

Report on UGA Campus Data Storage, Backup, and Recovery Needs

June 2010

Committee Membership

Jeff Teasley, EITS (Chair)

Wayne Crotts, College of Agricultural and Environmental Sciences

Brad Hunt, Terry College of Business

David Matthews-Morgan, EITS

Jim Metcalf, Terry College of Business

Jerry NeSmith, OVPR

Mark Walters, College of Education

Michael Wiesenberg

Introduction

At the core of our business model and education delivery processes is the collection, management, preservation, and redistribution of data in one form or another. Unlike twenty years ago, all of our data is not cloistered in one physical location under the watchful oversight of a specialized team of systems analysts. Instead, a multitude of data centers have emerged within the confines of individual colleges, departments, and service units. The aggregate data storage capacity continues to expand in these diverse computing environments. There are at least 90 such units at UGA, and each one must manage the storage and preservation of institutional data.

Managing the storage and preservation of data is not always highly valued within the overall context of the various information analysis and delivery processes that comprise our research, education, and outreach business models. However, it is extremely important to the short-term and long-term continuance of these processes and the effectiveness of our institution. Data storage backup and recovery is a necessary yet costly and time consuming function.

Challenges abound. Conservatively, data custodians estimate a storage growth rate of 18% annually. Increasingly stringent security measures and regulatory requirements are being imposed, requiring greater compliance diligence and increased resources. Clientele expectation for uninterrupted access to information is on the rise. Budget cuts have limited the ability to properly staff departmental IT units that also have inadequate equipment and software tools to meet the new access demands and regulatory mandates for data retention. The need for a better understanding of these issues and research into possible solutions and efficiencies stimulated the call from the CIO for a committee to research the options surrounding this issue and recommend a five-year strategy for the creation of a centrally provided data storage and backup environment to be included in the UGA IT Master Plan.

Methodology/Technology Overview

The committee thought it important to accurately frame the specific objective from the outset. The following is the objective statement, which the group used to guide its effort:

“Develop an institutionally focused recommendation for safeguarding the University against data loss due to disk or server failures while also lowering the total Cost of Ownership for UGA.”

The challenges faced when considering a centralized backup solution include identifying a solution that can be:

- OS agnostic and preferably using open standard protocols to accommodate the widest range of operating systems on campus;
- Deployable within the Athens campus network infrastructure and supportable to extended campus locations;
- Scalable. A centralized backup solution would not only be implemented in phases, but also need to be flexible to add to the storage environment in a time responsive manner;
- Manageable. A monitoring/management system would need to be in place to insure that:
 - successful backups and recovery of data can be verified by the college/unit/department;
 - the solution is flexible in order to support colleges/units/departments initiation of new backup plans and/or editing existing ones.
- Secure. The backup solution would need to meet established security criteria for handling the transmission and archiving of the stored data.
- Cost Effective: A centralized backup solution should offer identifiable cost benefits relative to current departmental backup and recovery practices.

The committee looked at the feasibility of current industry standard technologies for supporting large-

scale backup and recovery. These included:

- SAN (Storage Area Network) Technology
- NAS (Network Attached Storage) Technology
- iSCSI (internet Small Computer System Interface)
- 3rd party Solutions (which may include a combination of the above technologies).

The options for delivering the above technologies in the provision of a centrally managed storage/backup solution are many. The committee examined seven such options to determine the ultimate viability of each and the 'best fit' for the University at large (see Appendix A). These options were:

- Centralized solution for data backups only;
- Centralized solution for data storage (backups of stored data would result);
- 3rd Party backup solutions;
- 3rd Party storage solutions;
- A common, decentralized solution (i.e., standards based and driven);
- A meshed option that would combine central and departmental resources;
- Departmental partnerships.

A peer review/survey (See Appendix B for results) was accomplished for this effort, which revealed the following high-level information:

- 64% of survey respondents provide centralized server level data backup services to their campus
- 75% offer file-level backups, while only 25% provide DR level service
- 50% of peer centralized backup services are managed in a way that provides local, departmental control of their backups
- 75% of those who provide backup services provide an offsite option for data storage
- Infrastructure used to provide backup services is widely spread, but includes tape as a part of the service
- The majority of schools reported an annual data growth rate between 10 and 25+% per year
- Over 62% of schools that provide a centrally managed service manage more than 500 TB's for their campus
- On average, schools used 1 to 3 FTE's to support the central data store
- 50% of schools use a cost recovery model to support their central data store, while 38% use a hybrid model
- Campus buy-in for schools that provide this service is relatively low with only 25% of respondents reporting a greater than 50% campus utilization of the service

UGA Infrastructure Overview

The UGA Athens campus network consists of three major components – core, distribution and access layers. UGA's core layer consists of 16 Brocade routers interconnected via partially meshed fiber links. The distribution layer is composed of a router in the main wiring closet (MDF) in a building that connects to the core via fiber and to building networks via fiber or Cat 5E twisted-pair copper cabling. The core and distribution layers have a fixed amount of bandwidth capacity, and the access layer has varying capacity depending on the cabling that is used and the network switches in building wiring closets (IDFs). At the present time, the theoretical bandwidth capacity between core locations is two Gigabits per second (Gbps). However, with network overhead, a conservative estimate of the actual network capacity is 80% of the theoretical amount. This means that the inter-core capacity is approximately 1.6 Gbps. For most buildings on campus, the theoretical bandwidth capacity at the distribution layer is 1.0 Gbps with an actual capacity of 0.8 Gbps (or 800 Megabits per second [Mbps]).

It is important to realize that networked devices share the bandwidth capacity from their IDF to the building MDF. All of the IDFs in a building share the capacity of the distribution layer connection. Multiple distribution layer connections in turn share the bandwidth capacity of the core layer network.

Utilization data for inter-core and building distribution connections indicate that the average amount of bandwidth on the most active links is 18 Mbps and 13 Mbps, respectively (with peaks of approximately 30 Mbps). Although the current core and distribution bandwidth utilization is not near its capacity, future application needs such as research computing and centralized backup services could potentially exhaust that capacity. EITS is in the process of implementing UGA's next-generation core network for the Athens campus, which will consist of a fully meshed 10 Gbps, four-core network with redundant Gigabit links from fiber connected buildings. There is also a pair of fibers from buildings on the Athens campus (with existing fiber connectivity) to the Boyd Data Center (BDC) that can be used to facilitate high-bandwidth data transfers between these buildings and the BDC.

Network connectivity at extended UGA locations such as the Griffin, Tifton and Gwinnett campuses varies by location. The Griffin and Tifton campuses have their own core, distribution and access networks, and the Gwinnett campus (since it is housed in one building) only has distribution and access networks. The three extended campuses, as well as the Terry College Buckhead facility, connect to the Board of Regents managed Peachnet core network at various network speeds – Griffin (77 Mbps), Tifton (42 Mbps), Gwinnett (20 Mbps) and Buckhead (4.6 Mbps). All of the UGA campuses have a maximum amount of inter-Peachnet traffic allocated to them – Athens (200 Mbps), Griffin (50 Mbps), Tifton (25 Mbps), Gwinnett (4 Mbps) and Buckhead (1.2 Mbps). All other extended UGA locations (e.g., Cooperative Extension and Public Service and Outreach offices, international campuses, etc.) connect at either broadband speeds ranging from one to six Mbps or dialup speeds (56 Kbps max).

The UGA central IT organization currently provides storage for more than 90% of its high-profile, distributed server services on a Compellent SAN device. Considerable resources, both staff and financial, have been devoted to the creation of this Enterprise storage fabric which presently offers over 90 TB's of combined storage capacity between Tier I and III disk storage. The environment is managed by a full-featured Compellent management software suite which allows relatively easy assignment of storage containers, management of day-to-day storage operations and automatic data progression technology, which allows more efficient allocation between Tier I and III storage layers. Additionally, an Enterprise class tape backup framework is attached to the Compellent, which allows for production DR tape rotation to a secure, offsite location.

Summary

Various levels of backup and recovery services are being performed at scores of locations throughout both the Athens residential campus and extended UGA campus locations. Equipment, software, media, local infrastructure, and personnel must be in place to perform these necessary functions. Thus, ongoing backup and recovery costs are a significant portion of IT expenditures. The following points are known to impact any decision made:

- Cumulative campus expenditures exceed \$350,000 annually for equipment, software, and media.
- Substantial staff time is also devoted to maintaining this infrastructure in a unit-by-unit manner. Estimates place this at over 10 FTE's for the campus.
- Current campus wide data estimates (excluding Research) indicate a need for over 70 TBs of storage space.
- There are no official UGA institutional standards for administration of backup and recovery services. Data security practices, data management methodologies, retention and recovery practices vary from unit to unit.
- A large percentage of campus units do not include offsite data storage as part of their data backup management practice.
- Technology is available to address the need of centralized backup and recovery services.
- A combination of NAS, iSCSI and Fiber SAN technologies can be put in place to meet UGA centralized storage and backup needs.

The advantages to establishing a centralized backup and recovery service include:

- overall reduction in operation costs;
- better positioning to meet regulatory compliance;

- improved service levels;
- improved and more consistent performance;
- standardization;
- enhanced monitoring and analysis;
- greater security control;
- positioning to link to an electronic offsite storage facility.

After substantial deliberation the committee created a scoring methodology (Appendix A), which resulted in the following priority ranking for the top four options¹:

1. Central Storage Solution;
2. Central Backup Solution;
3. Departmental Partnership Solution;
4. Common/Decentralized Solution.

Subsequent deliberation, evaluation of a net present value calculation performed on the top four options (see Appendix C), consideration of other IT Master Plan committee efforts, along with the combined expertise of the members of the group and information gathered from peer and aspirants provided the recommendation as outlined in the final section below. Appendix D outlines the high-level expectations for the provision of such a service.

Recommendations

As part of an overall institutional data management plan, we recommend that the University initiate a pilot effort that would provide a scalable centralized solution for a data storage **and** backup environment. Given an emerging open source market in this area we believe it would be prudent to begin this effort by leveraging existing central storage resources to the maximum extent possible with a concurrent low-cost purchase of hardware to enable the pilot. We further recommend that an in-depth study be performed to ascertain the potential bandwidth impact to the campus network and any enhancements to the residential and extended campus networks that may be needed to fully utilize such a service. Work to develop an institutional policy for data storage and retention should be immediately started which will outline the operating parameters for the service followed closely by the development of a hybrid cost recovery model, which would help to support the service into the future. Based on the range of backup practices, sensitive data categories, and equipment being used on campus, we also recommend a training class be developed to promote recommended backup and retention standards for the campus.

We recommend that this effort be associated with the result of the IT Master Plan Disaster Recovery committee since electronic and staffing resources between the two would be heavily integrated.

Minimally, the following table represents estimates for the startup needs for the pilot, with concurrent years estimated as campus buy-in is achieved. Ultimately, the service should be self-supporting through cost recovery revenue.

¹ 3rd party solutions were examined and found to be not only cost prohibitive, but complicated by the lack of policy and security around locating potentially critical and sensitive data on resources external to campus control.

Expense Category	FY11	FY12	FY13	FY14	FY15
Hardware					
NAS Controllers	\$11,000		\$11,000		\$11,000
ZNAS Controllers	\$20,000		\$20,000		\$20,000
Network	\$2,500		\$2,500		\$2,500
Storage					
NAS & zNAS Disk			\$27,000		\$27,000
Licensing/Maintenance					
Licensing	\$25,000		\$25,000		\$0
Hardware	\$3,500	\$3,675	\$7,600	\$7,980	\$8,379
Maintenance	\$5,000	\$5,250	\$11,000	\$11,550	\$12,128
Personnel					
Backup Administrator	\$17,000	\$17,000	\$34,000	\$34,000	\$34,000
Total	\$84,000	\$25,925	\$138,100	\$53,530	\$115,007

Appendix A

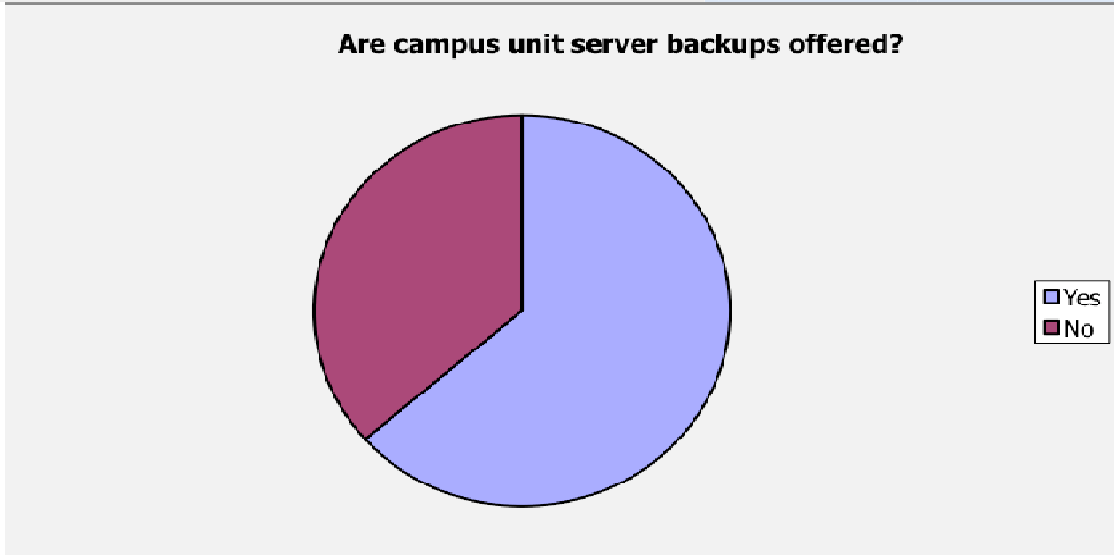
Considerations	Weighting	do nothing	Centralized b/u solution	Centralized Storage	3rd Party Backup	3rd Party Storage	Common/Decentralized Solution	Mesh	Departmental Partnerships
Reliability/resilience	5	1	3	3	3	3	2	1	3
Proprietary limitations	4	1	1	3	1	1	1	1	3
Cost	5	4	3	2	2	2	3	3	2
Scalability	4	1	3	3	3	3	3	3	3
Complexity	4	1	3	3	3	3	3	0	3
Supportability	4	0	3	1	1	1	2.5	1	3
Authority granularity	3	3	3	3	1	1	2	1	3
Client base impact	3	2	1	3	2	2	1	1	2.5
Security (system/data)	5	1	3	2.5	1	1	3	1	2
Recovery options/feature	4	0	3	3	3	3	3	1	3
HIPPA/FERPA et al	3	0	3	3	1	1	2	1	2
Shell access potential and Threats	2	2	unknown	unknown	1	1	2	1	1.5
Authentication mechanism	3	2	1	3	1	3	3	2	3
Institutional Focus	5	1	3	3	1	1	1	1	1
Network Impact	4	2	1	1	1	1	1	2	1
		80	140	144.5	100	106	127	79	138.5

- = 1
 Neutral = 2
 + = 3

UGA EITS Enterprise Backup Survey

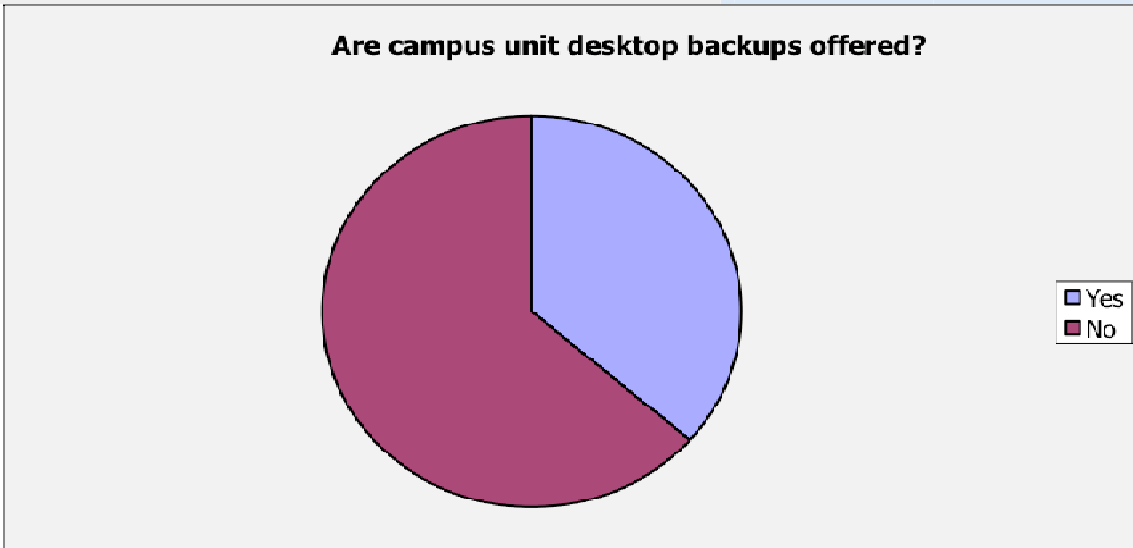
Are campus unit server backups offered?

Answer Options	Response Percent	Response Count
Yes	63.6%	7
No	36.4%	4



Are campus unit desktop backups offered?

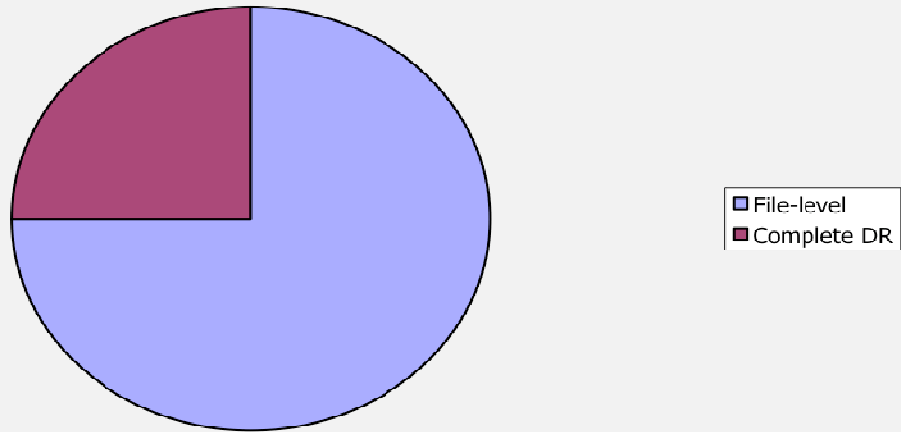
Answer Options	Response Percent	Response Count
Yes	36.4%	4
No	63.6%	7



What type of restoration is available?

Answer Options	Response Percent	Response Count
File-level	75.0%	6
Complete DR	25.0%	2
Any other comments on this?		3

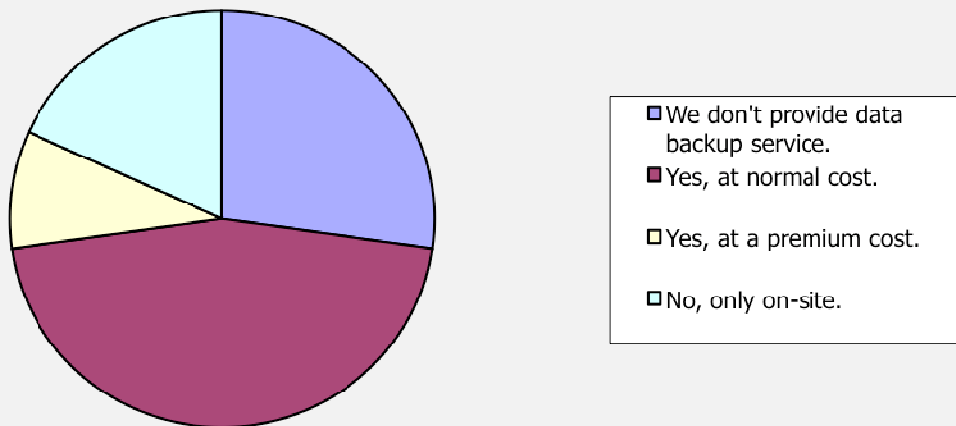
What type of restoration is available?



Is off-site data backup provided?

Answer Options	Response Percent	Response Count
We don't provide data backup service.	27.3%	3
Yes, at normal cost.	45.5%	5
Yes, at a premium cost.	9.1%	1
No, only on-site.	18.2%	2

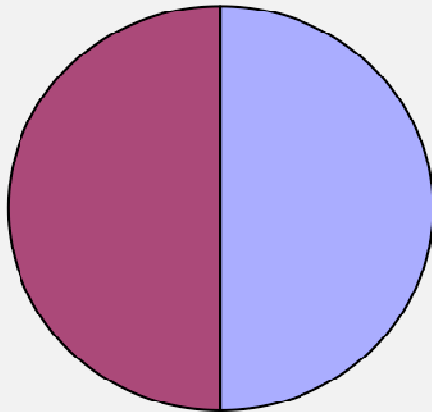
Is off-site data backup provided?



Is client management of storage "areas" utilized?

Answer Options	Response Percent	Response Count
Yes, clients can manage their own backup area.	50.0%	4
No, backup areas are centrally managed.	50.0%	4

Is client management of storage "areas" utilized?

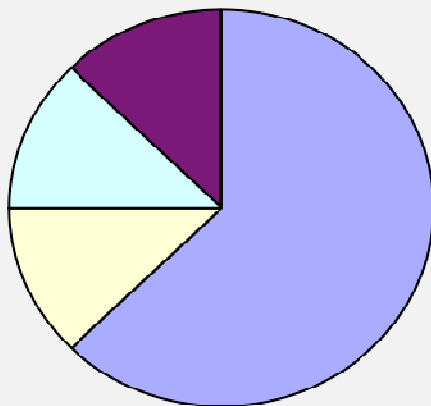


- Yes, clients can manage their own backup area.
- No, backup areas are centrally managed.

How much total enterprise backup storage are you currently managing?

Answer Options	Response Percent	Response Count
> 500,000 GB	62.5%	5
> 250,000 GB	0.0%	0
> 100,000 GB	12.5%	1
> 50,000 GB	12.5%	1
> 10,000 GB	12.5%	1
0 - 10,000 GB	0.0%	0

How much total enterprise backup storage are you currently managing?

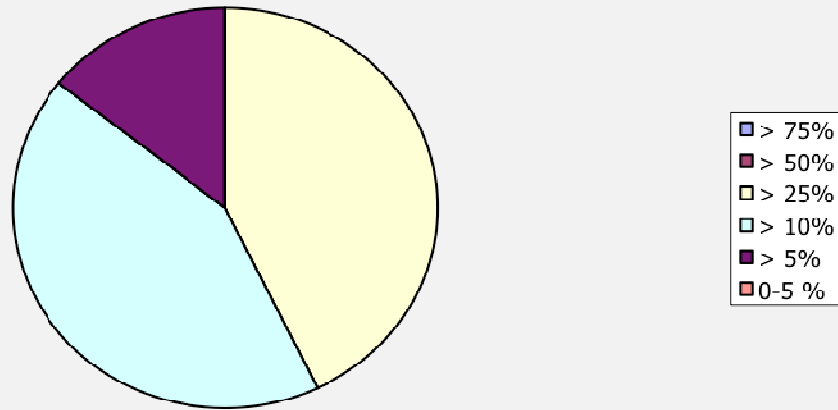


- > 500,000 GB
- > 250,000 GB
- > 100,000 GB
- > 50,000 GB
- > 10,000 GB
- 0 - 10,000 GB

What has been your average growth rate in backup storage space?

Answer Options	Response Percent	Response Count
> 75%	0.0%	0
> 50%	0.0%	0
> 25%	42.9%	3
> 10%	42.9%	3
> 5%	14.3%	1
0-5 %	0.0%	0

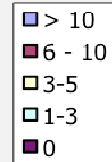
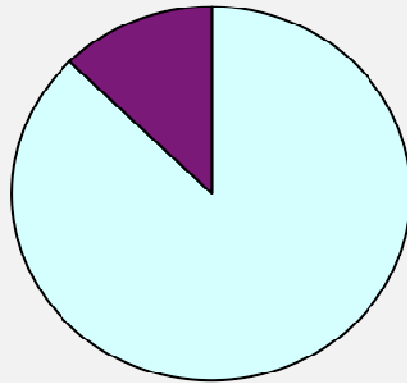
What has been your average growth rate in backup storage space?



How many central IT staff (full-time employees) do you devote to management/provision of the backup service?

Answer Options	Response Percent	Response Count
> 10	0.0%	0
6 – 10	0.0%	0
3-5	0.0%	0
1-3	87.5%	7
0	12.5%	1

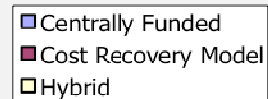
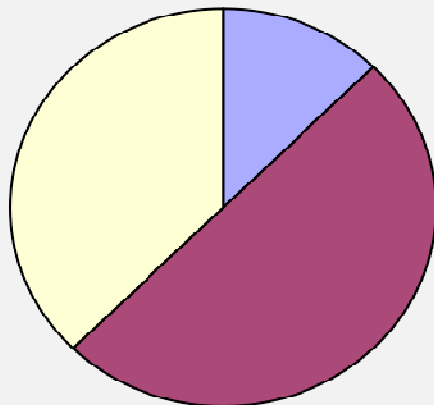
How many central IT staff (full-time employees) do you devote to management/provision of the backup service?



How is the service funded?

Answer Options	Response Percent	Response Count
Centrally Funded	12.5%	1
Cost Recovery Model	50.0%	4
Hybrid	37.5%	3
Other (please specify)		0

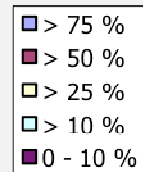
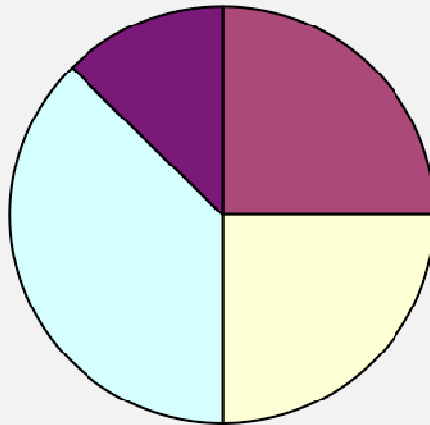
How is the service funded?



What is your estimated % campus participation rate in the program?

Answer Options	Response Percent	Response Count
> 75 %	0.0%	0
> 50 %	25.0%	2
> 25 %	25.0%	2
> 10 %	37.5%	3
0 - 10 %	12.5%	1

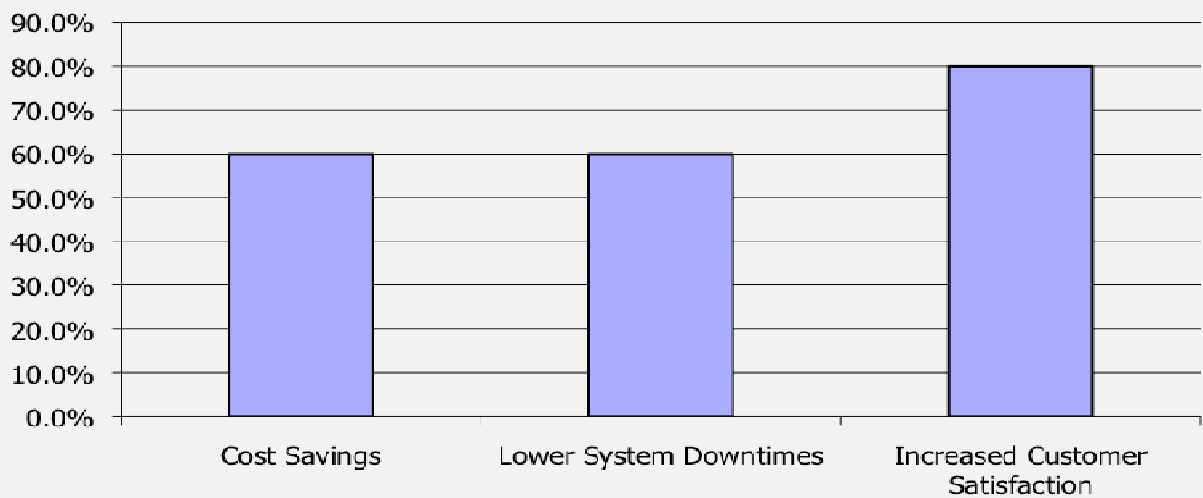
What is your estimated % campus participation rate in the program?



What benefits have you realized as a result of the service?

Answer Options	Response Percent	Response Count
Cost Savings	60.0%	3
Lower System Downtimes	60.0%	3
Increased Customer Satisfaction	80.0%	4
Other (please specify)		1

What benefits have you realized as a result of the service?



Appendix C

Scenario	NPV	Average Annual TCO	Initial Equipment Investment	Annual Equipment Costs	Annual Labor Costs	FTE
Current Decentralized	\$(6,300,000)	\$(940,000)	\$-	\$(300,000)	\$(636,500)	10
Centralized Backup	\$(3,400,000)	\$(480,000)	\$(185,000)	\$(73,500)	\$(402,000)	6
DIY Central BU	\$(3,500,000)	\$(520,000)	\$(27,000)	\$(33,750)	\$(482,400)	7
Dept Partnerships	\$(5,600,000)	\$(830,000)	\$-	\$(300,000)	\$(529,300)	8
MozyPro	\$(6,600,000)	\$(970,000)	\$(66,720)	\$(666,720)	\$(301,500)	5

Appendix D

Considerations	
Reliability/resilience	System must have four 9's availability (i.e., unscheduled downtime = < 53 minutes/ year)
Proprietary limitations	System must not be proprietary in nature unless price is substantially in our favor
Cost	Cost needs to be equal to or less of what departments would pay for an equivalent solution
Scalability	System needs to be scalable to multiple PB's
Complexity	System needs to be as simple as possible in order to provide maximum UGA location coverage
Supportability	System needs to be supportable within context of UGA IT expertise
Authority granularity	System needs to be supportable in terms of data access and data restore by local departmental IT support
Client base impact	System must be accessible to the widest UGA client population possible, OR needs to be accessible to the largest percentage of UGA data
Security (system/data)	Data must be transmitted and stored in an encrypted format
Recovery options/feature	Recovery options must be able to be initiated by dept. IT staff and be accessible at the file level
HIPAA/FERPA et al	Data storage will be HIPPA compliant
Authentication mechanism	System must be accessible via LDAP MyID authentication
Institutional Focus	System must provide the widest acceptability possible, allowing access by the largest number of constituents and acceptable to the institution
Network Impact	Bandwidth impact must not negatively affect regular, day-to-day departmental computing functions and systems
Not in scope for this effort:	Research data storage/backup Desktop data storage/backup