

Lawrence Berkeley National Laboratory

LBL Publications

Title

Midland College Requirements Analysis Report

Permalink

<https://escholarship.org/uc/item/0q85n3mm>

Authors

Zurawski, Jason

Meade, Brenna

Publication Date

2022-07-12

Peer reviewed



Midland College Requirements Analysis Report

July 12th, 2022



U.S. DEPARTMENT OF
ENERGY
Office of Science



ESnet

ENERGY SCIENCES NETWORK, INDIANA UNIVERSITY



Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor the Trustees of Indiana University, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California or the Trustees of Indiana University. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California, or the Trustees of Indiana University.

Midland College Requirements Analysis Report

July 12th, 2022

The Engagement and Performance Operations Center (EPOC) is supported by the National Science Foundation under Grant No. 1826994.

ESnet is funded by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. Carol Hawk is the ESnet Program Manager.

ESnet is operated by Lawrence Berkeley National Laboratory, which is operated by the University of California for the U.S. Department of Energy under contract DE-AC02-05CH11231.

This is a University of California, Publication Management System report number LBNL-2001468¹.

¹<https://escholarship.org/uc/item/0q85n3mm>

Participants & Contributors

Thomas Ready, Midland College: Chemistry
Lori Thomas, Midland College: Mathematics
Daniel Ortega, Midland College: Information Technology
Victor Ramos, Midland College: Information Technology
Shawn Shreves, Midland College: Information Technology
Charles Velasquez, Midland College: Information Technology
Byron Hicks, LEARN
Brenna Meade, Indiana University
Jason Zurawski, ESnet

Report Editors

Brenna Meade, Indiana University: meadeb@iu.edu
Jason Zurawski, ESnet: zurawski@es.net

Contents

1 Executive Summary	8
Deep Dive Review Purpose and Process	8
This Review	8
This review includes case studies from the following campus stakeholder groups:	8
The review produced several important findings and recommendations from the case studies and subsequent virtual conversations:	9
2 Deep Dive Findings & Recommendations	10
3 Process Overview and Summary	15
3.1 Campus-Wide Deep Dive Background	15
3.2 Campus-Wide Deep Dive Structure	16
3.3 Midland College Deep Dive Background	18
3.4 Organizations Involved	19
4 Midland College Case Studies	20
4.1 Midland College Department of Chemistry	21
4.1.1 Use Case Summary	21
4.1.2 Collaboration Space	21
4.1.3 Instruments & Facilities	21
4.1.4 Data Narrative	21
4.1.4.1 Data Volume & Frequency Analysis	23
4.1.4.2 Data Sensitivity	23
4.1.4.3 Future Data Volume & Frequency Analysis	23
4.1.5 Technology Support	23
4.1.5.1 Software Infrastructure	23
4.1.5.2 Network Infrastructure	24
4.1.5.3 Computation and Storage Infrastructure	24
4.1.5.4 Data Transfer Capabilities	24
4.1.6 Internal & External Funding Sources	24
4.1.7 Resource Constraints	24
4.1.8 Ideal Data Architecture	25
4.1.9 Outstanding Issues	25
4.2 Midland College Department of Mathematics	26
4.2.1 Use Case Summary	26

4.2.2 Collaboration Space	26
4.2.3 Instruments & Facilities	26
4.2.4 Data Narrative	26
4.2.4.1 Data Volume & Frequency Analysis	26
4.2.4.2 Data Sensitivity	26
4.2.4.3 Future Data Volume & Frequency Analysis	26
4.2.5 Technology Support	27
4.2.5.1 Software Infrastructure	27
4.2.5.2 Network Infrastructure	27
4.2.5.3 Computation and Storage Infrastructure	27
4.2.5.4 Data Transfer Capabilities	27
4.2.6 Internal & External Funding Sources	27
4.2.7 Resource Constraints	27
4.2.8 Ideal Data Architecture	27
4.1.9 Outstanding Issues	28
4.3 Midland College Information Technology	29
4.3.1 Use Case Summary	29
4.3.2 Collaboration Space	29
4.3.3 Capabilities & Special Facilities	29
4.3.4 Technology Narrative	29
4.3.4.1 Network Infrastructure	29
4.3.4.2 Computation and Storage Infrastructure	30
4.3.4.3 Network & Information Security	31
4.3.4.4 Monitoring Infrastructure	31
4.3.4.5 Software Infrastructure	31
4.3.5 Organizational Structures & Engagement Strategies	31
4.3.5.1 Organizational Structure	31
4.3.5.2 Engagement Strategies	31
4.3.6 Internal & External Funding Sources	31
4.3.7 Resource Constraints	31
4.3.8 Outstanding Issues	31
Appendix A – The Lonestar Education And Research Network (LEARN)	32
Introduction	32
Network Services	32

1 Executive Summary

Deep Dive Review Purpose and Process

EPOC uses the Deep Dive process to discuss and analyze current and planned science, research, or education activities and the anticipated data output of a particular use case, site, or project to help inform the strategic planning of a campus or regional networking environment. This includes understanding future needs related to network operations, network capacity upgrades, and other technological service investments. A Deep Dive comprehensively surveys major research stakeholders' plans and processes in order to investigate data management requirements over the next 5–10 years. Questions crafted to explore this space include the following:

- How, and where, will new data be analyzed and used?
- How will the process of doing science change over the next 5–10 years?
- How will changes to the underlying hardware and software technologies influence scientific discovery?

Deep Dives help ensure that key stakeholders have a common understanding of the issues and the actions that a campus or regional network may need to undertake to offer solutions. The EPOC team leads the effort and relies on collaboration with the hosting site or network, and other affiliated entities that participate in the process. EPOC organizes, convenes, executes, and shares the outcomes of the review with all stakeholders.

This Review

In April of 2022, staff members from the Engagement and Performance Operations Center (EPOC) met with researchers and staff from LEARN and Midland College for the purpose of a Deep Dive into scientific and research drivers. The goal of this activity was to help characterize the requirements for a number of campus use cases, and to enable cyberinfrastructure support staff to better understand the needs of the researchers within the community.

This review includes case studies from the following campus stakeholder groups:

- Department of Chemistry
- Department of Mathematics
- Campus Information Technology

Material for this event included the written documentation from each of the profiled research areas, documentation about the current state of technology support, and a write-up of the discussion that took place via e-mail and video conferencing.

The case studies highlighted the ongoing challenges and opportunities that Midland College has in supporting a cross-section of established and emerging research use cases. Each case study mentioned unique challenges which were summarized into common needs.

The review produced several important findings and recommendations from the case studies and subsequent virtual conversations:

- Field research conducted at Midland College could benefit from technology components that are portable to facilitate data transfer and analysis workflows.
- Future research use cases at Midland College may require support for scientific instrumentation that produces data locally, where current use cases rely on external ways of approaching the same problem.
- LEARN and Midland will work to understand future research needs that could leverage cloud resources, or extensive collaborations with external partners.
- The availability of a research testbed could aide in the development of advanced technology components.
- Midland College will explore options of the long-term storage and sharing of research and educational data sets, including videos produced during coursework.
- Midland College and LEARN will collaborate on ways to provide critical research and educational software packages.
- Midland College and LEARN will continue to explore options for wide-area, and metro-area connectivity, including ways to leverage community relationships to accomplish tasks such as path diversity and data backups.

2 Deep Dive Findings & Recommendations

The deep dive process helps to identify important facts and opportunities from the profiled use cases. The following outlines a set of findings from the Midland College Deep Dive that summarize important information gathered during the discussions surrounding case studies, and possible ways that could improve the CI support posture for the campus:

- The Midland College Chemistry Department use case performs periodic field research in remote locations. Due to the remote nature of this research, most equipment related to the scientific process must be transported. This includes, but is not limited to, instruments, sensors, cameras, and storage materials.
- The Midland College Chemistry Department use case performs a number of steps of the scientific process in the field. Examples include processing samples (e.g., distilling to essential elements to prevent having to transport biological material across international boundaries), and gathering measurements and photographs onto transportable digital media.
- The Midland College Chemistry Department use case performs the remaining steps of the scientific process outside of the field research activity. These include steps associated with analysis and synthesis. Examples include sending DNA samples to commercial laboratories that can perform sequencing, uploading processed DNA sequences to shared databases that can match against known patterns, and disseminating results after publication.
- The Midland College Chemistry Department use case will be investigating acquisition of DNA sequencing hardware for use at Midland College, or during field research.
- The Midland College Chemistry Department use case relies on several remote resources that provide access to DNA sequences. Using these resources involves uploading data samples via web-based forms, or accessing cloud-provided storage and processing resources. Uploading to these resources requires stable network connections due to the volume of data.
- The Midland College Chemistry Department use case is constructing sensors that can be used during field research. These devices are able to capture measurements about the surrounding environment; measurements include temperature, salinity, PH, and other chemical and biological signatures.
- The Midland College Chemistry Department sensors are currently deployed in the field in short-term patterns that typically last days. After retrieving data must be manually pulled from system memory. Research to investigate ways to pull data via automated methods (e.g., relying on wireless, cellular, or satellite methods) is underway.

- The Midland College Chemistry Department use case produces GB to TB of research data when in the field over the course of weeks; this activity is typically done yearly. The majority of the data comes in the form of image that are captured when sequencing coral reefs.
- The Midland College Chemistry Department use case has a future research use case to compare their field research data against data collected by federal agencies (e.g., NOAA and NASA). Using satellite image data from these entities, along with advanced processing algorithms run on computing resources, it would be possible to validate a number of research hypothesis for data they have collected during field research.
- The Midland College Chemistry Department use case will face storage issues in coming years as the volume of research data increases on campus. They would like to investigate ways to better leverage local and cloud resources.
- The Midland College Chemistry Department use case wants to investigate ways to utilize advanced software (on premises or within a cloud environment) that can facilitate education on the structure of molecules. Typically, these software packages can be used to translate 2D information into a 3D model that can be manipulated and changed. This software may have license fees associated with it.
- The Midland College Mathematics Department use case covers faculty in the mathematics department that record teaching material on a regular basis. Each faculty member records lectures and other special activities (e.g., working out problems, or training in the use of tools) and makes the recordings available to students.
- The Midland College Mathematics Department use case videos are stored in a number of locations during their life cycle:
 - Cloud storage after recording (e.g., within the Microsoft Teams or Zoom ecosystem)
 - Local (private) storage when downloaded for editing or processing
 - Local (campus) storage when private storage is not available
 - Institutional cloud storage via a Microsoft OneDrive account
 - YouTube, via personal or institutional accounts
- The Midland College Mathematics Department use case videos can vary in size from short clips (e.g., 5-10 minutes when split to explore a single concept), or entire lectures that are over an hour. The resulting file sizes are MBs to low-order GB. It is possible that the entire department can produce 100s of videos over the course of a year, requiring TBs of storage in the coming years.
- The Midland College Mathematics Department use case must leverage the campus network to upload content to cloud providers. Often this requires

multiple hours to send larger videos, and faculty may schedule these uploads to start at the end of the day given the amount of time it takes.

- The Midland College Mathematics Department use case notes that not all cloud use has been smooth. A particular area of friction when using YouTube is related to video ownership and account management. Videos must be associated with a Google account, and there is no way to transfer this to other accounts at this time. This will cause problems as faculty rely on a mixture of their personal and institutional credentials.
- The Midland College Information Technology networking environment is compromised of the main campus, and a satellite facility: the Advanced Technology Center (ATC).
- The Midland College Information Technology use case relies on a 1Gbps connection to LEARN, a 500Mbps to Grande (a commercial ISP), and a 1Gbps connection to Suddenlink (a commercial ISP). The Grande connection is run directly to the ATC, LEARN and Suddenlink are run to the main campus. There is a backup networking path between campus and ATC provided by Grande that is capable of 1Gbps speeds.
- Midland College Information Technology and LEARN could explore ways to leverage DR and backup use cases with other LEARN members in the future, particularly those that that may be able to use certain commercial providers via LEARN connectivity.
- The Midland College Information Technology network profile is heavily weighted toward accessing cloud and commodity resources, many of which support educational activities (e.g., Microsoft Teams, One Drive, and Zoom). Typically, the network can reach 800Mbps of traffic (or beyond) during busy parts of a day. This congestion may result in reduced network experiences for faculty, staff, and students.
- Midland College Information Technology is exploring ways to upgrade connectivity to LEARN: this could be additional 1G connections, or a moving to a 10G connection.

The following outlines a set of recommendations from the Midland College Deep Dive is as follows:

- Midland College Information Technology and the Chemistry Department will investigate ways to simplify the in-situ analysis requirements for field research. This may involve the provisioning of traveling computational resources (e.g., laptop with additional storage, and a satellite internet option) that would allow analysis and sharing of data when in the field.

- Midland College Information Technology and the Chemistry Department will investigate the implications of installing and operating DNA sequencing equipment on campus. This investigation could involve understanding the data and security requirements for operating these instruments long-term.
- Midland College Information Technology and the Chemistry Department will continue to understand the network patterns involved in research; in particular uploading to known cloud resources, or downloading research data from partners such as NASA and NOAA.
- Midland College Information Technology and the Chemistry Department will collaborate to set up a testbed environment to facilitate sensor research and development. This can take the form of a segregated network connection (e.g., secured wireless SSID, dedicated VLANs) that can have a different QoS and security profile versus the campus network.
- Midland College Information Technology, the Chemistry Department, LEARN, and EPOC can explore ways to incorporate emerging computing technologies (e.g., ML, AI) into the research process. Examples include how to utilize NOAA and NASA satellite data, construct processing pipelines, and leverage community resources such as TACC.
- Midland College Information Technology and the Chemistry Department will help develop better workflows and best common practices for long-term storage of research data.
- Midland College Information Technology and the Chemistry Department will investigate options for chemical analysis and modeling software.
- Midland College Information Technology and the Mathematics Department will help to simplify the workflow for capturing, processing, and storing videos. Recommendations include best practices and tools to use, and locations to store long term.
- Midland College Information Technology and the Mathematics Department will work to unify existing posted videos to a single location in the cloud, merging content that may span multiple accounts and providers
- Midland College Information Technology and the Mathematics Department will work to identify bottlenecks in the uploading process, which may include understanding if wireless or wired connections are being used, and the time of day that processing normally occurs.
- Midland College Information Technology and LEARN will evaluate ways to upgrade the LEARN connection in the coming years. This could be additional 1G connections, or a move to a 10G connection.

- Midland College Information Technology and LEARN will evaluate MAN fiber options to connect the campus to LEARN, and to satellite facilities.
- Midland College Information Technology and LEARN can develop a strategy to manage primary and backup connections to multiple providers, including ways to balance network traffic efficiently.
- Midland College Information Technology and LEARN will facilitate discussions on DR and backup options within the LEARN ecosystem. This could take the form of LEARN facilitating contracts with commercial providers, and providing dedicated network paths to ensure efficient operation.
- Midland College Information Technology and LEARN can investigate ways to better access popular cloud providers and services: either through improved peering or the use of caching.

3 Process Overview and Summary

3.1 Campus-Wide Deep Dive Background

Over the last decade, the scientific community has experienced an unprecedented shift in the way research is performed and how discoveries are made. Highly sophisticated experimental instruments are creating massive datasets for diverse scientific communities and hold the potential for new insights that will have long-lasting impacts on society. However, scientists cannot make effective use of this data if they are unable to move, store, and analyze it.

The Engagement and Performance Operations Center (EPOC) uses the Deep Dives process as an essential tool as part of a holistic approach to understand end-to-end research data use. By considering the full end-to-end research data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

EPOC supports five main activities

- Roadside Assistance via a coordinated Operations Center to resolve network performance problems with end-to-end data transfers reactively;
- Application Deep Dives to work more closely with application communities to understand full workflows for diverse research teams in order to evaluate bottlenecks and potential capacity issues;
- Network Analysis enabled by the NetSage monitoring suite to proactively discover and resolve performance issues;
- Provision of managed services via support through the Indiana University (IU) GlobalNOC and our Regional Network Partners; and
- Coordinated Training to ensure effective use of network tools and science support.

Whereas the Roadside Assistance portion of EPOC can be likened to calling someone for help when a car breaks down, the Deep Dive process offers an opportunity for broader understanding of the longer term needs of a researcher. The Deep Dive process aims to understand the full science pipeline for research teams and suggest alternative approaches for the scientists, local IT support, and national networking partners as relevant to achieve the long-term research goals via workflow analysis, storage/computational tuning, identification of network bottlenecks, etc.

The Deep Dive process is based on an almost 15-year practice used by ESnet to understand the growth requirements of Department of Energy (DOE) facilities². The EPOC team adapted this approach to work with individual science groups through a set of structured data-centric conversations and questionnaires.

² <https://fasterdata.es.net/science-dmz/science-and-network-requirements-review>

3.2 Campus-Wide Deep Dive Structure

The Deep Dive process involves structured conversations between a research group and relevant IT professionals to understand at a broad level the goals of the research team and how their infrastructure needs are changing over time.

The researcher team representatives are asked to communicate and document their requirements in a case-study format that includes a data-centric narrative describing the science, instruments, and facilities currently used or anticipated for future programs; the advanced technology services needed; and how they can be used. Participants considered three timescales on the topics enumerated below: the near-term (immediately and up to two years in the future); the medium-term (two to five years in the future); and the long-term (greater than five years in the future).

The case study process tries to answer essential questions about the following aspects of a workflow:

- **Research & Scientific Background**—an overview description of the site, facility, or collaboration described in the Case Study.
- **Collaborators**—a list or description of key collaborators for the science or facility described in the Case Study (the list need not be exhaustive).
- **Instruments and Facilities: Local & Non-Local**—a description of the network, compute, instruments, and storage resources used for the science collaboration/program/project, or a description of the resources made available to the facility users, or resources that users deploy at the facility or use at partner facilities.
- **Process of Science**—a description of the way the instruments and facilities are used for knowledge discovery. Examples might include workflows, data analysis, data reduction, integration of experimental data with simulation data, etc.
- **Computation & Storage Infrastructure: Local & Non-Local**—The infrastructure that is used to support analysis of research workflow needs: this may be local storage and computation, it may be private, it may be shared, or it may be public (commercial or non—commercial).
- **Software Infrastructure**—a discussion focused on the software used in daily activities of the scientific process including tools that are used locally or remotely to manage data resources, facilitate the transfer of data sets from or to remote collaborators, or process the raw results into final and intermediate formats.
- **Network and Data Architecture**—description of the network and/or data architecture for the science or facility. This is meant to understand how data moves in and out of the facility or laboratory focusing on local infrastructure configuration, bandwidth speed(s), hardware, etc.
- **Resource Constraints**—non-exhaustive list of factors (external or internal) that will constrain scientific progress. This can be related to funding, personnel, technology, or process.
- **Outstanding Issues**—Listing of any additional problems, questions, concerns, or comments not addressed in the aforementioned sections.

At a physical or virtual meeting, this documentation is walked through with the research team (and usually cyberinfrastructure or IT representatives for the organization or region), and an additional discussion takes place that may range beyond the scope of the original document. At the end of the interaction with the research team, the goal is to ensure that EPOC and the associated CI/IT staff have a solid understanding of the research, data movement, who's using what pieces, dependencies, and time frames involved in the Case Study, as well as additional related cyberinfrastructure needs and concerns at the organization. This enables the teams to identify possible bottlenecks or areas that may not scale in the coming years, and to pair research teams with existing resources that can be leveraged to more effectively reach their goals.

3.3 Midland College Deep Dive Background

In April 2022 EPOC organized a Deep Dive in collaboration with LEARN and Midland College to characterize the requirements for several key science drivers. The representatives from each use case were asked to communicate and document their requirements in a case-study format. These included:

- Department of Chemistry
- Department of Mathematics
- Campus Information Technology

3.4 Organizations Involved

The Engagement and Performance Operations Center (EPOC) was established in 2018 as a collaborative focal point for operational expertise and analysis and is jointly led by Indiana University (IU) and the Energy Sciences Network (ESnet). EPOC provides researchers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers. By considering the full end-to-end data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

The Energy Sciences Network (ESnet) is the primary provider of network connectivity for the U.S. Department of Energy (DOE) Office of Science (SC), the single largest supporter of basic research in the physical sciences in the United States. In support of the Office of Science programs, ESnet regularly updates and refreshes its understanding of the networking requirements of the instruments, facilities, scientists, and science programs that it serves. This focus has helped ESnet to be a highly successful enabler of scientific discovery for over 25 years.

Indiana University (IU) was founded in 1820 and is one of the state's leading research and educational institutions. Indiana University includes two main research campuses and six regional (primarily teaching) campuses. The Indiana University Office of the Vice President for Information Technology (OVPIT) and University Information Technology Services (UITS) are responsible for delivery of core information technology and cyberinfrastructure services and support.

Lonestar Education And Research Network (LEARN) is a consortium of 43 organizations throughout Texas that includes public and private institutions of higher education, community colleges, the National Weather Service, and K–12 public schools. The consortium, organized as a 501(c)(3) non-profit organization, connects its members and over 300 affiliated organizations through high performance optical and IP network services to support their research, education, healthcare and public service missions. LEARN is also a leading member of a national community of advance research networks, providing Texas connectivity to national and international research and education networks, enabling cutting- edge research that is increasingly dependent upon sharing large volumes of electronic data.

Midland College (MC) is a public community college in Midland, Texas. It was established as an independent junior college in 1972 and held its first classes on campus in 1975. The college offers over 100 degree and certificate programs, as well as a variety of Continuing Education programs. Midland College is accredited by the Commission on Colleges of the Southern Association of Colleges and Schools to award certificates and associate and baccalaureate degrees. It is also approved by the Texas Higher Education Coordinating Board.

4 Midland College Case Studies

Midland College presented a number use cases during this review. These are as follows:

- Department of Chemistry
- Department of Mathematics
- Campus Information Technology

Each of these Case Studies provides a glance at research activities, the use of experimental methods and devices, the reliance on technology, and the scope of collaborations. It is important to note that these views are primarily limited to current needs, with only occasional views into the event horizon for specific projects and needs into the future. Estimates on data volumes, technology needs, and external drivers are discussed where relevant.

4.1 Midland College Department of Chemistry

Content in this section authored by Thomas Ready, Midland College Chemistry Department and Marine Science Department.

4.1.1 Use Case Summary

Undergraduate Research in the Fox Science Building was greatly curtailed due to COVID-19 for much of 2020 – 2022. The marine science research effort at Midland College was non-existent for 2 years. Dr. Ready has re-started a modest research effort for this academic year (2021-2022) with 2 students participating.

Midland College is returning to the Caribbean Sea this Summer (2022) to resume research studying coral reef ecosystems. There are no less than 5 research projects planned for the trip; Prospectively, three of our marine research projects could utilize Midland College's IT-capabilities in a significant fashion.

This section will profile 4 use cases:

- The first research project (headed by Prof. Marlana Mertens) we will execute is the collection of coral mucus from both healthy and diseased coral species.
- A second research project is the quantification of planktonic (suspended) biomass in the ocean water at different depths.
- A third research project is the enumeration of coral populations on the coral reef surrounding Bonaire Island.
- A fourth project will be developing low-cost sensor packages for monitoring the water quality in coral ecosystems.

4.1.2 Collaboration Space

The research collaboration space is operated entirely out of Midland College.

4.1.3 Instruments & Facilities

Most of the instrumentation and facilities used are also operated out of Midland College. Some of the use cases rely on the use of external resources in the form of chemical analysis (e.g., LC mass spectrometry that is used to analyze the sugar/polysaccharides makeup) located at UT Arlington, as well as genomics sequencing provided by commercial entities. Overall, these instruments are being used to understand how diseased corals may be producing different mucus profile, and how it can be quantified.

Field work on the longitudinal time study of the coral reefs on Rhoitan, Honduras, and Bonaire Island date back 5 years (e.g., pre-pandemic). As such there are some instruments that belong to Midland that were left on site, and will be leveraged in future years. In particular, samples are reduced to an inert form (e.g., passed through a centrifuge to create a “pellet”) that can be transferred back to Midland, without having to worry about transfer of organic material across international boundaries.

4.1.4 Data Narrative

One of the research projects (headed by Prof. Marlana Mertens) we will execute is the collection of coral mucus from both healthy and diseased coral species. DNA will be

isolated from the coral mucus, and sent to a commercial laboratory for sequencing. Once the DNA sequence(s) are in hand, they can be uploaded into several data-bases for comparison with known sequences. This process uses enormous amounts of computing capability. The end result will hopefully identify any/all species of bacteria, algae, Zooanthella, or Zooplankton present in the coral mucus. Researchers will look for differences in coral mucus populations between healthy and diseased coral.

A second research project is the quantification of planktonic (suspended) biomass in the ocean water at different depths. Biomass is comprised of a vast mixture of algae, micro-bacteria, cyanobacteria, and Zooplankton. Our goal is to discover any trends between ocean biomass and the prevalence of coral disease. Although the biomass data will contain depth profiling, our surface biomass measurements can, theoretically, be compared with satellite data imagery that NASA/NOAA acquire to monitor algae blooms on the ocean surface. If Midland College can get access to the satellite data (other universities have), this may utilize Midland College's Internet capabilities in a big way.

A third research project we will be undertaking this Summer is the enumeration of coral populations on the coral reef surrounding Bonaire Island. The Midland College team will acquire hundreds of photos of the coral reef in a systematic fashion called a "transect belt". The photos will be uploaded into a cloud-based software package called CoralNet. In the beginning, photos are manually annotated and coral populations counted by human vision. After sufficient number photos (between 50 and 100, each being around 1GB each to get a full survey of the population) have been annotated, the CoralNet software uses machine learning and annotates the remainder of the photos itself. This computing power will save hundreds of man-hours in enumeration of the coral population. The software relies on sufficient band-width to transmit the photographic data set(s) to the cloud and return the results. In practice this produces nearly a TB of data and takes hours to upload to cloud resources.

A fourth project will be developing low-cost sensor packages for monitoring the water quality in coral ecosystems. These are similar to SONDEs, which is a commercial product that can cost around \$15,000 USD with the software that controls and manages data. Midland's pods are estimated to cost around \$1,500 USD to fabricate: these will offer a lower cost way to collect similar data. The fabricated pods will utilize a Raspberry Pi/Arduino platforms. The current thoughts regarding the way to retrieve data will be to anchor and a buoy with a float line that can be positioned 20ft below the water surface. The pods will be attached to this line every 20 feet (e.g., bottom of the anchor [60ft], 40ft, 20 ft. This single column will be deployed for around 1-2 total days, and collect data continuously. After deployment they will be retrieved and data will be extracted (the format is to be determined, but will resemble time/datum couplets that may include temperature, water salinity, water PH, dissolved oxygen, and possibly more (proof of concept will focus on core sensors, more can be added later). The first generation of sensors will be deployed during the summer of 2022. The ideal case for transmitting data would utilize some form of wireless (e.g., 5G, Satellite, Wi-Fi, or other radio signal). An area of research will be the balance of battery life for transmission and sensor readings.

All of the previously mentioned use cases make an assumption that data will be collected locally, and not examined in any substantial way while on site due to a lack of available computing or networking resources. The ability to leverage either portable computing, or non-wired based networking technologies (e.g., satellite, cellular) would enable a more ‘in-situ’ approach to analysis and would help to steer research outcomes while still in the field.

4.1.4.1 Data Volume & Frequency Analysis

It is estimated that previous field research efforts have produced GB of data on a yearly basis, with videos and pictures comprising the largest of the volume. Backups of this data could also push the storage requirements to the small TB range. This is generated yearly, when field research is conducted.

4.1.4.2 Data Sensitivity

There are no sensitive aspects to the use case's data.

4.1.4.3 Future Data Volume & Frequency Analysis

As camera resolutions increase, the data requirements could be solidly TB, but still produced on a yearly basis. The addition of external resources (e.g., datasets from NASA and NOAA – namely fluorescent imagery of algae blooms captured from space) could also bolster these requirements, and may also increase the frequency to monthly as more sophisticated analysis pipelines are planned.

4.1.5 Technology Support

The majority of technology resources used in this work are provided by and operated by Midland College Information Technology. Some aspects may be cloud-based.

4.1.5.1 Software Infrastructure

The CoralNet platform is an online resource that allows for the upload of coral reef research data, along with limited search and analysis capabilities.

The BLAST platform, funded by the NIH, is used to match sample genetic fragment sequences to known genetic sequences with the aim of identifying the species of origin for the sample genetic sequence.

Many of the instruments that are used for chemical analysis are proprietary and have software that must be used.

There are limited amounts of numerical analysis that can be done using software like SPSS, MatLab, and Excel.

There is an interest in exploring more of what AI/ML may have to offer, but the staff and researchers at Midland College are not experts in this area.

4.1.5.2 Network Infrastructure

No special networking requirements exist today. There are two possible future use cases to consider:

- It would be desirable to consider a dedicated wireless testbed (e.g., sectioned off from the campus network) that can be used to design, build, and test the sensor networks.
- The ability to share data while in the field (through a resource such as a satellite receiver, or a cellular network) would allow the ability to analyze and steer research directions before the on-site work completes.

4.1.5.3 Computation and Storage Infrastructure

Removable media is used extensively while in the field. Once transferred back to Midland, local data storage can be used, or the cloud storage allocations the college makes available.

4.1.5.4 Data Transfer Capabilities

Historically it has been the case that uploads to the cloud from Midland can be hard. It often takes hours/days to transfer GB/TB of the collected field research data that is shared with collaborators. This is not often (e.g., only once a year), but it can be challenging. Would like to consider ways to make this easier – either by having faster tools, or other methods to upload from the field, versus a bulk upload when back on site.

4.1.6 Internal & External Funding Sources

Midland is a part of an NSF Grant (grant numbers HRD-1202008 and HRD-1826745): The UT-System Senior Alliance; Louis Stokes Alliance for Minority Participation [in research] (LSAMP). The University of Texas at El Paso is the administrative institution for this grant.

The SCUBA - Student Cohort for Undergraduate for Bio-Marine Studies Abroad - program is a part of that grant, and features collaboration with UT Arlington (Dr. Laura Mydlarz) as major collaborator with that aspect of marine research.

Future work will involve more collaboration with international partners, along with the possible inclusion of NOAA and NASA research products.

4.1.7 Resource Constraints

The current largest constraint to research productivity is the inability to process research data when in the field. Having the ability to send data while still on site, and having a processing pipeline that may provide some insight into the findings they are discovering, would help to steer the process of science and influence results immediately, versus having to wait a full year to return and test a new hypothesis.

4.1.8 Ideal Data Architecture

A portable computer that offered data processing, storage, and networking capabilities that can be taken on-site to field research stations.

4.1.9 Outstanding Issues

In the general sense, slow uploads to cloud resources remains the largest issue for the primary research use cases.

A secondary problem, within the Chemistry Department, is the ability to access tools that can be used to teach general and organic chemistry. In particular some software packages can be used to teach these concepts, in particular 3D drawing and modeling of molecules, but the cost for students remains high. Researching either cloud-based or educational discount options would allow students a way to explore this aspect of education. ACD Labs³ would be an option to explore that allows for the drawing of molecules in 2D, convert to 3D, and then simulating the NMR or IR spectrums

³ <https://www.acdlabs.com>

4.2 Midland College Department of Mathematics

Content in this section authored by Lori Thomas, from the Midland College Department of Mathematics

4.2.1 Use Case Summary

The Midland College Mathematics Department has 9 full time and adjunct faculty. Most of the faculty now record lectures so they can be available immediately after a class: this is done to allow for remote education, along with helping students who may need to see materials as they study or may miss a class.

The creation of videos, where they can be stored, and how they can be uploaded, is a serious point of friction for this use case. Many instructors rely on local storage (e.g., portable hard drives) when they run out of space on a personal computer. There is general confusion about which locations video should be stored (e.g., Canvas, Google Drive, OneDrive, Box, YouTube), and this is compounded because some instructors may have an institutional account from Midland, along with personal accounts to the cloud providers.

4.2.2 Collaboration Space

The research collaboration space is operated entirely out of Midland College.

4.2.3 Instruments & Facilities

The primary instruments used are commercial video platforms (e.g., Teams, Zoom) to record video and sound, along with iPEVO document cameras. All of the videos are captured at either Midland, or on occasion an instructor's home.

4.2.4 Data Narrative

The mathematics videos add up fast: each could be an hour or so in length (for a class), or several minutes for the teaching of an independent concept. 9 Faculty will do this twice a week for a semester, resulting in GBs (approaching TBs) of storage requirements. Once uploaded to cloud resources (e.g., YouTube or OneDrive) the local copies can be purged from local storage resources. In addition to videos of an instructor, document cameras may be used to capture the process used for working out math problems. These video length and storage requirements are similar.

4.2.4.1 Data Volume & Frequency Analysis

It is estimated that GB of videos are produced on a weekly basis, and TB may be created by the end of a semester.

4.2.4.2 Data Sensitivity

There are no sensitive aspects to the data featured in this work. On occasion there may be risks in showing content from textbooks or other teaching resources that are copyright protected.

4.2.4.3 Future Data Volume & Frequency Analysis

Future videos may be slightly larger in size, but the frequency of creation, and number created, should remain stable in the future.

4.2.5 Technology Support

The Midland College Mathematics Department relies entirely on Information Technology to support their software, hardware, and networking needs.

4.2.5.1 Software Infrastructure

Beyond commodity services (e.g., video conferencing, productivity software), the department on occasion can use:

- SPSS
- R
- Mathematica
- LaTeX
- StatCrunch - via textbook
- Cloud-based Canvas/publisher web sites for testing/assignments.

4.2.5.2 Network Infrastructure

Networking is mostly accomplished through wireless access, although faculty do have access to wired infrastructure in their offices.

4.2.5.3 Computation and Storage Infrastructure

Midland has a local storage infrastructure (e.g., a “R-Drive” that can be accessed from personal computers) that is tied to the MS Active Directory infrastructure, but this does not have enough space to handle all of the video requirements. Midland also uses OneDrive, Google/YouTube, and Canvas for cloud-based storage.

4.2.5.4 Data Transfer Capabilities

Uploading video content can take hours. Many faculty have gotten into the habit of scheduling uploads to occur at the end of a day (or overnight) due to the multiple hours required.

4.2.6 Internal & External Funding Sources

The Midland College Mathematics Department is based funded, and does not have any external grants.

4.2.7 Resource Constraints

The most critical aspects of this case study are:

- Having a unified process and set of tools to handle the production and sharing of videos. This should be a single location that can be used, along with a “playbook” on the right way to perform the operation
- Adequate local and remote storage for growing video library.

4.2.8 Ideal Data Architecture

Having Midland College Information Technology create a workflow for the Mathematics Department so that they know how to perform the steps of video upload consistently and performantly. This could involve a HOWTO guide, instruction on the use of wired

versus wireless networking, and suggested times to perform network-centric operations such as uploading to cloud resources.

4.1.9 Outstanding Issues

Some faculty have lots of content that is not unified, e.g., it may be located in a personal YouTube account versus the institutional Midland cloud storage. Migrating this content is time consuming, and could be done by a student employee. Additionally, helping to clean up local storage (once content is cloud accessible) will free up resources on the Midland Campus

4.3 Midland College Information Technology

Content in this section authored by Shawn Shreves, the VP of Information Technology at Midland College

4.3.1 Use Case Summary

This use case covers the networking, computational resources, and software packages used by Midland faculty and staff.

4.3.2 Collaboration Space

No external collaborations exist at this time.

4.3.3 Capabilities & Special Facilities

Midland offers standard classroom and office setups for faculty.

4.3.4 Technology Narrative

Midland offers a campus infrastructure of 1 to 10 Gbps on campus, and the use the LEARN network connection for Internet2 access which is 1 Gbps.

4.3.4.1 Network Infrastructure

Figure 1 shows the physical infrastructure of Midland College.

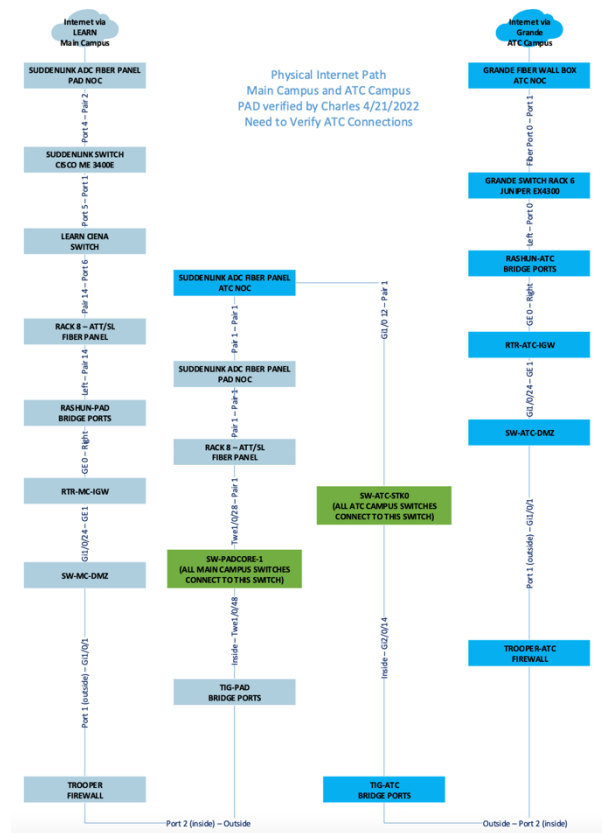


Figure 1 – Physical Network

Figure 2 shows a logical breakdown of the campus network infrastructure.

Basic Campus Design

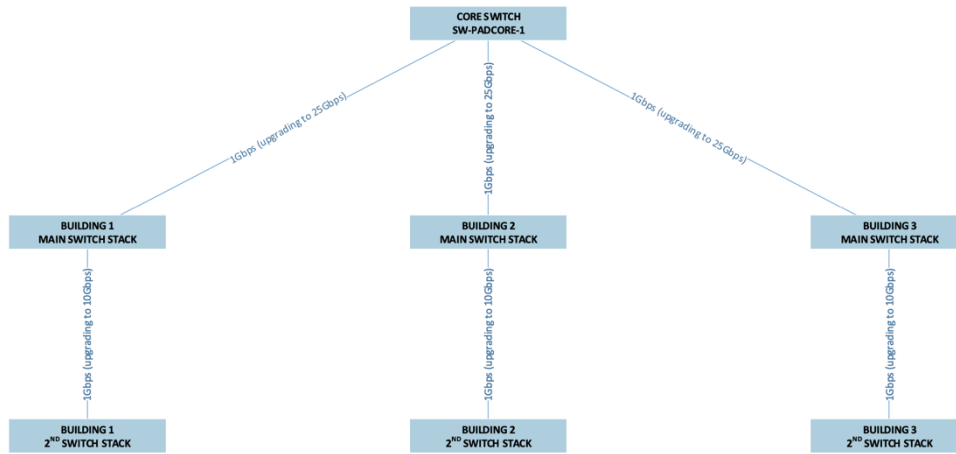
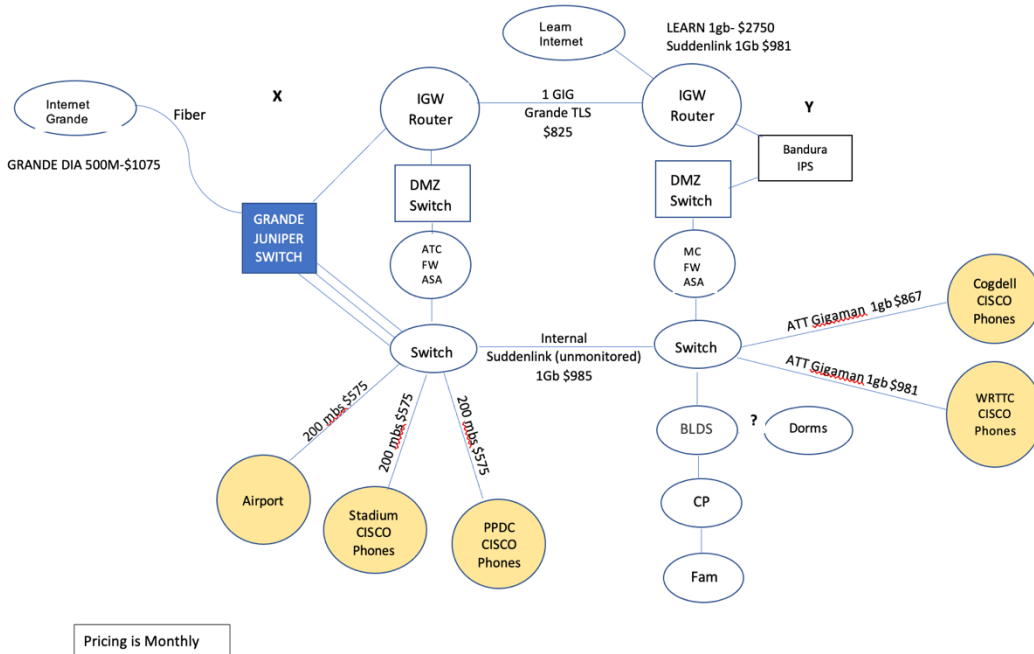


Figure 2 – Logical Network Diagram

Figure 3 shows the wide area connectivity to support Midland College.

Data Circuits



Pricing is Monthly

Figure 3 – WAN Connectivity for Midland College.

4.3.4.2 Computation and Storage Infrastructure

Midland has a local storage infrastructure (e.g., a “R-Drive” that can be accessed from personal computers) that is tied to the MS Active Directory infrastructure, but this does

not have enough space to handle all of the video requirements. Midland also uses OneDrive, Google/YouTube, and Canvas for cloud-based storage.

4.3.4.3 Network & Information Security

We provide a multi-layered approach to security with end-point protection, firewalls, intrusion prevention systems, and e-mail quarantine systems.

4.3.4.4 Monitoring Infrastructure

We use Solar Winds and also the perfSONAR supplied by the LEARN network.

4.3.4.5 Software Infrastructure

Midland College Information Technology is responsible for all software purchases and installations. Midland also provides OneDrive access for local users to facilitate a cloud-based anywhere anytime service for our faculty.

4.3.5 Organizational Structures & Engagement Strategies

4.3.5.1 Organizational Structure

Midland College IT is managed under the Vice President of Information Technology who reports directly to the President of the college and serves on the administrative council.

4.3.5.2 Engagement Strategies

There is no formal engagement strategy at this time, however all faculty work with Information Technology with respect to hardware and software needs.

4.3.6 Internal & External Funding Sources

The Midland College Information Technology is based funded, and does not have any external grants.

4.3.7 Resource Constraints

Midland is approaching the need to increase bandwidth and should be able to increase based upon budgeting. In particular it will be critical to support the cloud streaming use cases in the future, as the current 1 Gbps circuit peaks during the day. Most (if not all) of this traffic is related to video.

4.3.8 Outstanding Issues

Bandwidth remains the primary issue for most, if not all, of the use cases at Midland. Due to the location of the campus, connectivity is rare and expensive, and often must be competed against private industry (e.g., energy companies).

Appendix A – The Lonestar Education And Research Network (LEARN)

Introduction

The Lonestar Education And Research Network (LEARN) is a consortium of 43 organizations throughout Texas that includes public and private institutions of higher education, community colleges, the National Weather Service, and K–12 public schools. The consortium, organized as a 501(c)(3) non-profit organization, connects its members and over 300 affiliated organizations through high performance optical and IP network services to support their research, education, healthcare and public service missions. LEARN is also a leading member of a national community of advance research networks, providing Texas connectivity to national and international research and education networks, enabling cutting-edge research that is increasingly dependent upon sharing large volumes of electronic data.

LEARN's Mission

Empower non-profit communities to execute their missions through technology and collaboration.

LEARN's Vision

LEARN will be the most efficient and effective enabler of research, education, healthcare, and public service communities in Texas using technology and shared services.

Network Services

Members are entitled to appoint an individual to the Board of Directors and to acquire network services from LEARN at member rates. Network services are designed and provisioned based on the needs of individual members through collaboration between those members and the LEARN staff.

Network services, which are funded by the members who consume the services at rates which are set by the Board, sustain current and future network requirements including capital refresh at periodic intervals to keep the network state-of-the-art.

Network services include:

- Layer 1 Dedicated Transport Services Between LEARN Points-of-Presence (POPs),
- Layer 2 IP/MPLS Transport Services,
- Service Level Agreement (SLA) based Layer 2 connections to Cloud Service Providers (AWS, Google, & Azure),
- Routed Layer 3 IP Services,
- Connection Gateways to the National Research and Education Networks (Internet2 and Energy Sciences Network, and on 100G ramps to reach Pacific Wave International Exchanges),
- Seamless access to on-net data centers,

- Inter-POP Port aggregation & Co-location Services
- Commodity Internet Services (100G burst capacity spread across 4 POPs),
- Low-Latency High-Capacity Access to Content and Application Providers (Peering and Caching Services),
- DDoS Mitigation Service,
- Managed Network Service and Consultation, and
- Unmetered Network Service.

LEARN is currently listed as a telecommunication/Internet service provider with the Universal Service Administration Company (USAC). Becoming a USAC telecommunications/Internet service provider permits LEARN's school, library, and rural healthcare customers to receive significant discounts through the Universal Services Fund.

The Board and the staff are committed to ensuring LEARN remains the trusted and preferred means by which its members obtain network services in Texas. There is a broad consensus among LEARN's members that the organization has a unique role to play in the state in providing highly reliable, cost-effective network services to the higher education, K–12, research institutions, healthcare, city and county governments, libraries and museums, and not-for-profits and public service entities. LEARN is a trusted partner and convener in these communities.

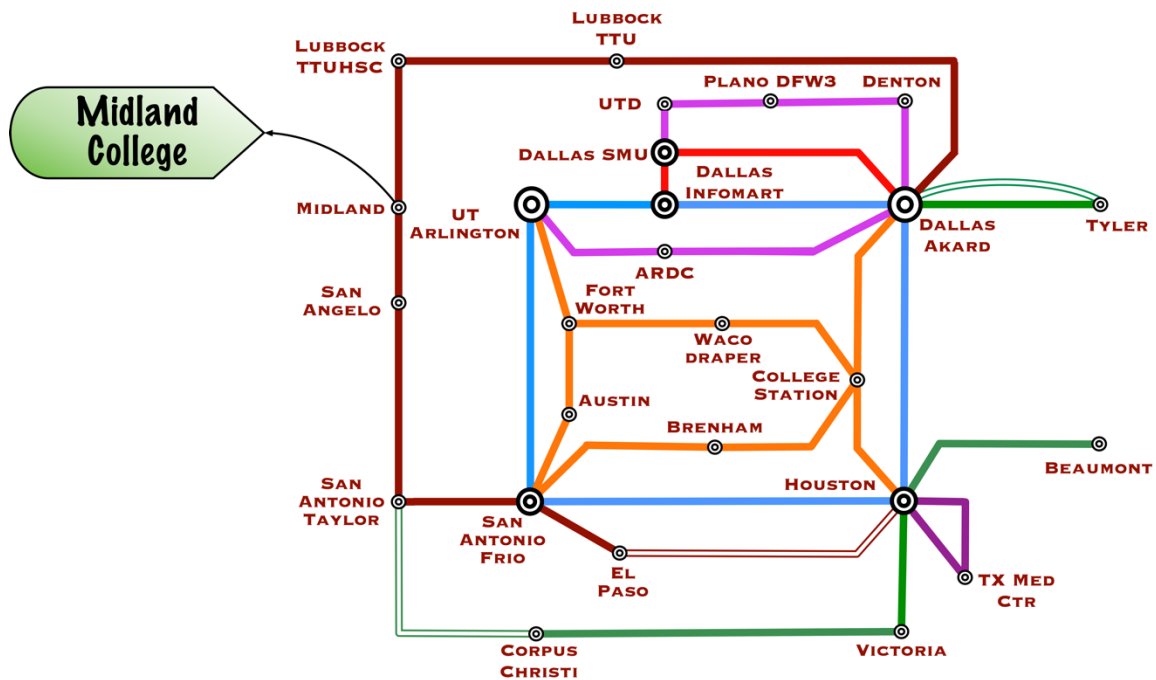


Figure 3: LEARN Connectivity Serving Midland College

CC* Funding

In 2019, LEARN was awarded NSF Awards #1925553: “CC* Regional: Accelerating Research and Education at Small Colleges in Texas via an Advanced Networking Ecosystem Using a Virtual LEARN Science DMZ”.

LEARN is partnering with national organizations in the implementation of this project. Projected impacts include increased opportunities for students to learn about and gain experience in advanced aspects of science, technology, engineering and mathematics (STEM) for which they might not otherwise have had an opportunity, for extension of the project to students and faculty at other campuses in Texas, and for the extension of the LEARN model to other regional networks and smaller campuses throughout the United States.

Objectives:

- Establish a small college collaborative environment within the LEARN community
- Improve network connectivity/services at each college campus for research and education
- Establish a network performance monitoring infrastructure
- Establish a means to facilitate the transfer of large data sets
- Deliver technical training to personnel at each campus
- Develop and implement an outreach program for informing/educating faculty, staff, and students at each college, and develop and disseminate project results