



University IT Open House for Research IT Services

Philip Baczewski, Executive Director, University IT
Ravi Vadapalli, Director of Research IT Services

UNT

UNIVERSITY
OF NORTH TEXAS

EST. 1890

High-Performance Computing

DaMiri Young
Charles Peterson
Yuguang Ma

Data Science & Analytics

Richard Herrington
Jonathan Starkweather

RITS Overview and Organization



- UNT Computing: 1962 – present
- Strategy & Impact
- What are Research IT Services? What's In It For Me?
- Capabilities
- Programs & Opportunities
 - NSF XSEDE
 - ASSUR²E
 - NSF Cloud and Autonomic Computing Center
 - Texas GLO for Disaster Resilience
- Q & A

Research Computing 1962

1962 – Computing at UNT (known at the time as NTSU) officially begins! [IBM 1620](#) purchased to support academic users.



IBM 1620

First Director of Academic Computing – Gene Milner – hired

Research Computing 1972



IBM 360/50

1970 – [IBM 360/50 CPU](#) purchased to support both academic and administrative computing.

Research Computing 1982

1980: National Advanced System AS/5000 4 megabyte CPU and 3.5 billion byte disk storage subsystem purchased; academic disk storage was set to 1.1 billion bytes, 4.7 times that available on the IBM 360/50, which was retired.

SOFTWARE ACQUISITIONS SINCE JANUARY 1981

Following is a list of software that has been acquired by the Computing Center since January 1981, along with the Operating System(s) that it operates under. Some of the items on the list are renewals or updates of software that was already available.

ADABAS (MVT)
COM-LETE (MVT)
Strong-Campbell Interest
Inventory (MVT)
LINPACK (MVT)
BMDP (MVT)
VM-CMS Interface (MVT)
COBOL ANS Version 4 (MVT, CMS, MUSIC)
SAS/ETS (MVT)

NATURAL (MVT)
WATFIV-S (MVT)
Multi-Dimensional Scaling
Package (MVT)
PRIME (MVT)
Comput-A-Charge (MVT)
WATERLOO SCRIPT (MVT, CMS)
SAS Release 79.5 (MVT)
SPSS Version 9 (MVT)

Research Computing 1992

1991: Solbourne 52/902 Sun-compatible UNIX system (Sol) acquired for academic and research use.

SOLBOURNE USAGE STATISTICS

November Top Ten Programs: CPU Time Used

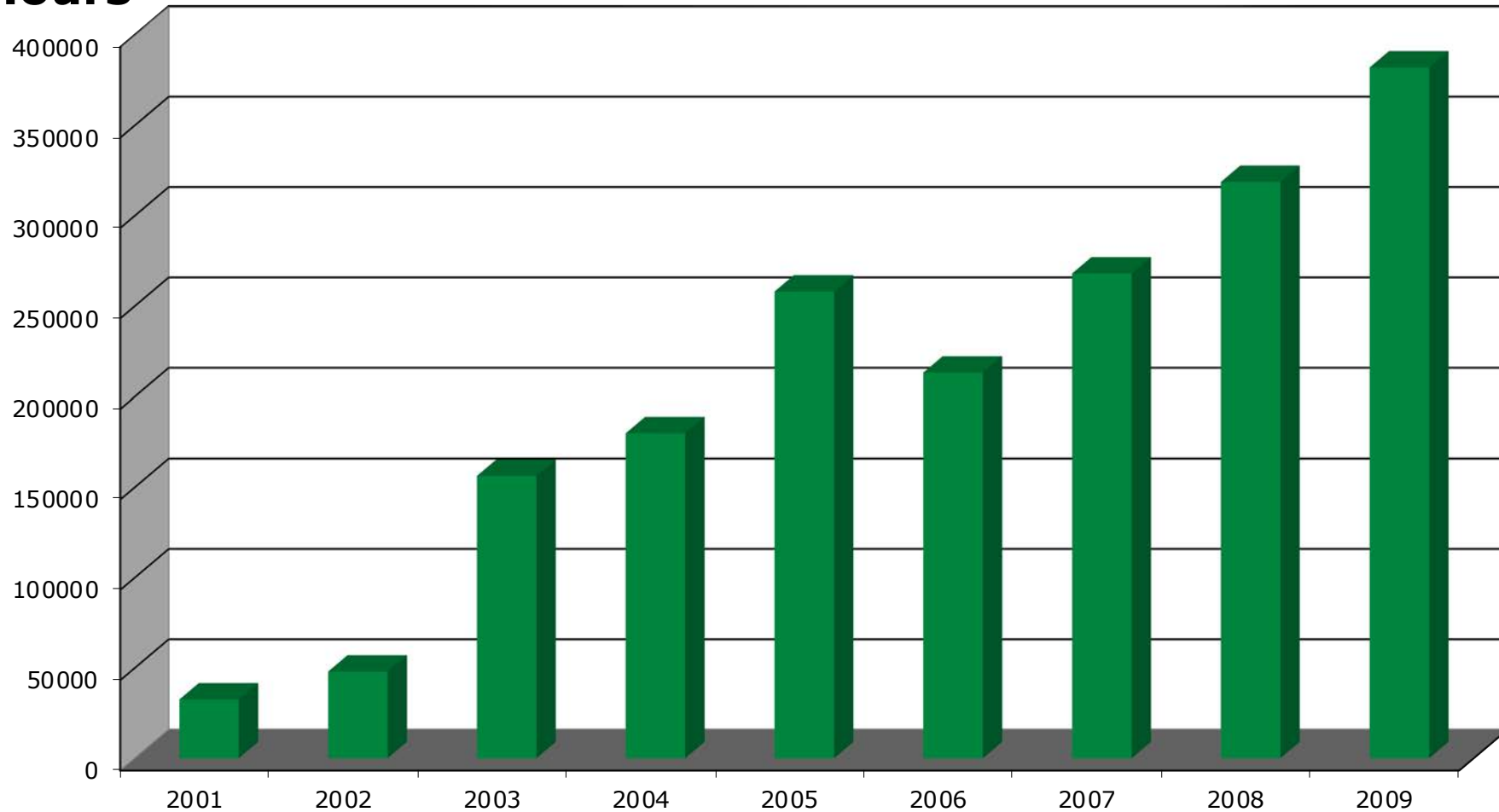
Program	Description	CPU Minutes	Percent of Total
1. 19999.es	User Program	10417.4	30.2
2. g90	Gaussian 90	6499.8	18.8
3. wod2b	User Program	3445.7	10.0
4. wod2a	User Program	3347.9	9.7
5. wod2c	User Program	2819.4	8.2
6. nntpd	USENET News Xmit Daemon	639.5	1.9
7. find	UserProgram	565.6	1.6
8. in.telne	User Program	392.1	1.1
9. relaynew	User Program	385.0	1.1
10. 11002.ex	User Program	321.7	0.9
Total		28834.1	



Research Computing 2002

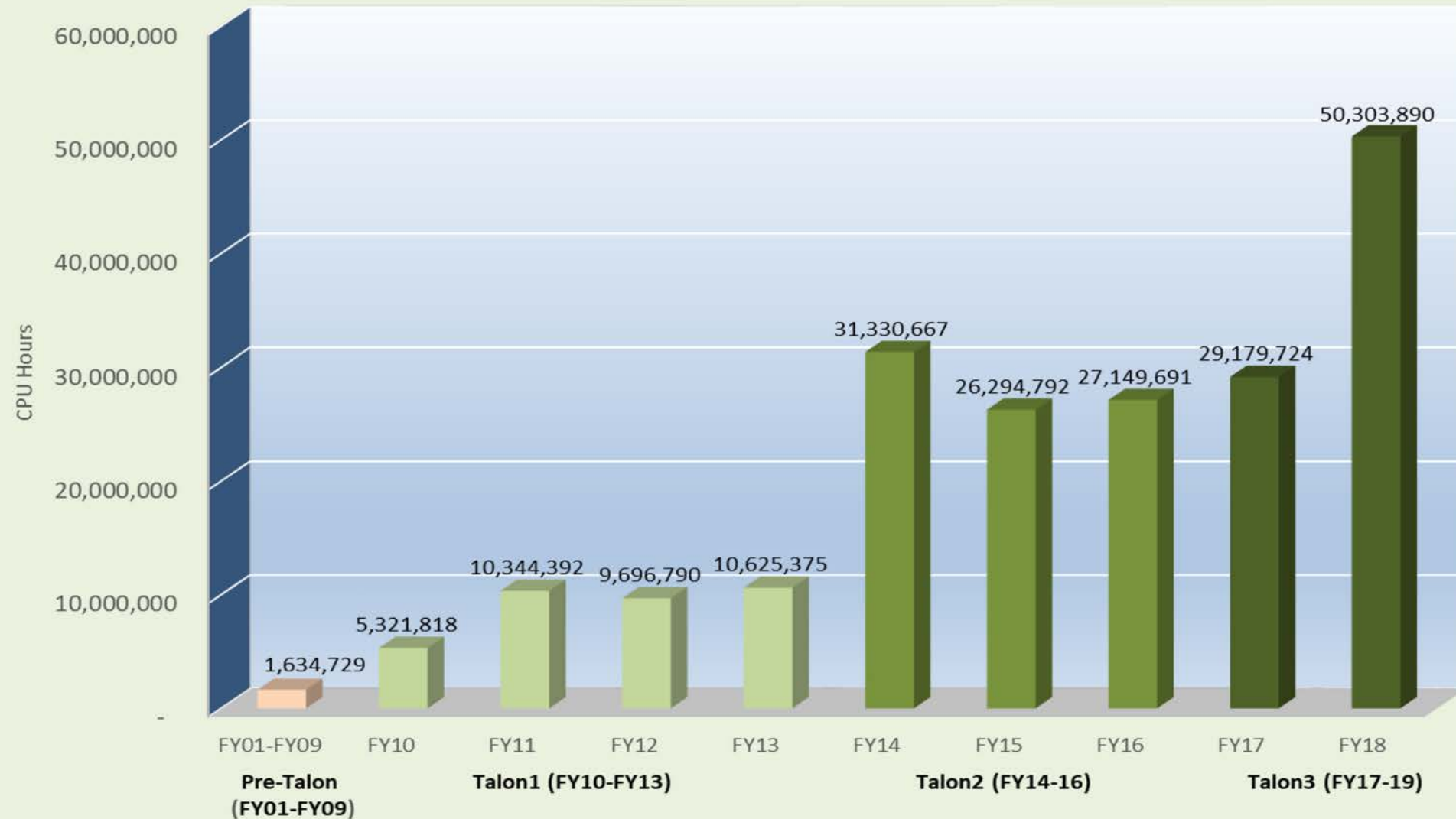
HPC Usage by Year—Before Talon

CPU Hours



Research Computing 2012–2018

HPC Usage by Academic Year/Cluster Generation



Research Computing 2022



Talon 2020

- Nodes: TBD
- Cores: TBD
- Memory: TBD
- Storage: TBD

Your input will help create the next generation of research computing systems at UNT.

Strategy and Impact

Convergence of computing and analytics is an emerging paradigm in research IT

■ Facts about UNT

- Integration of HPC and DSA in one office = Research IT Services;
- 32,000+ students enroll in non-STEM programs annually;
- Nine non-STEM colleges out of 12;
- Societal value — key aspect in grant applications;
- Paradigm shift: UNT's unique position.

■ How do we go about it?

- STEAM: STEM + Arts
- Train future workforce in leveraging IT in their domain areas!
- Foster industry-research partnerships in convergence areas!
- Develop new research and training opportunities



Vision, Mission, Goals

- To empower research through information technology.
- To promote, support, and strengthen computational research and training capabilities on campus through integrating leading-edge high-performance computing, data science & analytics, and visualization capabilities for faculty, staff, and students.
- To serve as a common denominator and support UNT's innovation capacity across both STEM and non-STEM areas.
- To help developing industry-research partnerships leading to identify, prototype, and accelerate convergence of both computational and data intensive applications and capabilities.
- Create industry-savvy workforce that can help strengthen research innovations and economy.

What are Research IT Services?

- ❑ Staff support through University Information Technology, University Information Services & Division of Finance and Administration;
- ❑ Office of the Vice President for Research and Innovation supports research computing and storage hardware;
- ❑ We manage research computing, storage and visualization capabilities—both university-owned and researcher-owned computing equipment.

Services

- ❑ Install, monitor and maintain Talon HPC cluster hardware and software services;
- ❑ Provide consulting services, periodic and customized trainings in leveraging current and emerging research computing capabilities and programming environments;
- ❑ Collaborate, design, and support computational and data-intensive research, and co-mentor students;
- ❑ Support and collaborate with faculty and researchers in new research and grant development opportunities.

it.unt.edu/RITS

Meet the RITS team

High-Performance Computing



DaMiri Young, HPC Team Lead
System Admin
User Support



Charles Peterson, PhD
Computational Scientist
System Admin
User Support
Research Development



Yuguang Ma, PhD
Computational Scientist
System Admin
User Support

Data Science and Analytics



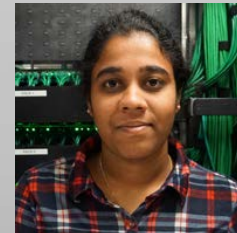
Richard Herrington, PhD
Data Science & Analytics
User Support
Research Development



Jonathan Starkweather, PhD
Data Science & Analytics
User Support
Research Development

Part-time Staff / GRA / Student Assistants

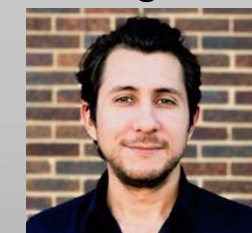
Sharanya



Ramya



George



High-Performance Computing

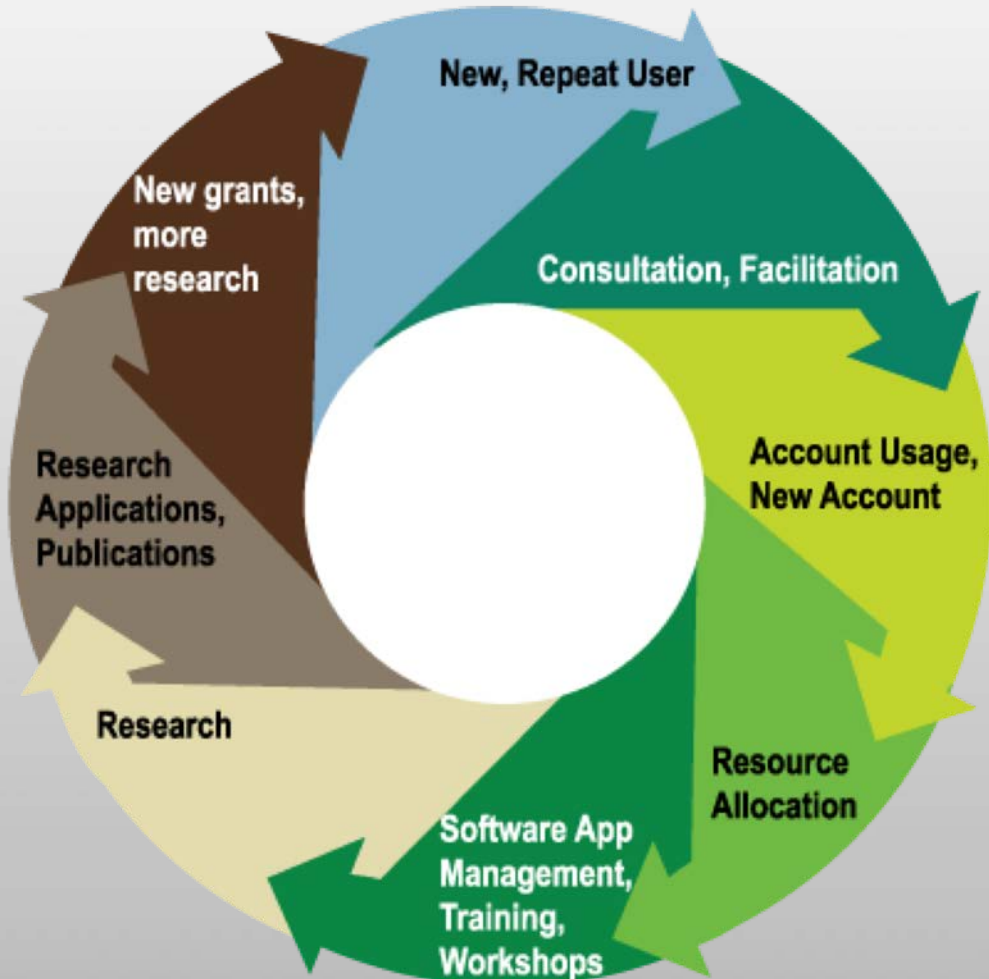


Talon—UNT's flagship computing system

- ❑ Over 8,300 CPU cores
- ❑ Over 150,000 GPU computing cores
- ❑ Mellanox 56 Gbps (FDR) InfiniBand network
- ❑ 1.4 Petabytes of Lustre computing data storage
- ❑ 700TB object storage for general data storage
- ❑ Workload manager for job scheduling with SLURM handles computing resource submissions.
- ❑ Various programming languages and libraries
 - ❑ Fortran, C/C++, Python, Java, CUDA, etc.
 - ❑ Mathematica, MATLAB, R, TensorFlow, Spark, Hadoop, and others
- ❑ Scientific software
 - ❑ Gaussian, VASP, AMBER, NWChem, among others
- ❑ Open-ticketing request system for issue management.

Research IT Services

HPC research life cycle



HPC services

- ❑ Provides consulting services in performing research on HPC
- ❑ Provides end-user support for using HPC resources
- ❑ Community training and outreach with workshops
- ❑ Provides HPC resources for UNT courses

The HPC team supports

- ❑ 350+ users,
- ❑ 80+ research groups,
- ❑ 18+ STEM and Non-STEM departments.

What is HPC and why should I care?

What is High-Performance Computing?

- HPC aggregates computing power and techniques in order to solve large or *intractable* problems—fast computing.
- Enables model size to scale manifold.

Why should I care?

- Is your research computing taking hours or days to complete on a desktop? Are you limited by model size due to its run-time limitations?
- Is your research data growing or disparate across sources and potentially unmanageable with current tools?
- Do you anticipate scaling the problem beyond what you currently work with?

Why is HPC important?

HPC enhances research outcomes by facilitating:

- ❑ Collaboration
- ❑ Cost reduction
- ❑ Enhanced Productivity
- ❑ Performance
- ❑ Reduced time-to-solution
- ❑ Scalability

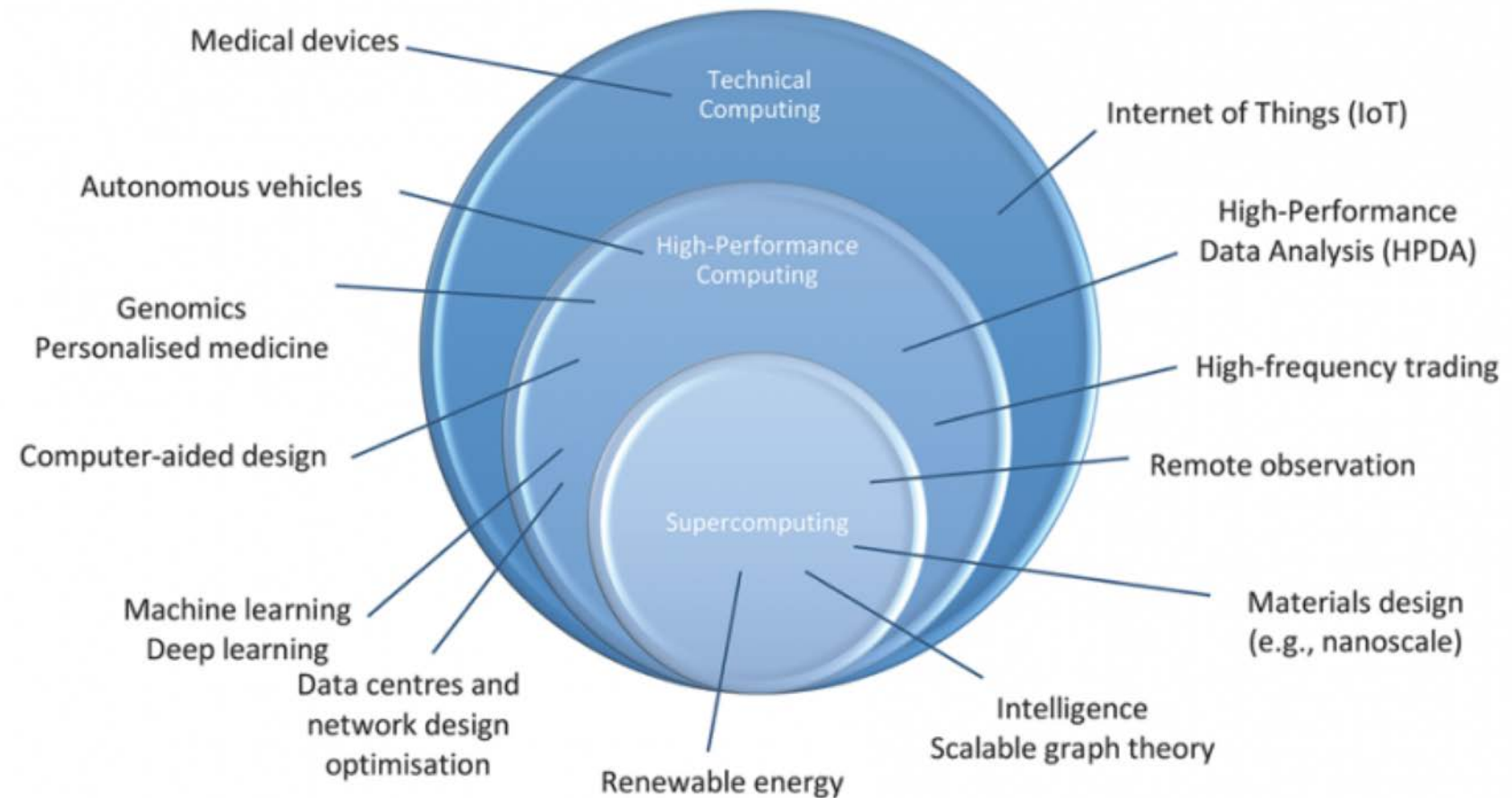
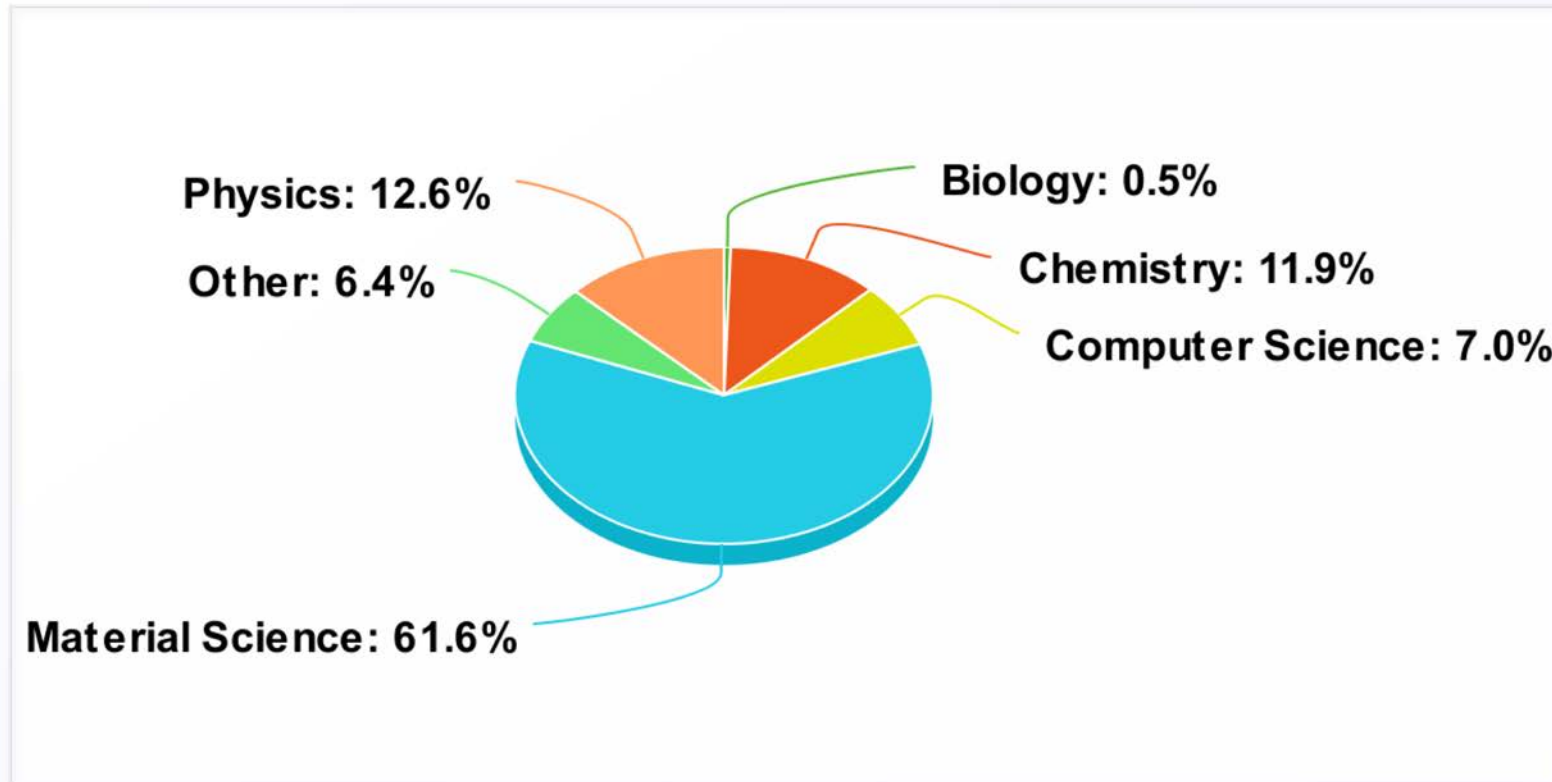


Figure: Relevance of HPC to strategic and emerging domains

Who uses central HPC on campus?

FY18 CPU Core Hours Utilized by Department

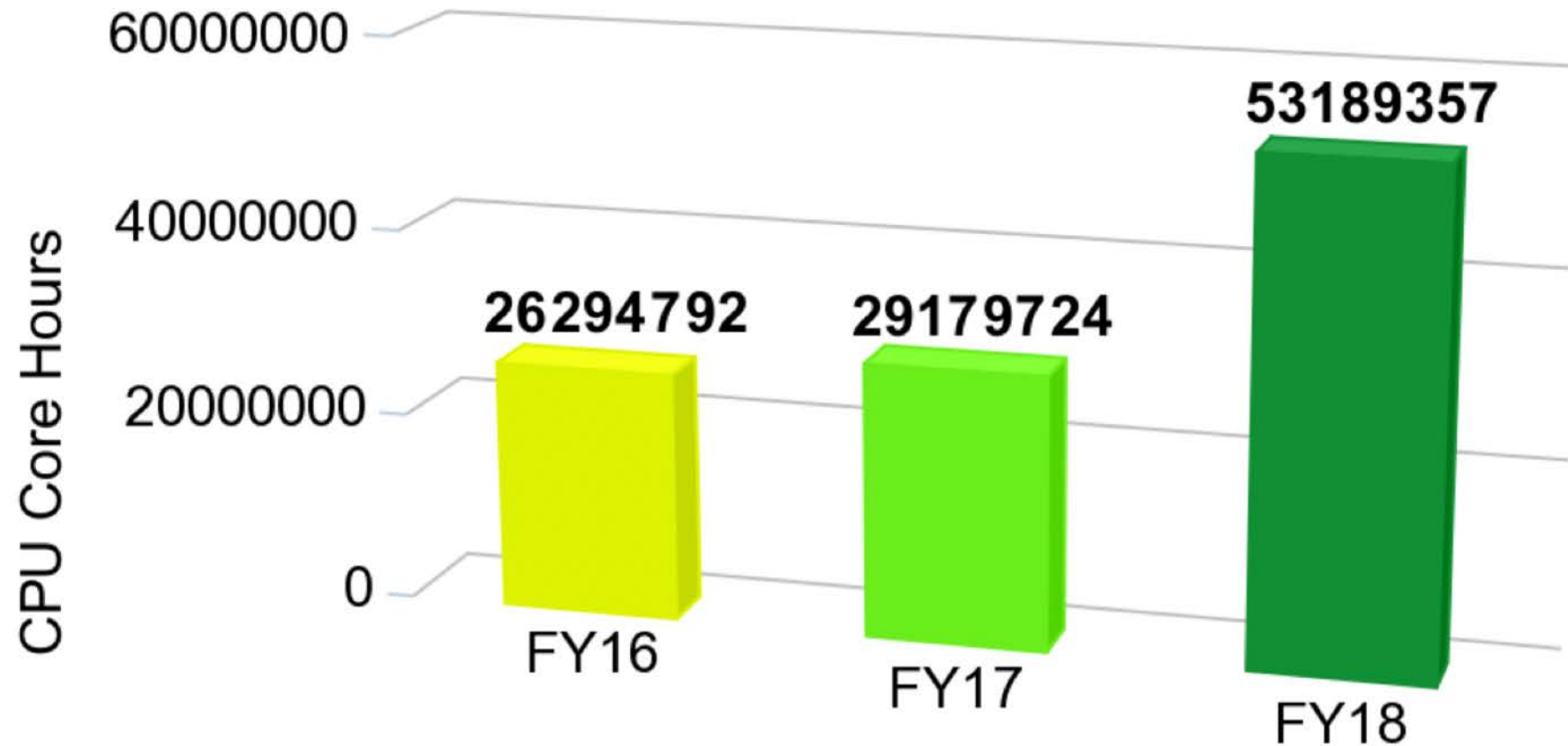


■ Biology ■ Chemistry ■ Computer Science ■ Material Science ■ Other
■ Physics

meta-chart.com

Who uses central HPC on campus?

CPU Core Hours FY 16-18



More of why HPC is important...

Return on Investment Summary	
FY16-FY18 Total Cost (Talon 3 infrastructure + Staff)	\$3,043,938
FY18 Annual Cost	\$1,014,646
Computational Research Funding Generated in FY18	\$7,683,269
FY18 Return on Investment	\$7.54/\$1
Itemized Cost	
Number of researchers	380
Cost per researcher / year	\$2,670
Number of CPU Cores	8,304
Cost per core	\$366
Cost per core / year	\$122

NSF XSEDE: We Can Help!

- ❑ eXtreme Science and Engineering Discovery Environment
- ❑ NSF funded virtual organization of advanced computing and storage infrastructure
- ❑ Ranks Under 10 in Top 500 in the world
- ❑ Science Gateways
 - ❑ Collaborative research & grant opportunities
- ❑ UNT campus champions: Charles Peterson and DaMiri Young
- ❑ We can help with:
 - ❑ Getting you an account on XSEDE
 - ❑ Allocations up to \$50K SUs
 - ❑ Migrating your codes!
 - ❑ Various research computing related training courses

More information at:
<https://www.xsede.org>



²Contact HPC

General Academic Building,
Room 535

Homepage: [*hpc.unt.edu*](http://hpc.unt.edu)

Email: [*hpc-admin@unt.edu*](mailto:hpc-admin@unt.edu)

Data Science & Analytics

What we can do for you!

DSA Consulting

- DSA personnel have expertise in mathematics, computer science, education, and psychology;
- DSA engages in critical thinking and problem solving with our clients; during this process, we start as a resource but become mentors, collaborators, and sometimes end up serving on student dissertation committees;
- Consultation sessions can be of short duration, multiple sessions, perhaps across multiple semesters;
- Consultation involves intensive, interactive, multi-directional conversations with interactive use of software, demonstrative simulations, visual displays, and (an increasingly) the use of cluster computing (i.e., Talon 3);
- Most of our customers to date are from non-STEM areas; we are very interested to explore opportunities in leveraging STEM and non-STEM areas.

Who are DSA's Clients and How Do We Support Them?

- Students, faculty, and administrators involved in research, theses and dissertations, grant writing and administrative projects
- We provide assistance with mathematical and statistical theory, research design, modeling techniques, interpretation of analyses, software implementations, and data management
- DSA maintains expertise on emerging trends in computational statistics, computing technologies, and best-practice standards
- DSA disseminates these technologies through teaching and training: one-on-one sessions, short course workshops, web-based tutorials, hands-on tutorials, guest lecturing, serving as adjunct instructors, and UIT's online publication, *Benchmarks*

DSA Software Application Support

■ **General purpose software**

- R (general purpose statistical programming language and statistical analysis environment – similar to MATLAB or Octave);
- Windows-based platforms, Debian and Cento-OS based Linux, Python, web technologies (HTML/CSS, JavaScript, Linux server administration);
- RStudio (a desktop and server-based R and Python IDE for browser-based analysis and visualization);
- SPSS, Stata, and SAS (popular statistical software packages).

■ **Specialized Software**

- MATLAB, Mathematica (numerical and symbolic analysis);
- Mplus, LISREL, AMOS, EQS (structural equations).

Research(er) Challenges

- **Typical data are messy and modelling with messy data is problematic**
 - Small data or really large data (so-called “curse of dimensionality”);
 - Missing data (drop-out, unit non-response) and outlying data;
 - Unidentified mixtures in the data (multiple modes or “bumps”);
 - Unidentified sources of hidden correlation in the data;
 - Complex data structures or data collection designs (hierarchical, unbalanced, and non-replicated designs);
 - Non-random sampling and assignment designs (observational data versus experimental)
 - Correlational versus causal modeling research goals – the mismatch;
 - Selection bias and measurement error.
- **Software algorithms can be fragile**
 - Many software algorithms can fail to converge or provide satisfactory solutions;
 - DSA will either find resolutions to these software convergence issues or suggest alternatives that are more robust and meet the goals of the researcher.

Meeting the Challenges of a Typical Researchers

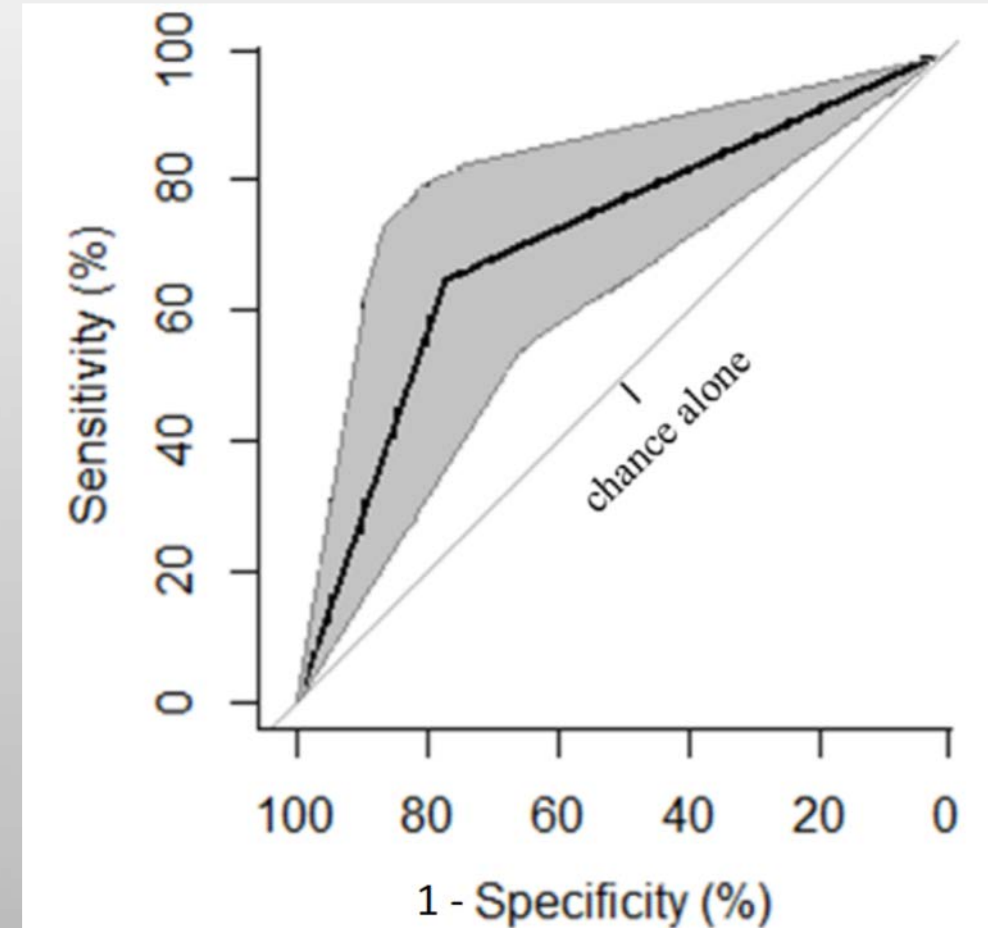
An Example of a Recent DSA Collaboration

- Biology graduate student (Sheela Sadruddin) working on dissertation under supervision of Warren Burggren – she graduated Spring 2018
- *Optimization of in Vitro Mammalian Blastocyst Development: Assessment of Culture Conditions, Ovarian Stimulation and Experimental Micromanipulation*
- Study goal - compare four fertility treatment protocols and identify fertility conditions and optimal treatment regime to facilitate fertilization of embryos
- Data dimensions $n=104$, $p=29$
- Main effect analysis using classical regression and ANOVA based methods failed to detect any significant relationships; too many possible interactions and too many potential models when all variables are considered in a full model
- Final solution: classification decision tree learning using a genetic algorithm to determine globally optimal pruning and globally optimal node splits using BIC as a criterion for model selection
- Model tuning (parameter alpha - a complexity penalty parameter) was conducted on Talon using more than 200 cores and more than 10 hours during the optimization process

Meeting the challenges of typical researchers

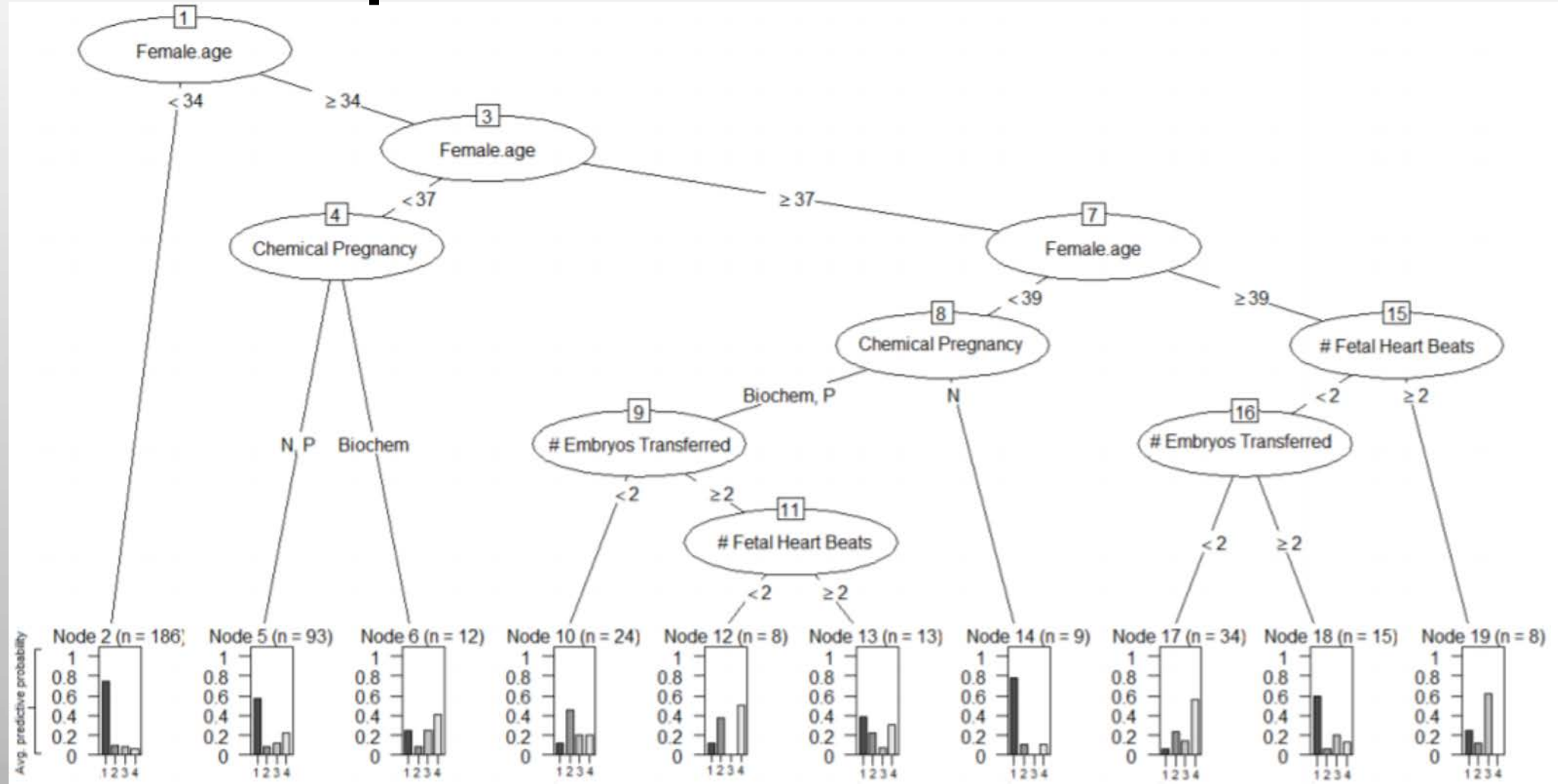
Example of a Recent DSA Collaboration

- The final classification tree yielded an accuracy of 71% with an error rate of 29%; the ROC curve had an area under the curve (AUC) of 0.71 with 95% confidence interval (CI) of 0.6- 0.8, indicating that any final node of the decision tree results in 10% greater success than chance alone as the lowest measure and up to 30% greater success than chance alone at its highest success.



Meeting the challenges of typical researchers

Example of a Recent DSA Collaboration



Recent DSA Grant Collaboration

Understanding Unplanned Hospital Readmissions: The Importance of Individual and Neighborhood Characteristics

- Grant was submitted to National Institutes of Health in fall of 2018
- \$434,000 over five years
- Dale Yeatts, Department of Sociology,
Chetan Tawari, Department of Geography,
Ami Moore, Department of Rehabilitation and Health Services,
Richard Herrington, RITS/Data Science & Analytics

More on DSA Instructional Services

- Creation of web-based software tutorials with realistic data science examples that use simulation or real-world data sets
- Interpretation and illustration of the core statistical or mathematical ideas underlying the software algorithms
- Interactive demonstration of software usage, modeling, analysis and interpretation, and in general, the workflow of a modelling session
- DSA provides customized and periodic training on using Talon 3 cluster computing (parallelization with SLURM batch submission); visualization nodes, RStudio, R, and Python based programming
- These resources and training experiences are increasingly integral to STEM and non-STEM researchers interested in soft computing, AI, and machine learning methods

Examples of other DSA data support services

- Past services have included:
 - Transferring data from a rudimentary electronic form into a statistical software package
 - Providing tools for scraping web data and transferring that data into statistical software packages
- DSA supports locating and downloading archived data, e.g., ICPSR data repositories
- DSA can supply realistic simulated data for classroom examples or homework
- These simulated data can provide researchers with a different view towards a better understanding of data structures and their related models
- Occasionally the simulation becomes integral to and part of the model itself, e.g., so-called generative models

⁵Contact DSA

Richard.Herrington@unt.edu

Jonathan.Starkweather@unt.edu

General Academic Building, Suite 535

DSA Homepage: <https://it.unt.edu/research>

Programs & Opportunities

RITS: ASSUR²E

- Application Support Services for User Recruitment and Retention
- Jointly funded through support from Offices of the Provost and Research Innovation.
- Supports student-leveraged research development and grant collaborations

Funding

- Supports one graduate and one undergraduate student costs
- Up to five competitive scholarships for graduate student in-state tuition
- Faculty and RITS Staff co-mentor students & share student costs: tuition or stipend

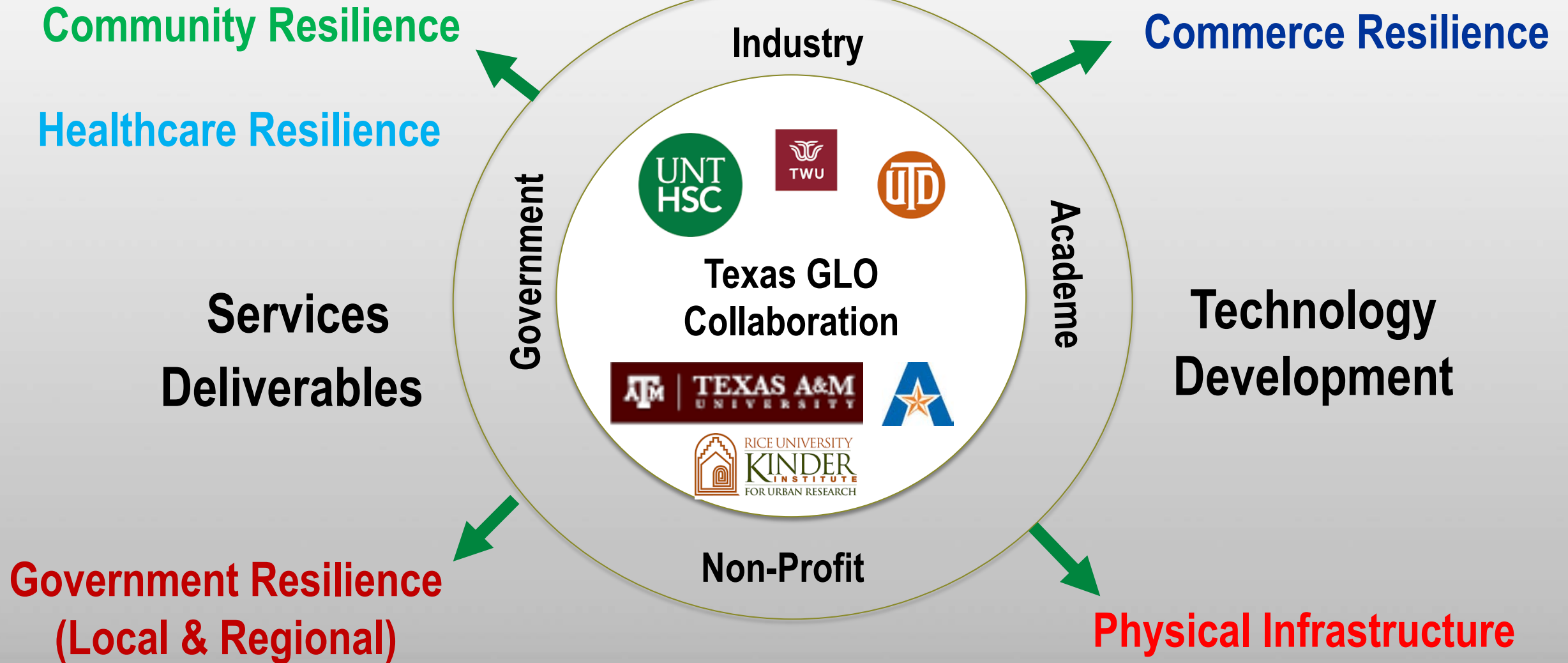
Metrics for Success

- Results in new research development / interdisciplinary collaborations
- At least one grant application submitted with a potential to generate 3x or more ROI
- Author/co-author at least one peer-reviewed publication to a reputed journal or conference

NSF Cloud & Autonomic Computing Center

- UNT was approved as an affiliate site: Oct. 23, 2018
- An MOU was executed with TTU that allows cost-sharing of indirect (up to 10% maximum)
- This new affiliation facilitates
 - New collaborative opportunities with other CAC member sites – TTU, UA, and Universidad de Sonora, Sonora, Mexico
 - Center's industry members,
 - Federal grant opportunities accessible to IUCRC programs, and
 - Recruiting industry members in the Dallas-Fort Worth region
- UNT is actively developing NSF-funded site on “Integrated Modeling and Analytics in the Cloud”
- Together with this program, Research IT services can support
 - STEM and non-STEM collaborations
 - New research and grant opportunities, and
 - Industry savvy workforce training and economy
- We are recruiting project leaders with research aligned with industry scope!

North Texas Consortium for Data, Analytics, Research & Engagement



Disaster Resilience & Community Engagement

Across STEM and Non-STEM Areas



Healthcare Resilience

Economic impact of health care disruptions

Open-drone for emergency response

Event mapping using LiDAR & social media

Virtual reality for emergency responders

Integrated Modeling and Analytics

Commerce Resilience

Labor market recovery

Local government finance recovery

Environmental quality assessment

High-performance Computing

Edge-cloud big data analytics

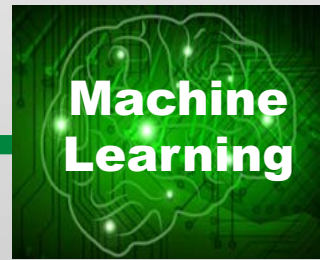
Trustworthy information sharing

Government Resilience

Autonomous vehicles

Capability building for disaster resilience

Community Resilience





Thank you!

Research IT Services

UNIVERSITY
INFORMATION
TECHNOLOGY
UNT[®]
EST. 1890