AE Initiative Summary Business Case – Data Center Aggregation

Business Sponsorship & Ownership

Project Name:	IT – Data Center Aggregation
Team Members	Edward Van Gemert (team lead), Melissa Amos-Landgraf, Phillip Barak, Terrence Bradshaw, Kevin Cherek, Rick Konopacki, Steve Krogull, Nancy McDermott, Dan Motl
Business Unit(s):	Vice Chancellor for Administration – Administrative Excellence
Business Process Owner(s):	CIO/proposed organization under the CIO
Preliminary Cost Estimate:	Approximately $2.4M in cost over five years; annual cost based on level of virtualization and co-location; initial capital investment (if necessary) to be determined by implementation team
Preliminary Savings Estimate:	~$6.8M in savings over 5 years

Background

The Administrative Excellence (AE) Phase 1 assessment indicated that the University had the opportunity to generate savings, reduce risk, and improve service through the aggregation of servers and data centers. In January 2012, the AE Steering Committee chartered a team to deliver detailed, action-oriented recommendations on how to accomplish server and data center aggregation. The specific charge, as articulated in the team's charter, described the objective as follows:

> "Develop a new model of server and data center structure to serve the needs of the University's academic, research, and administrative communities, leveraging industry-leading practices for server administration, virtualization, and management to save costs, improve service levels, and minimize data security risks."

The team conducted its work over 22 weeks, recently presented its recommendations to the AE Advisory and Steering Committees, and received Advisory Committee endorsement and Steering Committee approval.

Approach

UW-Madison's data center and server infrastructure has grown across the organization with the increasing demand for IT services. Unlike other areas of the University, the growth of servers and data centers has not been monitored, measured, or documented. At the time the team was launched, no comprehensive data existed about the number of servers on campus, little data was available about campus data centers, no data was available on their respective operating costs, and no organization was responsible for collecting this information. This absence of data, led the team to prioritize the collection information to better understand the current state of servers and data center operations. Data gathering efforts included: (1) collecting server and server infrastructure; and (2) identifying and classifying existing facilities used to house servers. The team also engaged data center administrators across campus to evaluate stakeholder impact of potential future changes.

Collecting Server/Infrastructure Data

In order to complete its work, the team needed data, including the number, type, and location, on the servers on campus. Previous attempts to gather campus-wide data of this nature had relied on self-reporting and yielded incomplete data sets. In an attempt to provide a more objective baseline, the team engaged the Office of Campus Information Security (OCIS) to perform a scan of campus networks to identify possible servers.[1] Results of the scan were then shared with distributed IT staff to refine and vet the data collected.

Though time- and labor-intensive, this approach to data collection and analysis provided critical data supporting the team's work and ultimately to its recommendations.

Identifying Existing Server Facilities

In addition to understanding the landscape of servers across campus, the team prioritized the collection of data about the facilities used to house servers. In order to understand more about the existing server rooms and their operation, the team distributed a survey to data center administrators, identified by the Madison Technology Advisory Group (MTAG), to collect data center/server room characteristics on a variety of attributes including size, customers served, power/cooling, and security.

To supplement the survey, the team engaged DoIT and FP&M to collect data to support the calculation of power use effectiveness (PUE), an industry-standard metric used to evaluate the

[1] The primary tool used to perform this scan was "nmap" (nmap.org). Subnets which were not scanned include Computer Sciences (many servers dedicated to research and not candidates for virtualization or aggregation), ResNET (network segment for students living in residence halls), WiscNET and WiscVPN (not locations of servers within our scope)

efficiency of data center power and utility consumption. The team collected data on power and utility consumption from six data centers across campus, selected to be representative of the diversity of facilities. These PUE data were then used to estimate the reduction in utilities that would occur both through migrating physical servers into more efficient data centers and through reducing the number of physical servers through virtualization.

The combined information from the two approaches supported the development of a financial model to quantify the impact of alternative future state scenarios.

Evaluating Stakeholder Impact

In order to understand the how the current state of server infrastructure evolved and to determine the service needs of future-state alternatives, the team engaged data center administrators in a focus group/ listening session.

Ultimately, engagement with administrators, MTAG, and other stakeholders allowed the team to develop a more refined sense customer needs and expectations related to servers and facilities.

Observations

The absence of data on existing servers and server infrastructure presented a significant challenge. This challenge was addressed through a range of approaches described above, the results of which have produced data that, while not entirely comprehensive, are complete enough to support reasoned decision making.

Data collected indicate that the University's server infrastructure includes approximately 5,000 servers and at least 97 data centers/server facilities. Over $9 million per year is spent on operating this infrastructure.

UW-Madison has no central campus approach for providing data center services and little coordination and communication of distributed server provisioning and data center creation. As a result, the University is duplicating services, spending more on hardware purchases, consuming more utilities (power / cooling), and upgrading and maintaining more facilities than it would if server and data center activities were more coordinated.

The data collected indicate that there are a multitude of data centers/server rooms that provide a wide range of hosting services. In the 97 identified facilities, approximately 52,000 square feet of space is dedicated to housing servers, of which 23 percent is reported as unutilized. Facilities vary widely in terms of size (25 percent are larger than 500 square feet while 16 percent are smaller than 100 square feet). These facilities exhibit significant variance in power/cooling efficiency, security profiles, and redundant/ back-up systems.

While the degree of server virtualization (running multiple servers on a single physical machine) varies by department, an estimated 58 percent of servers are already virtualized to some degree. While the biggest return on investment is in encouraging the virtualization trend, some smaller departments may not have the resources to fully take advantage of this technique.

Access to data center space varies; some departments have purpose-built data centers to support instructional, administrative, or research computing needs, often with excess capacity, while other departments need additional space, relying on under-built, borrowed, or leased space to meet their needs. Ultimately, local resource constraints force the utilization of sub-optimal spaces for server hosting, resulting in what might be lowest up-front cost for that particular department, but higher cost (particularly in higher utilities consumption) to the University.

The current state of server and data center infrastructure and the lack of consistent practices and policies limits the University's ability to maximize the use of its computing resources, manage enterprise risk, and minimize the environmental impact of server infrastructure. The AE data center aggregation team's recommendations of a new model address these four issues.

Core Recommendations

The lack of data on all current server rooms and data centers, particularly data on relative efficiency and capacity, combined with the changing landscape of server hosting (e.g., cloud-based services) have constrained the scope of the team's recommendations and shifted the focus to creating institutional capacity, rather than articulating a specific facility configuration, to realize efficiencies, enhance services, reduce environmental impact, and mitigate risk.

The team recommends that UW-Madison establish a campus data center service to achieve the following goals:

1. Eliminate duplicative infrastructures and substandard facilities (efficiency, sustainability)
2. Match optimized, high-efficiency hosting facilities with service needs, including off-campus/cloud-based options (efficiency, sustainability)
3. Provide consistent service levels across campus (service enhancement & risk management)
4. Promote and encourage best practices, including virtualization (efficiency, service enhancement, sustainability)
5. Align with the UW-Madison Advanced Computing Infrastructure (MACI) and the research community through the provision of foundational services (service enhancement)
6. Develop and maintain appropriate data on existing servers and data centers (risk management, service enhancement)

In order to achieve these goals, the central service provider would oversee a reduction in the number of campus facilities dedicated to housing servers. All data centers under consideration for continued operation would have requisite minimum levels of energy efficiency, security, and projected up-time. Over time, inefficient, suboptimal, and/or underutilized facilities would be repurposed, with servers moving to alternative hosting locations. In conjunction with corresponding local units, the central service provider would manage these alternative, high-efficiency facilities as a shared resource across campus. The process for classifying and repurposing facilities, setting standards for high-efficiency facilities, and establishing decision rights will be determined by subsequent implementation teams.

Through its data gathering process, the team identified types of facilities that may be candidates for the high-efficiency data centers that the campus would seek to invest in and maintain. In addition, off-campus facilities, including public-private partnerships and cloud services would be considered where appropriate, with a priority placed on identifying and meeting customer needs and maintaining service continuity.

The central service provider would also serve as a one-stop shop for data center services including consultation with high-level staff and the virtualization and co-location of servers. Because there are additional efficiencies possible through the creation of specialized staff, the team recommends that a virtual server service be provided, ranging from automatic virtual server creation via web interface to physical to virtual conversions. Units that require physical servers would be provided a co-location service where those servers can be situated in a high-efficiency server room and supported by a range of services from simply the rack space to a more comprehensive management of servers in a highly secure and energy efficient environment.

The central service provider would also be responsible for updating and maintaining the inventory of facilities and servers, for understanding the economics of alternate hosting arrangements (across the range of on-campus physical, virtual, and cloud-based services), and for regularly reporting on these data to inform future decision making.

Ultimately, the creation of a campus shared data center service provider would facilitate campus-wide optimization of the server infrastructure. Over five years, through incremental virtualization (70 percent of total servers virtualized) and co-location (70 percent of servers in high-efficiency data centers), the University has the opportunity to generate $6.8 million in savings. This includes savings related to server replacement, annual backup and software, utilities, and the reallocation of labor. Savings are net of approximately $2.4M in variable costs over five years but do not account for any capital investment. These costs, where required, will be determined by the implementation teams.

Implementation

To address the considerations as outlined in the observations and recommendations sections, the team recommends that three implementation teams are formed with a focus on: (1) defining the organization/governance of the central service; (2) defining the set of services to be offered; and (3) identifying the set of hosting facilities where these services will be provided. Responsibilities of these teams include:

1. Central Service Organization/Governance Team
    a. Develop funding model
    b. Develop operational model
    c. Develop staffing model
    d. Develop policies
2. Services Team
    a. Interview customers to determine what services must be provided
    b. Set service level expectations
    c. Define server hosting facility attributes/requirements
3. Facilities Team
    a. Complete campus data center inventory
    b. Identify spaces suitable for providing co-location and virtualization services
    c. Determine level of investment needed to bring existing spaces to required campus standards
    d. Evaluate cost/benefit of alternative hosting options
    e. Identify services to be provided by off-campus hosting vendors

Customer Readiness and Change Management

The diverse nature of the IT and user community at UW-Madison was reflected in the team's stakeholder engagements. The team recognizes the following as key considerations for implementation:

1. Eight percent of identified facilities are already offering paid server hosting to other internal departments/divisions, with 18 percent interested in doing so. This indicates that select facilities may already have the resources, willingness, and infrastructure to meet customer needs and partner with a central service provider.
2. A number of facilities were identified as candidates for high-efficiency data centers based on key characteristics. These facilities may be suitable for the backbone of a central service, but may require additional investment. The facilities implementation team should conduct site visits to assess readiness, engage with stakeholders, and collect additional data to support decision making regarding space required, space available, and any capital investment required to upgrade existing space or to create additional data center space.

3. Data collected indicate varying degrees of willingness to utilize a hypothetical central service provider. Some users have no need for physical servers and would be early adopters, some are virtualizing already, and others may require significant incentives to participate. The team expects willingness to use a central provider to increase once details of the service are defined and a comprehensive communication plan is launched. However, explicit policies may be required for particular individuals or units that are skeptical of the efficacy of the central service model.
4. Policy will be needed to define minimum standards for data centers and when a new data center is created, when an existing data center is modified or enhanced, or when a data center should be decommissioned.
5. "Virtualization first" policies are in place across many campus IT service units and should continue to be encouraged, though, the team recognizes that there are many systems that are not candidates for virtualization.
6. Data center policies should be consistent across all campus facilities. Access controls, auditing, space utilization and planning, networking, facilities management, inventory/asset management, and more will need to be defined. These policies will help to ensure that different organizational units sharing common facilities can work together to efficiently utilize the space and not create risk to the facility, data, network security, devices or other staff.

The domains of data centers and virtualization are large and complex. The communication plan will need to concisely describe a wide-ranging subject, inform stakeholders throughout the assessment and project implementation phases, and provide benchmarks so the larger campus community can readily assess progress towards transition goals.

Review and Approval

Advisory Committee	Endorsed	May 17, 2012
Steering Committee	Approved	June 5, 2012