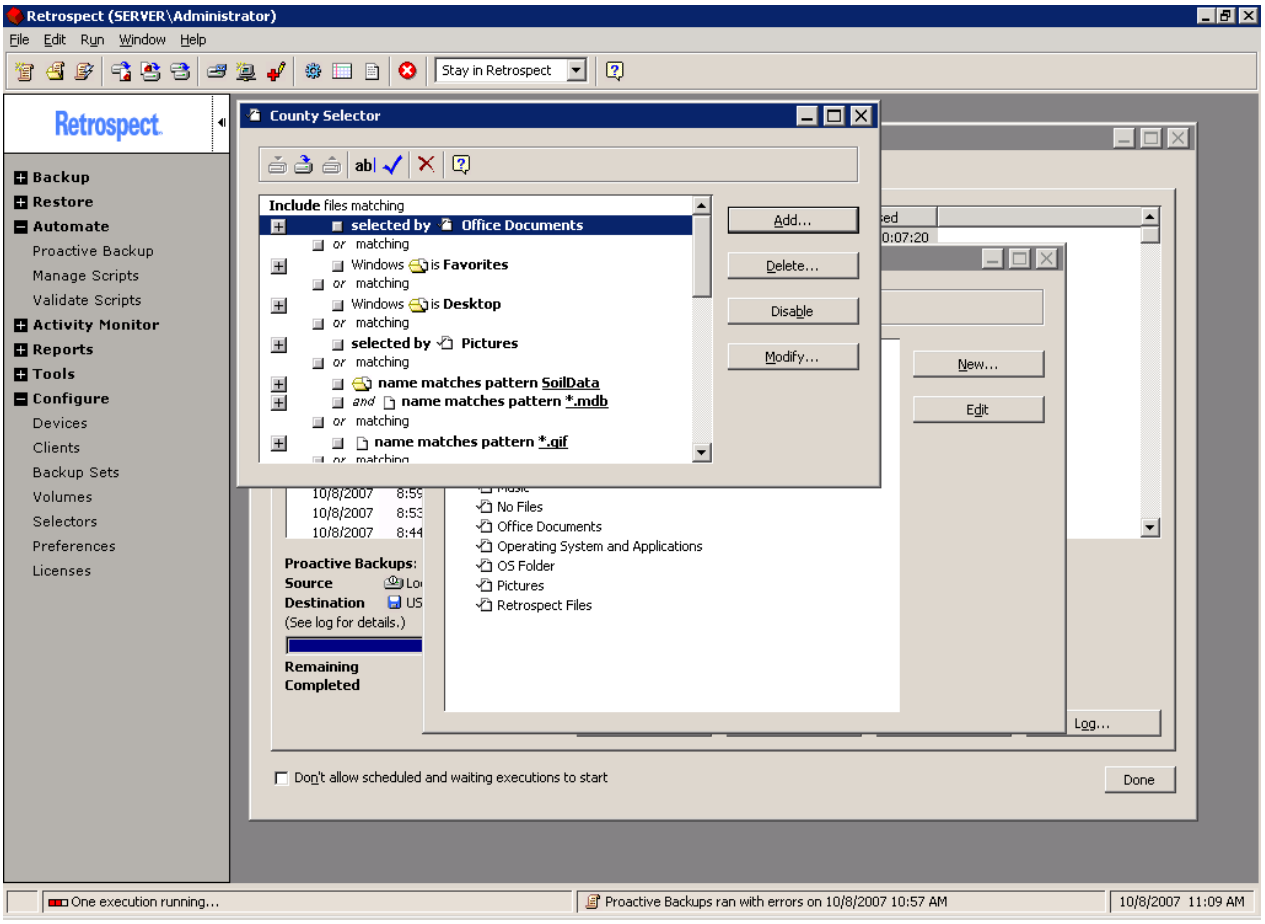


Figure 4: Screenshot: Client data selection filter

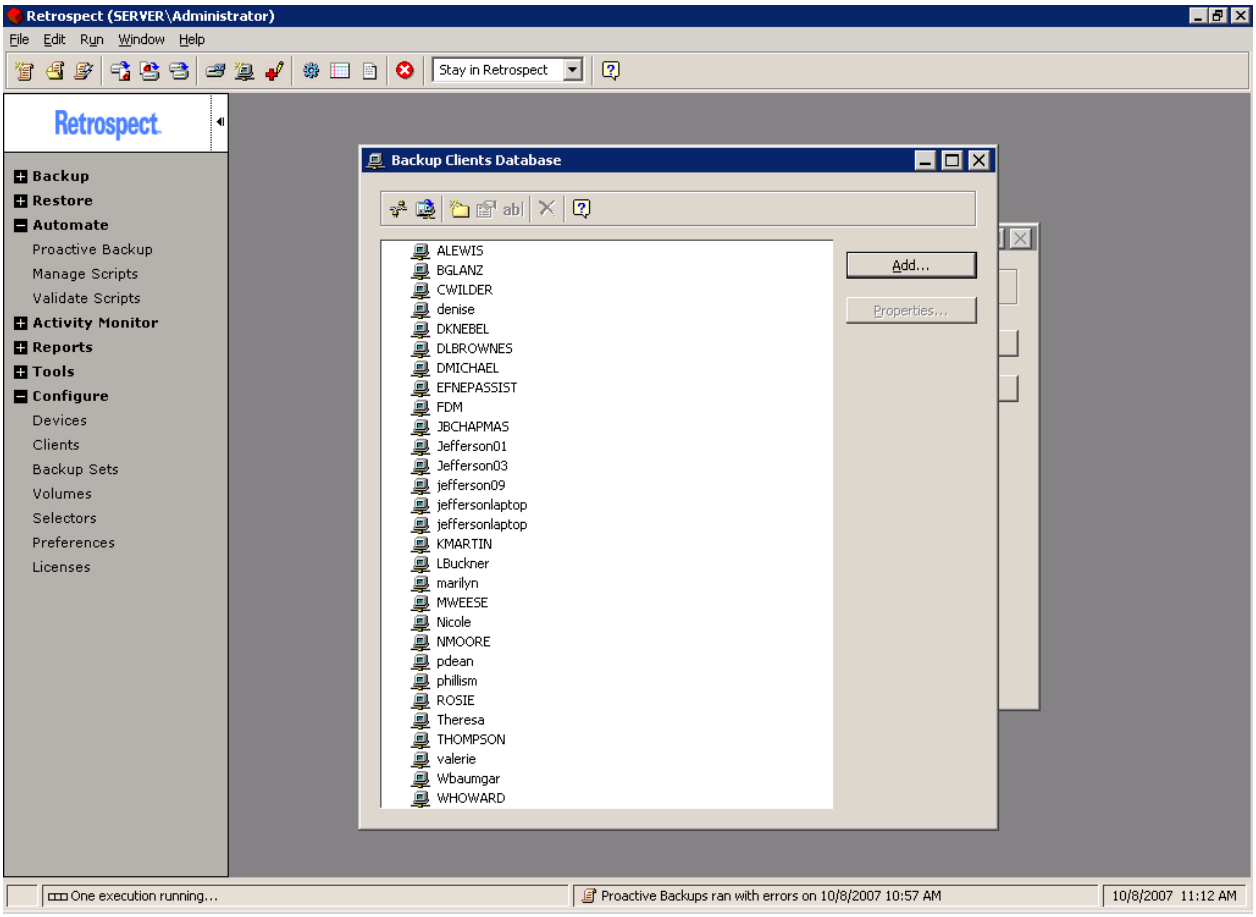


Figure 5: Screenshot: Clients database

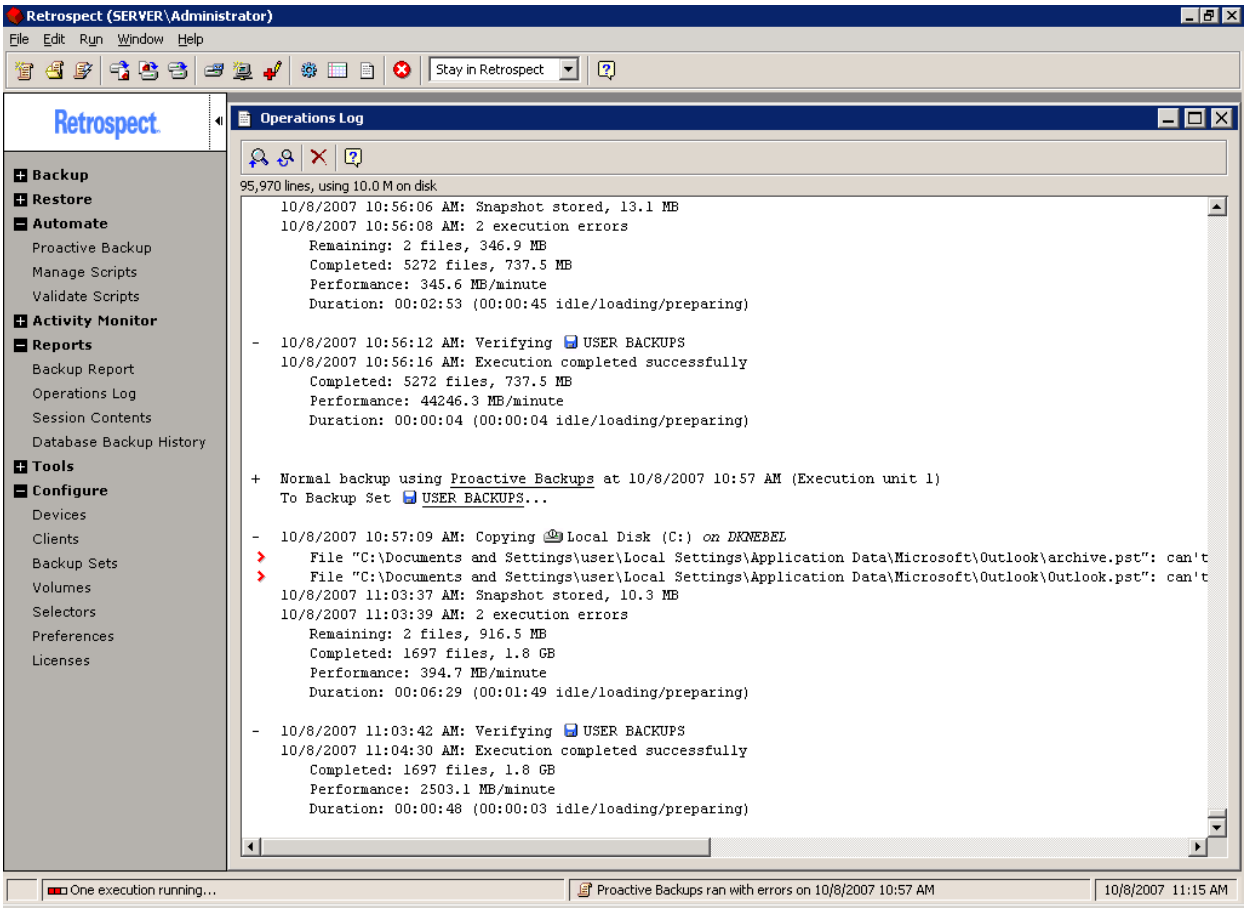


Figure 6: Screenshot: Operations log

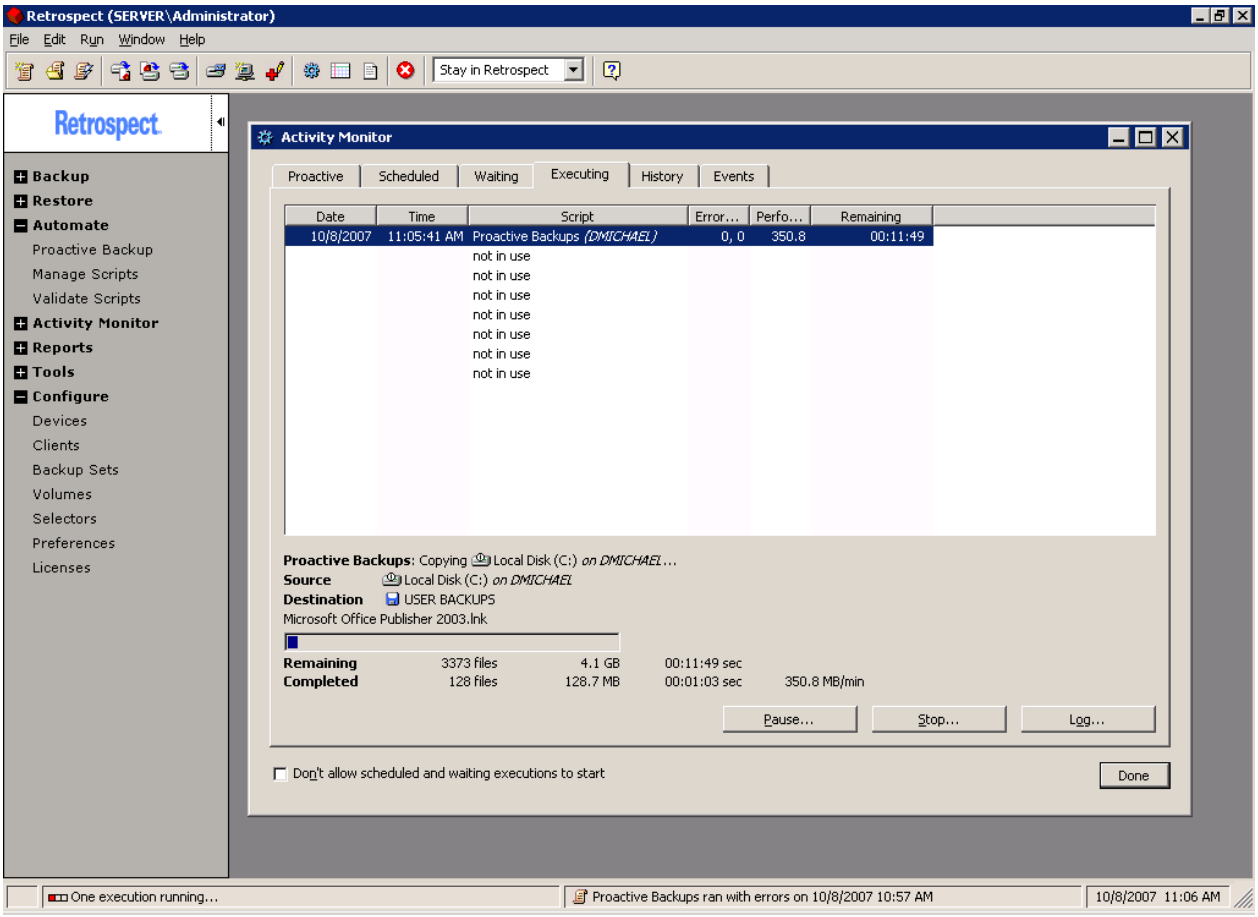


Figure 7: Screenshot: Client D2D backup in progress

Appendix B: Best Practices / Backup Methods for CES Offices

When developing the best practices, it was necessary to analyze the perceived threats to the data as well as current business process requirements (Sandhu, 2002). After consulting with project stakeholders and reviewing applicable research, the following practices and methods were identified during the design phase. Where applicable, notes are included on each practice was specifically implemented in the Retrospect software. It should be noted that these practices are not necessarily generally applicable in all situations; rather, they were identified with CES offices specifically in mind.

Workstation Backup Strategy

Disk-to-disk-to-tape (D2D2T) backups are an ideal solution to meet the need for relatively fast backup and restore procedures, as well as the need for reliable long-term storage. It decreases backup and recovery times and increases overall efficiency, and has even been recently touted as a "data savior" (Pascarelli, 2004).

Workstation backups in each office utilized a D2D2T strategy. Individual workstations were backed up once per 24 hours to the server's RAID array. The Retrospect

software was configured to store at least the last ten snapshots for each client. Scheduling a specific time for each workstation backup was not practical, as it is nearly impossible to predict when any given employee workstation might be powered on or off. Thus, Retrospect's "proactive backup" option was used. This option simply ensures that workstations are backed up once per specified interval (in this case, once per 24 hours).

These disk-based backup sets were eventually archived to tape (and subsequently moved offsite) as part of the server backup strategy (described below).

Server Backup Strategy

The data on the server itself, including user and application data as well as workstation data (inside D2D backup sets) will be backed up to tape on a daily basis, with tapes regularly being rotated offsite.

The newly purchased servers feature an LT0-2 tape drive. LTO-2 tapes feature 200GB of native storage capacity; when a 2:1 data compression ratio is used, the capacity doubles to 400GB. The large capacity will enable all of the servers' data to fit onto a single tape for the foreseeable future.

Because it is necessary for designated CES personnel

to make weekly tape changes, the tape rotation strategy was designed with simplicity as a priority. Three tape backup sets were defined: Red, Blue, and Green. The backup set names correspond with the color of the label on the physical tape, to ensure that tapes are easily located on not confused with one another.

A backup script was created in Retrospect to backup to the *Red* backup set every three weeks on Monday, Tuesday, Wednesday, and Thursday. A backup was also scheduled for Friday, but configured as a "recycle" backup: essentially Retrospect's version of a full backup, wherein all of the media on the tape is erased and all of the server data is newly copied.

Next, identical scripts were created for the *Green* and *Blue* tape sets, but scheduled to start one week later, respectively. This strategy effectively resulted in a daily backup to tape, with tapes rotating weekly.

Finally, to meet the offsite backup requirement, a "new media" backup was scheduled for the *Red* backup set, occurring every six weeks on Friday. With new media backups, Retrospect requests a new tape for the Red set before performing the backup. When the new tape is inserted, a full backup is performed. Thus, the older Red tape can be rotated offsite.

While somewhat complex in description, these scripts resulted in a remarkably simple tape backup strategy, especially from the perspective of the personnel designated to change the tape. A typical six weeks in the tape backup process are illustrated in Table 8.

WEEK	DAY	ACTION
Week 1	Monday	Move old red tape offsite Insert new red tape New media backup to red
	Tuesday	Normal backup to red
	Wednesday	Normal backup to red
	Thursday	Normal backup to red
	Friday	Recycle backup to red
Week 2	Monday	Insert green tape Normal backup to green
	Tuesday	Normal backup to green
	Wednesday	Normal backup to green
	Thursday	Normal backup to green
	Friday	Recycle backup to green
Week 3	Monday	Insert blue tape Normal backup to blue
	Tuesday	Normal backup to blue
	Wednesday	Normal backup to blue
	Thursday	Normal backup to blue
	Friday	Recycle backup to blue
Week 4	Monday	Insert red tape Normal backup to red
	Tuesday	Normal backup to red
	Wednesday	Normal backup to red
	Thursday	Normal backup to red
	Friday	Recycle backup to red
Week 5	Monday	Insert green tape Normal backup to green
	Tuesday	Normal backup to green
	Wednesday	Normal backup to green
	Thursday	Normal backup to green
	Friday	Recycle backup to green
Week 6	Monday	Insert blue tape Normal backup to blue
	Tuesday	Normal backup to blue
	Wednesday	Normal backup to blue
	Thursday	Normal backup to blue
	Friday	Recycle backup to blue

Table 8: Typical six weeks in tape backup schedule

As illustrated by the table, under normal circumstances, human intervention is only required once a week on Mondays in order to change the tape. All of the other processes are automated.

Backup Metadata

Because the onsite disk and tape backups, along with the offsite tape backups, will eventually grow into a large quantity of raw data, the creation and maintenance of metadata is crucial. The metadata system will serve as an "index" for the backup data itself, and will allow objects to be easily located for restoration when necessary (Farley, 2001).

The Retrospect software was configured to maintain a "catalogue" (Retrospect's term for metadata) for all backups, snapshots, and media. The catalogue can be searched or simply browsed to locate specific backup data. It was stored in a common area across each server, and was itself backed up on a daily basis.

Offsite Component

Due to the risk of data loss due to unforeseen circumstances such as theft, fire, or natural disaster, an offsite backup component is absolutely crucial.

As described in the Backup Strategy section, Retrospect was configured to utilize a rotating library of tape media, wherein tapes were eventually rotated offsite for permanent, secure archival storage.

Snapshot Backups

"Snapshot" backups are an effective alternative to full, incremental, and/or differential backups. While these three traditional backup types are certainly effective when correctly applied, they can also create unnecessary complexity and long restoration times. Snapshots are essentially incremental backups, but with the use of metadata and software logic, they are made to resemble an ongoing set of virtual "full" backups. From the Administrator's perspective, when browsing stored snapshots, each resembles a full backup of the disk as it existed at the time of the backup operation. In reality, a static file might only exist once in the physical backup set, with each snapshot merely containing a "pointer" to it.

Retrospect, by default, implements snapshot backup technology.

Integrity Verification

If a restoration is necessary and it is discovered that the backups themselves are damaged or corrupted, the situation can quickly go from bad to very much worse (Piedad & Hawkins, 2001). It is absolutely necessary to implement a method of verification to ensure backup integrity.

Retrospect offers data integrity verification by default. While the feature can be disabled to speed backups, that is not an option for this project.

Fault Tolerance / RAID

Via usage of a RAID array, the server's internal storage hardware must be redundant, in order to provide an effective "first defense" against failed hardware. Hard drives are often the first components of a system to fail, and fault tolerance begins with RAID (Cougias, 2003).

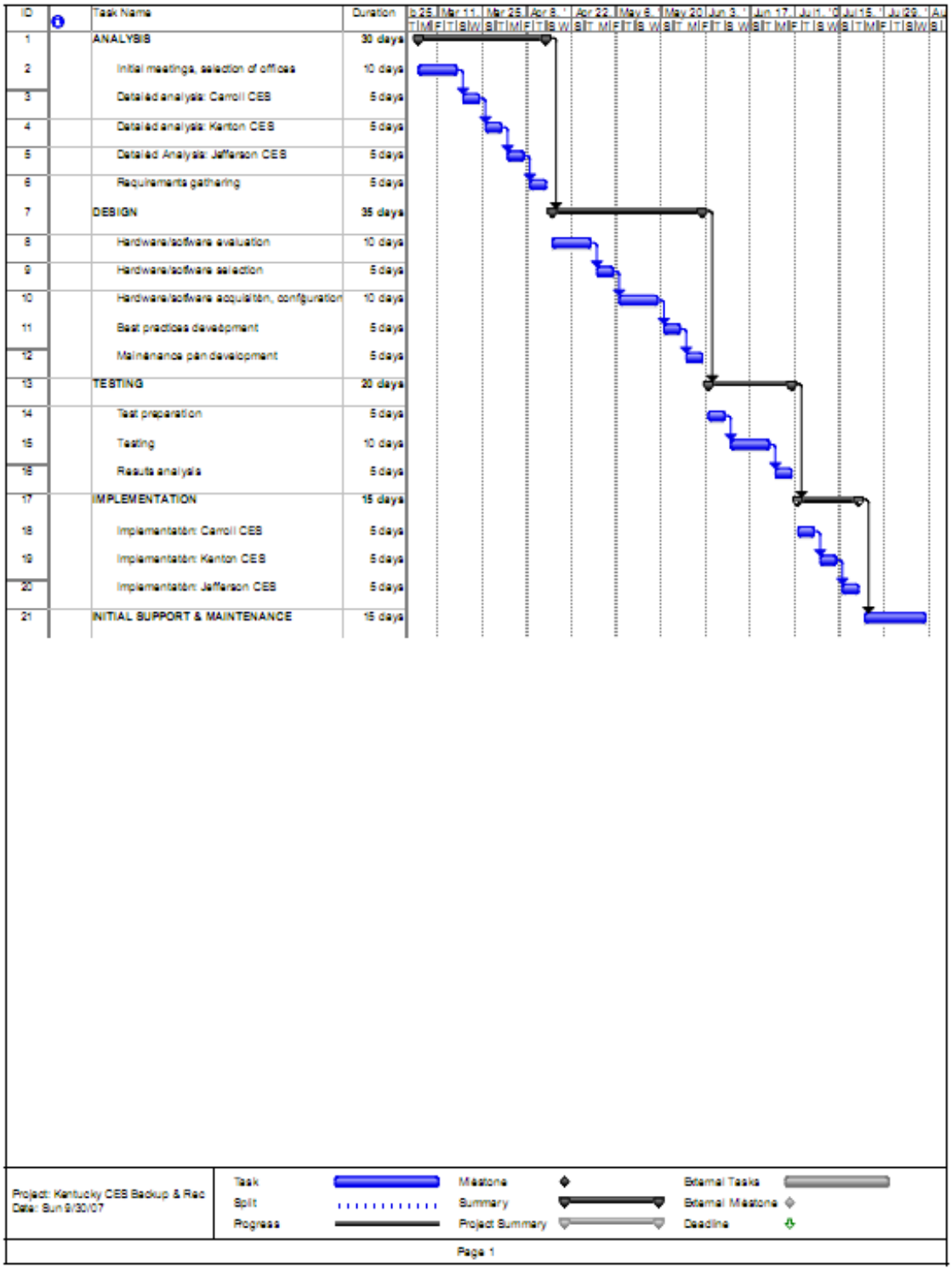
The servers purchased during the project's design phase utilized a RAID 5 implementation, wherein data is striped across multiple disks (in this case, three). In the event of a single disk failure, the data on the failed disk can be reconstructed using corresponding parity information stored on the other disks (Thomasian, 2005).

Physical Security

Physical security is an important consideration when identifying backup practices. The server itself, along with any onsite removable media, must be kept in a secure, environmentally sound location. While offsite backups exist to guard against circumstances such as theft due to physical intrusion, it is certainly desirable to keep potentially sensitive data safe.

Fortunately, each CES office involved in the project had an existing area that provided adequate physical security.

Appendix C: Project Plan



Appendix D: Maintenance Plan

Despite occasional vendor claims of "set it and forget it" backup technology, regular management and maintenance is a crucial requirement for an organization to fully benefit from a backup system (Schultz, 2007). The maintenance plan, developed during the project's design phase, identifies scheduled procedures used to keep the system running at an optimal pace. These procedures include:

- Automated notifications: The Retrospect software will be configured to notify the system administrator, via e-mail, of any alerts that need attention. (Example: bad backup media, failed backups, etc.)
- Weekly review: While the e-mail updates noted above will provide a "first line of defense," the system administrator will make a weekly status check of each backup server, and handle any items that need attention. Ideally, the review should be performed outside of regular CES office hours, thus if the review identified maintenance actions that require a reboot, the effect will be

minimal.

- Weekly updates: During the weekly review, the administrator will apply any necessary updates to the Retrospect software and operating system.
- Integrated media validation: The Retrospect software will be configured to perform automatic media validations after each backup, to ensure data integrity. Manual validations may also be performed as deemed necessary.
- Monthly restore tests: Every four weeks, during the weekly review, the system administrator will perform a restore of selected data from the disk-based and tape-based backups. This process will provide an additional data integrity check, and help to ensure that the system is operating as expected.
- Support process: As is the case with all other technical issues in CES offices, initial support requests will be processed at the College of Agricultures Computing & IT Help Desk, located on campus. Issues not resolved at the help desk will be passed on to the CES office's designated District Extension IT Contact. (For the project's proof-of-concept implementation, the

project manager also serves as the DEITC for each county chosen for implementation.)

- Annual assessment: The implemented system will be subject to an annual assessment by the project manager and management, in order to continually evaluate current value, processes, needs, and long-term viability regarding the backup system.

Appendix E: Case Study

Prior to the project start, Mrs. Rosie Allen, a Family & Consumer Sciences Agent in the Kenton County CES office, experienced a hard drive failure. Because no adequate backup system was in place, Mrs. Allen lost access to a substantial amount of data, including years of documents, photos, and e-mail archives. The project manager, having exhausted all other options, attempted the "freezer method," essentially a last-resort method involving placing the drive in a freezer overnight in hopes that the mechanical components, contracting due to the low temperature, will put the drive into a temporary working state. Fortunately, this was the case. The drive was revived for about one hour – just enough time to copy the data onto another drive – and then failed again. While Mrs. Allen was quite happy to have her data back, it was clear that a backup method other than relying on the "freezer method" was urgently needed.

Because the system was still under warranty, Dell provided a refurbished drive to replace the failed one. Nearly a month after the project ended, the replacement drive itself failed. However, due to the presence of the

backup system, circumstances were quite different after this latest drive failure. There was no anxiety concerning a potential catastrophic data loss, and no freezers were necessary. After receiving yet another replacement drive from Dell, the drive was installed and imaged, and the data was easily restored from the Retrospect server. Mrs. Allen was back up and running very quickly.

The events in this case study happened outside of the project itself – the initial drive failure occurred before the project start, and the subsequent failure occurred after the initial support and maintenance phase was completed. Nonetheless, the incidents described herein provide a compelling illustration of the effectiveness of the backup system, particularly in comparison to the situation prior to the project's inception.

References

Chernicoff, D. (2005, May 1). 6 essential storage strategies to meet compliance needs. *Windows IT Pro, 11*(5), S10.

Cougias, D. J. (2003). *The backup book: Disaster recovery from desktop to data center.* Lecanto, FL: Network Frontiers.

Dell, Inc. (2007). *Dell enterprise product specifications guide.* Round Rock, TX: Author (October 4, 2007).

Dow, J. (2004, March 1). Planning for backup and recovery. *Computer Technology Review, 24*(3), 20-21.

Farley, M. (2001, November 19). *Backup best practices*. Retrieved April 22, 2007, from http://searchstorage.techtarget.com/ateQuestionNResponse/0,289625,sid5_cid423460_tax286191,00.html.

Gerber, B. (2004, June 28). So many bytes, so little time -- disk-to-disk-to-tape backup solutions emerge not a moment too soon. *VARbusiness, 2014,* 52.

Hope, M. (2005, November 1). Users take different approaches to D2D. *InfoStor, 9*(11), 1-2.

Kay, R. (2006, July 3). Backup strategies. *Computerworld, 40*(27), 29.

Martech Systems, I. (2007). *Blue Ribbon Youth Enrollment*. Retrieved October 4, 2007, from http://www.martechsys.com/software/youth.php.

Pascarelli, P. (2004, May 1). D2D2T: Is it quite right for you? *Computer Technology Review, 24*(5), 26.

Piedad, F., & Hawkins, M. (2001). *High availability: Design, techniques, and processes.* Upper Saddle River, NJ: Prentice Hall PTR.

Preston, W. C. (2007). *Backup & recovery.* Boston: O'Reilly Media, Inc.

Ray, R. (2004). *Technology solutions for growing businesses.* New York: American Management Association.

Sandhu, R. J. (2002). *Disaster recovery planning.* New York: Thomson Course Technology.

Satori Software. (2007). *Mailroom Toolkit Office: Address management for Microsoft Office*. Retrieved October 4, 2007, from http://www.satorisoftware.com/Products/MailRoomToolKit/office.aspx.

Schultz, B. (2007, May 21). Backup bonanza. *Network World, 24*(20), 58-59.

South Dakota State University. (2007). *SDSU: Recommended backup routine*. Retrieved March 5, 2007, from http://www3.sdstate.edu/TechnologySupport/PolicyDocuments/RecommendedBackupRoutine/Index.cfm.

TAMU Internal Audit Department. (2004). *Review of agricultural research and extension centers' business operations*. Retrieved March 7, 2007, from https://tamus.edu/offices/iaudit/Reports/HR%20Reports/TAES%20HR%20Reports/20030701%20TAES%20TCE%20Rev%20Research%20Ext%20Centers.pdf.

Thomasian, A. (2005, November 1). Clustered RAID arrays and their access costs. *The Computer Journal, 48*(6), 702.

The three pillars of data. (2007, March 12). *InfoWorld, 29*(11), 21-29.

United States Department of Agriculture. (2007). *Expanded Food and Nutrition Education Program (EFNEP)*. Retrieved October 4, 2007, from http://www.csrees.usda.gov/nea/food/efnep/efnep.html.

University of Arkansas. (2006). *Backing up your data in windows XP*. Retrieved March 5, 2007, from http://www.uaex.edu/depts/InfoTech/helpfulhints/backingupdata.asp.

University of Nebraska-Lincoln. (2007). *NSave: Campus-wide backup service for desktop computers & servers*. Retrieved March 5, 2007, from http://nsave.unl.edu/index.shtml.